International Journal of Education & Literacy Studies
ISSN: 2202-9478
www.ijels.aiac.org.au

Optimizing Oral Proficiency Assessment in Chinese as a Second Language: Challenges and Improvement Strategies of the OPI Test

Na Liu, Jeferd Saong*

Graduate School, University of Baguio, Baguio City, Benguet, Philippines

Corresponding author: Jeferd Saong, E-mail: jeferdsaong@e.ubaguio.edu

ARTICLE INFO

Article history
Received: January 02, 2025
Accepted: March 22, 2025
Published: March 31, 2025
Volume: 13 Issue: 2

Conflicts of interest: None
Funding: None

ABSTRACT

The study examined the use of the Oral Proficiency Interview (OPI) in assessing Chinese as a second language and the challenges faced by teachers in the Chinese Language Scholarship (CLS) program. The Concurrent Triangulation Mixed Method Research was used in the study in which qualitative and quantitative data are collected simultaneously, analyzed separately, and then compared or integrated to achieve a comprehensive understanding of the research problem. Quantitative data primarily consists of Oral Proficiency Interview (OPI) test scores, which supplied the admission, pre-test, and posttest scores. Semi-structured interviews with five teachers were conducted to identify challenges in administering the OPI. The findings revealed that technical and cultural differences, examiner subjectivity, and environmental factors influenced students' OPI performance. Key results emphasized the importance of understanding students' backgrounds for adaptive questioning while maintaining standardized assessment practices. To enhance OPI effectiveness, the study recommends further examiner training, continuous professional development, and the integration of technology, such as AI, into the assessment framework. Overall, the CLS program should adopt more effective teaching and learning strategies to improve language proficiency outcomes and literacy.

Key words: Oral Proficiency Interview, Language Scholarship, Second Language Literacy, OPI Challenges, Assessment, AI

INTRODUCTION

The recognized evaluation systems for Chinese as second language in the world include (a) Common European Framework of Reference for Languages (CEFRL), (b) American Council on the Teaching of Foreign Languages (ACTFL), and (c) International Chinese Language Education Chinese Proficiency level standards. The evaluation systems focus on the oral proficiency interview (OPI) for the assessment of language speaking ability. OPI assesses the ability to use language effectively and appropriately in real-life situations. The OPI under the ACTFUL framework is proficiency-based, and provide a means of assessing the proficiency of a foreign language speaker. OPI provides reference for Chinese as a second language oral assessment. American Council on the Teaching of Foreign Languages (ACTFL) is an organization aiming to improve and expand the teaching and learning of all languages at all levels of instruction.

ACTFL Proficiency Guidelines (2012) introduced Oral Proficiency Interview OPI, a live 15- to 30-minute telephone conversation between a certified ACTFL Tester and the candidate. The OPI assessment is a valid and reliable test that measures how well a person speaks a language. The procedure is standardized to assess global speaking ability, measuring language production holistically by determining patterns of strengths and weaknesses.

The assessment result in OPI served as a reference in the Critical Language Scholarship (CLS) program in China. International students have to take three OPI tests, during admission, the placement test after coming to China, and the assessment after the project. The results of these three OPI tests often exceed the expectations of teachers, which attracts the attention of project managers. The study seeks to put forward relevant strategies based on the actual OPI of the students.

LITERATURE REVIEW

Bijani (2019) pointed out "the Oral Proficiency Interview (OPI) is a widely used tool for assessing language learners' speaking abilities" This set of measurement methods can be used not only in English assessment, but also in Japanese, Chinese and other languages. Relevant scholars have evaluated and analyzed its applicability in Japanese assessment. "OPI's short test time and ease of use have contributed to its widespread adoption in language research and education, particularly in Korea for Japanese language studies" Kawaguchi et al. (2020). The advantages of OPI include (a)

Published by Australian International Academic Centre PTY.LTD.

http://dx.doi.org/10.7575/aiac.ijels.v.13n.2p.605

simple and flexible question setting, (b) short test time, (c) easy to use, and (d) low operation difficulty. On the contrary, McConnell (2022) critics argue that assessment activities, including OPI, have become burdensome accountability exercises rather than tools for improving student learning OPI poses (a) unreasonable grade determination, (b) lack of differentiation, and (c) inability to objectively judge the starting point of the test.

