

The Eurasia Proceedings of Educational & Social Sciences (EPESS), 2022

Volume 28, Pages 11-18

IconSE 2022: International Conference on Science and Education

Oral Proficiency Interview (OPI) for the Assessment of Speaking

Ahmet ONAL

Süleyman Demirel University

Abstract: The skill of speaking, arguably, plays the most essential role in the process of communication among people; thus, the teaching and assessment of speaking deserves special attention. The oral proficiency interview (OPI) has been utilized as a technique for the assessment of speaking for about 70 years and its traditional format has been revised and updated in line with the requirements of contexts and innovations in technology. Accordingly, this paper briefly introduces the OPI and discusses the advantages and disadvantages of the OPI. In addition, the reliability and validity of the OPI for the assessment of speaking skills has also been expounded on with specific reference to its strengths and weaknesses. More precisely, the paper argues that the OPI has the potential to assist foreign language teachers in teaching and testing the speaking skills of their learners as long as its weaknesses are compensated for, reliable rating scales are designed and raters are trained to score the performances reliably.

Keywords: Testing speaking, Assessment of speaking, Oral proficiency interview

Introduction

Knowing a language involves being able to listen, read (and understand), speak and write using that language and these four skills are usually integrated in daily life. For instance, when a person gets a text message, s/he writes a reply or when s/he watches the news on television, s/he may talk about the news item to the person that accompanies him/her. However, being able to speak the target language fluently is the main aim of most people learning foreign languages (Gültekin, 2022; Richards, 2015) and speaking is regarded as the primary skill when compared to listening, reading or writing (Boeru & Cizer, 2021; Namaziandost & Ahmadi, 2019; Suwarno, 2017) because people ask ‘*Do you speak English?*’ rather than ‘*Do you listen to/read in/write in English?*’ when they meet a foreigner.

Assessment of Speaking

Assessment of speaking skill is notorious for being time-consuming, resource-intensive and logistically-complex endeavor (Abi & Üstünel, 2018; Jankowska & Zielińska, 2015; Louma, 2004; Malone & Montee, 2010; Nakatsuhara et al., 2021; Namaziandost & Ahmadi, 2019; Önal, 2020) since it calls for the development and implementation of valid and reliable tests. Moreover, it has been argued that the best way of testing speaking is getting test-takers to speak (Fulcher & Davidson, 2007; Green, 2021) although it is also possible to assess speaking through indirect tasks (such as repetition and multiple-choice tests). However, as has been observed by Gültekin (2022), due to the challenges it presents, some teachers may tend not to teach and assess the skill of speaking and when they do teach and test speaking, they make use of traditional methods, which do not contribute much to the communicative competences of the learners. Consequently, foreign language learners complain that *they can understand but cannot speak* the target language.

Oral Proficiency Interview

- This is an Open Access article distributed under the terms of the Creative Commons Attribution-Noncommercial 4.0 Unported License, permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

- Selection and peer-review under responsibility of the Organizing Committee of the Conference

© 2022 Published by ISRES Publishing: www.isres.org

The term oral proficiency can be defined as “learners’ ability to converse with one or several interlocutors” (Sandlund et al., 2016, p. 16) and oral proficiency interviews (OPIs) have been increasingly and extensively used to assess learners’ oral proficiency levels since their introduction in 1952 (Abi & Üstünel, 2018; Brown & Abeywickrama, 2018; Chalhoub-Deville & Fulcher, 2003; Cho, 2004; Glisan et al., 2013; Green, 2021; Gültekin, 2022; Hırçın Çoban, 2017; Johnson, 2000; Kasper & Ross, 2007; Liskin-Gasparro, 2003; Louma, 2004; Malone, 2003; Malone & Montee, 2010; Nakatsuhara et al., 2021; Nakatsuhara et al., 2022; Richards, 2015; Staples et al., 2017; Swender, 2003; Thompson et al., 2006). OPI can be briefly defined as an interview between a trained tester and a test-taker “...designed to elicit a profile of the examinee’s oral proficiency through a sustained performance” (Malone, 2003, p. 491). The performance of the test-taker is then scored in line with the criteria in the rating scale employed (Malone & Montee, 2010). Thus, it would be safe to argue that the OPI is a criterion-referenced test in that it compares each test-taker’s performance to the assessment criteria rather than comparing the performances of the test-takers with each other (Swender, 2003).