Second language learners can accurately self-assess their oral proficiency using tools like the NCSSFL-ACTFL Can-Do Statements, particularly at Novice and Advanced levels (Ma & Winke, 2019). Peng et al. (2020) pointed out: "An independent oral test, Hanyu Shuiping Kouyu Kaoshi (HSKK), complements the HSK for speaking assessment." According to Akbari, (2020) "While standardized tests like OPI and HSK are widely used, there is a growing emphasis on authentic, performance-based assessments in classroom settings to provide ongoing evaluation integrated with instruction". These diverse assessment methods aim to comprehensively evaluate Chinese language proficiency across various skills and contexts.

ACTFL (2021) present colleges and universities are using OPI results for student placement in language courses and awarding college credit for demonstrated proficiency. This allows students to bypass introductory courses and enroll in advanced classes directly. (ACTFL, 2020) Institutions employ OPI results to evaluate the effectiveness of their language programs and make necessary adjustments to their curricula to meet proficiency standards.

The OPIc is owned by ACTFL and administered by LTI, which holds an exclusive license (ACTFL, 2018). While ACTFL and LTI retain all administrative and content control of the exam, this development, along with growing test-taker numbers, suggests that in the future the OPIc may expand its reach and relevance both in the United States and abroad.

Government agencies and the military utilize OPI results for certifying language proficiency, which is essential for roles requiring specific language skills such as foreign service officers, translators, and intelligence personnel (Language Testing International, 2021). Proficiency certifications obtained through OPI assessments are often prerequisites for security clearances and can influence promotion decisions within these agencies (ACTFL, 2021).

Oral tests are more appropriate for low-level test takers, and semi-direct tests for higher ability ones. Data analyses demonstrated no significant difference between the ratings of direct and semi-direct oral assessment by the raters. Consequently, semi-direct oral tests can be regarded as a reliable substitute for direct oral tests.

Bijani (2019) argues that Oral Proficiency Interviews (OPI) can be effective for assessing language learners' speaking skills, but rater training is crucial for reliable scoring. Studies show that rater training improves interrater consistency and reduces severity/leniency bias. Rater training is essential for maintaining rating consistency and understanding assessment criteria. Rossi and Brunfaut (2020) recommended to improve reliability and validity in classroom assessment, teachers can employ strategies such as multiple raters and communal rating sessions; however, while rater training can enhance consistency, it may also increase differences in severity. Overall, these studies suggest that OPI can be valuable for language assessment when combined with proper rater training and reliable rating scales.

Theoretical Framework

The establishment and development of the Oral Proficiency Interview (OPI) system have been influenced by multiple theories, all of which emphasize the communicative and contextual nature of language learning. OPI comprehensively assesses the oral proficiency of language learners through interactive tasks set in real-life scenarios, reflecting the latest principles and practices in modern language teaching and assessment.

Communicative language teaching (CLT)

CLT emphasizes the communicative function of language rather than just language forms. It posits that the ultimate goal of language learning is to enable effective communication in real-life situations. Wang and Zhang (2023) stated emphasizes the importance of assessing language learners' communicative abilities in context, beyond just grammar and vocabulary accuracy. The test assesses learners' language skills in different situations through a series of realistic conversational tasks, emphasizing fluency, appropriateness, and communicative function in language use.

Sociocultural theory

Proposed by Lev Vygotsky, this theory asserts that language learning is achieved through social interaction and cooperative activities, emphasizing the interaction between learners and more experienced language users.

Vygotsky's sociocultural theory emphasizes that language learning occurs through social interaction and cooperative activities. This theory posits that cognitive development is mediated by social relationships and cultural context. Language is viewed not only as a communication tool but also as an instrument of thought that enables individuals to internalize knowledge (Silva et al., 2024).

Task-based language teaching (TBLT)

TBLT stresses learning language through completing meaningful tasks, which should simulate real-life communication needs. Task-Based Language Teaching (TBLT) emphasizes learning through meaningful tasks that simulate real-life communication needs This approach engages learners' natural language acquisition abilities and focuses on form through task performance (Ellis et al., 2019).

OPI test process

In OPI testing, a trained and certified test officer will test the language proficiency of the test subject, which will have four processes to perform. The first part, warm-up, aims to make

students not feel unfamiliar and nervous, adapt to the exam environment, but also bring the subjects into a relatively daily natural state to "chat", such as: Tell me a little bit about yourself and your normal daily routine. Tell me a little bit about yourself and your best friend.

The second part usually asks about hobbies, places you visit, people and things in your community, or at your school, such as: Tell me how you first met one of your neighbors. Describe in detail when you met and everything that happened during your fist few meetings. Who is your favorite character from any movie/TV show? Why is that character your favorite? Describe him/her in detail. A series of questions are asked to see if the candidate can make a detailed description, can make the overall output of the sentence, how to organize the connection of the context, and what is the fluency and appropriateness of the language expression. In this process, the examiner seeks the highest level of its showed, like "ceiling" its ability could reach, meanwhile, the examiner determines the lowest level should reach;

In the third part, the examiner determines the grade of the subject according to the performance of the second part, so the second part is crucial. If it is determined that the subject is intermediate, the examiner may choose to set up a scenario according to which the subject will role-play, such as: 1) I'd like to give you a situation and ask you to act it out. Imagine you want to take dance lesson. Call a dance school and ask three or four questions about dance classes. 2) I'm sorry. There's a problem you need to resolve. The ticket agent tells you that the event is almost sold out. There are only a few tickets left, and the seats are not together. Call your friend and leave a message explaining what the situation is and offer three or four alternatives. If the test is determined to be advanced level, the examiner will ask the topic of comparison, talking about politics, history, news, sports reports and other current political issues, please talk about your views, make some abstract generalization or horizontal comparison. For example: Express (and sometimes support) opinions on abstract issues. In this part, the examiner will determine the language level of the subject based on the quality of the tasks completed.

In the fourth part, the examiner wind down the interview.

METHOD

Research Design

The study adopted Concurrent Triangulation Mixed Method Research design in exploring the professional development of Calligraphy teachers' discipline integration. In this study, quantitative data primarily consists of Oral Proficiency Interview (OPI) level. Qualitative data is collected through teacher interviews to explore the challenges they face in administering OPI assessments.

Sample and Sampling Procedures

Purposive sampling was used in the study for the selection of five teachers in the CLS project who are from the department of international Chinese group that were interviewed. The five interviewees received OPI test training and situational teaching training. Total sampling of the eight students in the CLS program was used in the study.

Instrument

Standard OPI tests administered by the teachers were used in determining the student's practical level of OPI. A set of researcher-made interview questions were utilized to explore the challenges faced by teachers when administering OPI (Oral Proficiency Interview) tests for Chinese language proficiency.

After the interview was finalized, the interview data was imported into a software produce the themes analyzed according to the subject division in turn. The study adopted admission pre-test and post-test records to determine the students' practical level of oral proficiency as the background and used NVivo software to help organize, manage and analyze the interview transcript in determining the practical level of oral proficiency and challenges.

RESULTS AND DISCUSSION

Students' Practical Level of Oral Proficiency

Oral Proficiency Interview (OPI) is a standardized assessment tool used to measure an individual's ability to speak a language. It evaluates oral communication skills in a natural and interactive format (Siraranghom, 2024). Foreign students enrolled in Chinese universities submits OPI (Oral Proficiency Interview) test scores obtained in their country of origin as part of their placement assessments. However, upon retaking the OPI test in China for placement purposes, discrepancies were observed between the scores from the U.S. and those obtained in China as shown in Table 1. Table 1 shows the OPI (Oral Proficiency Interview) levels during admission (students' OPI scores from their home countries), pre-test (OPI scores before class placement), and post-test (OPI scores at the end of the program). It is evident that seven (7) students OPI level from admission to port-test differs while one (1) remains the same (advanced-mid). The students OPI level between admission and pre-test yielded five (5) similar levels, two (2) with higher admission OPI level, and one (1) with higher pre-test admission OPI. Level. The difference in OPI level can be contributed by non-target and target language environment. The admission OPI level serves as the basis for the placement of students. However, the CLS program implemented flexible mechanism to retain or change the placement based on the OPI level at the pre-test.

This approach benefits not only individual students' language progression but also supports teachers' instructional strategies and ensures the program advances cohesively.

Comparing pre-test and post-test results indicates that the 8-week immersive program led to varying degrees of improvement based on students' personal efforts. However, it is crucial to note that language proficiency gains cannot solely be attributed to individual effort. While immersion offers opportunities for linguistic and intercultural growth, outcomes

Table 1. Students OPI level in admission, pre-test and post-test

Student	Admission	Pre-Test	Post-Test
1	Novice-high	Novice-mid	Intermediate-low
2	Advanced-mid	Advanced-mid	Advanced-mid
3	Novice-high	Intermediate-mid	Intermediate-high
4	Intermediate-mid	Intermediate-low	Advanced-mid
5	Intermediate-low	Intermediate-low	Advanced-low
6	Intermediate-mid	Intermediate-mid	Intermediate-high
7	Intermediate-low	Intermediate-low	Advanced -low
8	Novice-mid	Novice-mid	Novice-high

are highly variable due to environmental factors and individual differences (Jackson & Schwieter, 2019). Factors such as intercultural communication, cultural integration, and personal emotions also play significant roles. For instance, students with lower language levels did not improve as quickly as those in the intermediate range. This is partly due to their limited language skills, which hindered their ability to make local friends or participate in more cultural activities, slowing their language progress.

Similarly, advanced-level students showed less significant improvement compared to intermediate-level students. Their existing language knowledge and skills were already sufficient for their current intercultural lives, so they did not face as many challenges requiring rapid adaptation. However, their progress in specialized academic language areas was limited due to a lack of focused learning and interaction in such domains. While advanced students may adapt more easily to intercultural environments, their progress in specialized academic language can be limited (Achirri, 2021). A problem that project managers need to face is how to improve the academic and special Chinese of high-level Chinese learners in the project. These findings highlight the importance of tailoring language training to specific academic contexts and professional profiles (Candel-Mora, 2019).

Challenges Encountered at the Critical Language Scholarship Program

In the "techniques for assessing proficiency", varied strategies are employed by examiners to ensure accurate assessments. Interviewee 3 emphasized understanding the candidate's background, such as their learning history and cultural experiences, to design appropriate questions. Interviewee 4 detailed differentiation techniques, noting: "We assess based on sentence complexity, fluency, and naturalness." These techniques highlight the importance of flexibility and adaptability in the assessment process. "Measures of complexity, accuracy, and fluency as important dimensions for assessing second language performance." (Winke & Brunfaut, 2020). Secondary probing, as described by Interviewee 4, ensures that ambiguous responses are clarified, reducing the risk of misjudgment. In practice, flexible strategies allow examiners to adjust their assessment methods according to specific situations. However, it is crucial to ensure that such flexibility does not lead to inconsistencies in assessment standards or excessive subjectivity. The lack of techniques employed by the examiner will influence the results of OPI. In practice, standardized tools should be combined with personalized assessment methods to achieve a balance between fairness and accuracy. This balance is essential for both second language learners and examiners.

In the "reasons for score disparities", score disparities were attributed to factors such as examiner subjectivity and environmental differences. Interviewee 3 noted that familiarity with the test format could skew results: "Prepared answers during narration often give an impression of fluency, masking deeper issues." "Oral proficiency assessments face challenges in reliability and validity due to various factors. Examiner subjectivity can lead to discrepancies in scoring, with examiners potentially awarding the same score for different reasons or being influenced by familiarity with a candidate's accent" (Bucher, 2019). This further emphasizes the importance of dynamic questioning techniques. Through flexible question design, examiners can more effectively evaluate candidates' genuine language proficiency. This underscores the need for examiners to employ dynamic questioning techniques to uncover genuine proficiency. If the examiner does not enhance the awareness of secondary confirmation, the student will use his preparation for the exam to hide his true level, such as the reality of a student's performance is different from his true level.

Within the "challenges in OPI assessment", examiner subjectivity is a key factor affecting scoring reliability. Interviewees 1 and 2 emphasized that different examiners often have varying interpretations and focuses regarding scoring criteria. Kinney et al. (2021) found that assessor training increased awareness of subjectivity. For instance, some examiners may prioritize fluency while overlooking grammatical accuracy, whereas others might consider complexity or naturalness as core indicators. This subjectivity not only leads to inconsistencies in scoring standards but can also be exacerbated by cultural or linguistic biases. For example, examiners may be inclined to assign higher scores to candidates with familiar accents or expressions, while being stricter toward unfamiliar accents.

Moreover, Interviewee 4 highlighted that the inefficiency of remote training further intensifies this issue. Virtual sessions lack the interactivity and practical components necessary to provide comprehensive guidance for examiners.

Online training environments can lead to reduced social interaction, engagement, and attention from participants. Many dynamic language interaction scenarios cannot be realistically simulated in remote training, making it challenging for examiners to effectively handle complex situations during assessments. Additionally, technical and environmental constraints may reduce engagement and the overall effectiveness of training.

In suggestions for training, training is undoubtedly a cornerstone for enhancing the reliability and consistency of oral proficiency assessments. The emphasis on practical course demonstrations and regular retraining, as mentioned by the interviewees, reflects a growing awareness of the need for dynamic and experiential learning methods. Integrating hands-on exercises, as suggested by Interviewee 2, not only makes training more engaging but also ensures that examiners are better equipped to apply theoretical principles in real-world scenarios. "The importance of hands-on exercises in training across various fields, integrating practical scenarios from industry certifications can bridge the gap between theory and real-world challenges. This approach acknowledges the complexity of oral proficiency assessments, which often require quick judgments in dynamic and diverse language use situations.

Moreover, Interviewee 4's focus on addressing translation issues is particularly significant in a global context. Challenges arise in oral examinations when conducted in different languages, necessitating accurate translation processes to ensure validity (Carr & Sun, 2021). Non-native examiners may face unique challenges in interpreting assessment guidelines or candidate responses, which could inadvertently affect their evaluations. Ensuring clarity in training materials not only supports non-native examiners but also fosters inclusivity and equity in assessment practices. Overall, these insights highlight the importance of making training more adaptive, inclusive, and grounded in practical application.

In future improvements for OPI, innovative solutions such as AI and big data were proposed to streamline the assessment process. Interviewee 3 noted: "AI could assist in designing tailored question chains, reducing examiner workload and enhancing efficiency." The integration of Artificial Intelligence (AI) and Big Data in education offers promising opportunities for enhancing assessment processes, personalized learning, and administrative efficiency (Gardner et al., 2021). By leveraging big data, patterns in language usage and proficiency levels could be analyzed to refine scoring criteria and improve the consistency of evaluations across diverse candidates. Lifelong learning opportunities for examiners were also emphasized, ensuring that they remain updated on best practices and emerging trends in language assessment.

Interventions for the CLS Program

To mitigate the impact of OPI (Oral Proficiency Interview) discrepancies and challenges on the CLS program, the program team can adopt a series of detailed and effective measures.

First, improving examiner training and assessment consistency is crucial. Interactive and hands-on training programs, such as role-playing and simulated OPI scenarios, can better prepare examiners for diverse linguistic and cultural contexts. "Evaluated trainee performance by calculating their level of agreement with expert consensus ratings" (Ortega, et al., 2023). Standardized scoring rubrics and regular calibration sessions can further minimize subjective biases and discrepancies in evaluations. In the training, Chinese should be enhanced as a purpose-specific training, and not with the help of translators, but with senior experts who speak Chinese as their native language, so as to reduce learners' translation and misreading of foreign languages

Second, tailored support for immersion students at different proficiency levels is essential. Lower-level students can benefit from structured cultural interaction programs, while advanced learners need targeted workshops focused on specialized academic language development. Flexible placement mechanisms, such as mid-program reassessments, can also ensure students are aligned with their true language abilities. "The development process involved an in-depth analysis of the test's components to ensure that learners become familiar with and confident in the test format This can reduce candidates' doubts and anxiety"(Siraranghom, 2024). Be fair to students from different cultures and backgrounds "students hailing from diverse cultural backgrounds perceive partiality in the language and cultural context of evaluations"(Khasawneh & Khasawneh, 2023) Since the examiners are native speakers, they may exhibit bias when assessing heritage students, such as assuming their Chinese proficiency is higher than others due to their identity. This situation requires examiners to adopt a professional attitude and treat students from different cultural backgrounds with impartiality.

Additionally, leveraging modern assessment tools like AI and big data can enhance the efficiency and reliability of the OPI process. AI can assist in designing adaptive question chains and analyzing speech patterns, while big data can track performance trends to refine methodologies. Collect examiner feedback regularly: "The nature and impact of raters' feedback in language assessment remains underexplored"(Yang, 2023). Gather feedback from examiners about challenges and uncertainties they face during the scoring process to adjust training and guidance in a timely manner.

Finally, optimizing testing processes and scoring standards. "Rater reliability is the extent to which two or more raters agree on each other's scoring of the same candidate" (Karuppaiah et al., 2020). In addition to traditional interview evaluations, incorporate methods such as audio recordings and peer reviews to provide a more comprehensive assessment perspective and reduce biases associated with a single evaluation method.

CONCLUSION

The results of the study are summarized as follows: (1) discrepancies in the students OPI level at admission and pretest, and an increase in OPI level at post-test (2) uneven

progress across proficiency levels and difficulties in cultural and social integration.

Discrepancies between admission, pre-test, and post-test results reveal the significant impact of examiner subjectivity, environmental differences, and strategic manipulation by students. These inconsistencies underscore the need for standardized scoring practices and rigorous examiner training to enhance reliability and fairness in assessments.

Moreover, the challenges faced during the immersion program, such as uneven progress across proficiency levels and difficulties in cultural and social integration, emphasize the importance of tailoring instructional strategies to meet the specific needs of students. Lower-level learners require structured cultural engagement to foster language growth, while advanced learners need specialized academic language development to address their unique challenges.

The findings highlight the need for a comprehensive, multifaceted approach to address the identified challenges. By combining technological advancements, targeted training, and tailored instructional strategies, the OPI process can be transformed into a more equitable, efficient, and impactful system. These interventions will not only benefit students by providing accurate and meaningful evaluations but also support the CLS program in achieving its broader educational goals.

ACKNOWLEDGEMENTS

The success of this study was dependent on the unconditional support and understanding of the participants for their willingness to be part of the study and took time to complete the interview.

REFERENCES

Achirri, K. (2021). Life is splendid here in the U.S.: Intercultural learning in contemporary Chinese students' academic adjustment. *The Qualitative Report*. https://doi.org/10.46743/2160-3715%2F2021.3501

American Council on the Teaching of Foreign Languages [ACTFL]. (2018). 2017 Annual Report. www.actfl.org/sites/default/files/reports/annualreport2017/index.html

ACTFL. (2020). *ACTFL assessments*. https://www.actfl.org

ACTFL. (2021). *Oral Proficiency Interview (OPI).* https://www.languagetesting.com

Akbari, N. (2020). *Second language assessment in Persian*. https://doi.org/10.4324/9780429446221-25

Bijani, H. (2019). Evaluating the effectiveness of the training program on direct and semi-direct oral proficiency assessment: A case of multifaceted Rasch analysis. *Cogent Education, 6*(1). https://doi.org/10.1080/2331186X.2019.1670592

Bucher, T. (2019). The impact of the social dimension on assessing oral proficiency. In E. White & T. Delaney (Eds.), *Handbook of Research on Assessment Literacy and Teacher-Made Testing in the Language Classroom* (pp. 141-156). IGI Global. https://doi.org/10.4018/978-1-5225-6986-2.CH008

Candel-Mora, M. A. (2019). Intercultural communication competence in specialized languages and contexts: Research prospects and possibilities. *Advanced Linguistics, 4,* 4-9. https://doi.org/10.20535/2617-5339.2019.4.181340

Carr, S., & Sun, S. (2021). Ensuring accuracy and quality for oral examinations in translation. *Assessment & Evaluation in Higher Education, 47*, 830 - 842. https://doi.org/10.1080/02602938.2021.1972929

Ellis, R., Skehan, P., Li, S., Shintani, N., & Lambert, C. (2019). *Task-Based language teaching.* https://doi.org/10.1017/9781108643689

Gardner, J., O'Leary, M., & Yuan, L. (2021). Artificial intelligence in educational assessment: 'Breakthrough? or Buncombe and Ballyhoo?'. *J. Comput. Assist. Learn., 37*, 1207-1216. https://doi.org/10.1111/jcal.12577

Jackson, J., & Schwieter, J. W. (2019). Study abroad and immersion. In J. W. Schwieter & A. Benati (Eds.), *The Cambridge Handbook of Language Learning* (pp. 727 - 750). https://doi.org/10.1017/9781108333603.031

Karuppaiah, S., & Raof, A. H. A. (2020). The impact of rater training on rater reliability in an English Oral Test. *Asian Journal of Assessment in Teaching and Learning, 10*(2), 94-105. https://doi.org/10.37134/ajatel.vol10.2.10.2020

Kawaguchi, K., Sakoda, A., Kojima, K., & Goto, A. (2020). Current trends and prospects of studies on Japanese OPI in Korea. *The Japanese Language Association of Korea, 66*, 5-26. https://doi.org/10.14817/jlak.2020.66.5

Khasawneh, M., & Khasawneh, Y. (2023). *Achieving assessment equity and fairness: Identifying and eliminating bias in assessment tools and practices*. Preprints. https://doi.org/10.20944/preprints202306.0730.v1

Kinney, C. L., Raddatz, M. M., Robinson, L. R., Garrison, C. J., & Sabharwal, S. (2021). Interrater reliability in the American Board of Physical Medicine and Rehabilitation Part II Certification Examination. *American Journal of Physical Medicine & Rehabilitation, 101*, 468 - 472. https://doi.org/10.1097/PHM.0000000000001859

Ma, W., & Winke, P.M. (2019). Self-assessment: How reliable is it in assessing oral proficiency over time? *Foreign Language Annals, 52*(1), 66-86. https://doi.org/10.1111/FLAN.12379

McConnell, K.D. (2022). The devil is in the details: A response to Eubanks. *Journal of Assessment and Institutional Effectiveness*, *12*, 48 - 60. https://doi.org/10.5325/jasseinsteffe.12.1-2.0048

Ortega, P., Izquierdo, K., Iñigo, R., Gonzaliez, J., Gregorich, S., Karliner, L., Cordon, C., Diamon, L., & Figueroa, J. (2023). Development and effectiveness of a rater training curriculum for evaluating student medical Spanish Oral Proficiency using the physician oral language observation matrix. *Global Business Languages, 23*, 14–39. https://doi.org/10.4079/gbl.v23.3

Peng, Y., Yan, W., & Cheng, L. (2020). Hanyu Shuiping Kaoshi (HSK): A multi-level, multi-purpose proficiency test. *Language Testing, 38*, 326 - 337. https://doi.org/10.1177/0265532220957298

Rossi, O., & Brunfaut, T. (2020). Raters of subjectively-scored tests. *The TESOL Encyclopedia of English Language Teaching*, 1-7. https://doi.org/10.1002/9781118784235.eelt0985

Silva, C. D., Alexandre, B. H., Ferronato, R. F., Pontes, F. G., & Lima, O. D. (2024). Vygotsky E A Aprendizagem Sociointeracionista: O Papel Da Linguagem E Do Contexto Cultural Na Educação. *IOSR Journal of Business and Management, 26*(11), 8-18. https://doi.org/10.9790/487x-2611090818

Siraranghom, W. (2024). Developing a comprehensive preparation module for the Oral Proficiency Interview (OPI). *NKRAFA Journal of Humanities and Social Sciences, 12*, 117–129. https://so04.tci-thaijo.org/index.php/KANNICHA/article/view/274563

Wang, Z., & Zhang, J. (2023). Mediation and learner reciprocity. *Language and Sociocultural Theory, 10*(1), 82-105. https://doi.org/10.1558/lst.22181

Winke, P., &Brunfaut, T. (Eds.). (2020). *The Routledge handbook of second language acquisition and language testing* (1st ed.). https://doi.org/10.4324/9781351034784

Yang, H. (2023). Investigating stakeholders' needs to enhance raters' feedback on an English Oral Proficiency Interview Test. *Multimedia-Assisted Language Learning, 26*(4). https://doi.org/10.15702/mall.2023.26.4.98