In its traditional form, the OPI consists of four stages (Brown & Abeywickrama, 2018; Cho, 2004; Gültekin, 2022; Johnson, 2000; Malone & Montee, 2010):

- 1- *Warm-up*: The tester warms up the test-taker by getting acquainted with each other, introduces the test format and tries to relieve test-taker’s anxiety.
- 2- *Level check*: The tester tries to identify and confirm the proficiency level of the test-taker through a set of preplanned questions and tasks.
- 3- *Probe*: The tester aims to discover the upper limit (or *ceiling*) ability of the test-taker with the help of linguistically and cognitively challenging questions.
- 4- *Wind-down*: The tester ensures that the test-taker leaves the room with a feeling of achievement and satisfaction with the help of easier questions.

Basically, the OPI may be regarded as a kind of interaction between the tester and the test-taker. In this respect, when two or more interlocutors talk to each other, the process is referred to as interaction and involves such components as *turn management* (starting, maintaining and ending), *topic management* (initiating, extending, shifting and closing), *non-verbal behavior* (eye contact, facial expression and laughter), *breakdown repair* (joint utterance creation and recasts) and *interactive listening* (backchannelling, comprehension checks and continuers) (Galaczi & Taylor, 2021; Richards, 2015). As for the scoring of the OPI performances of the test-takers, global frameworks such as the American Council on the Teaching of Foreign Languages (ACTFL) Guidelines and the Common European Framework of Reference (CEFR) have been utilized to assess, identify, interpret and compare linguistic proficiency of individuals and OPI scores are linked to the ACTFL or CEFR (ACTFL, 2012; Glisan et al., 2013; Gültekin, 2022; Hırçın Çoban, 2017; Johnson, 2000; Isbell & Winke, 2019; Isbell et al., 2019; Kissau, 2014; Malone & Montee, 2010; Swender, 2003; Tschirner et al., 2012).

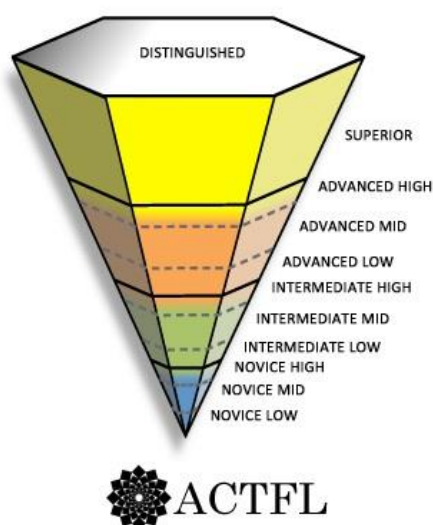


Figure 1. ACTFL proficiency guidelines–levels of proficiency (Source: <https://www.actfl.org/resources/actfl-proficiency-guidelines-2012>)

As can be seen in Figure 1, levels of proficiency in the ACTFL consist of 5 major levels and three of them (novice, intermediate and advanced) are further divided into three sub-levels (low, mid, and high) (ACTFL, 2012). As a result, the ACTFL features a total of 11-level structure (including the sub-levels) whereas the CEFR

consists of 6 main levels (i.e. A1, A2, B1, B2, C1, and C2). A comparison of ACTFL and CEFR levels has been presented in Table 1.

Table 1. Comparison of ACTFL and CEFR levels (Adapted from Tschirner et al., 2012; 12).

ACTFL	CEFR
Novice High	A1
Intermediate Low	A2
Intermediate Mid	B1(.1)
Intermediate High	B1(.2)
Advanced Low	B2(.1)
Advanced Mid	B2(.2)
Advanced High	C1
Superior	C2

Besides its traditional form, many different forms of OPI have been designed and implemented throughout the years. For example, *simulated oral proficiency interview* (SOPI) modeling the format of the OPI has been offered (Malone, 2000; Mikhailova, 2007; Sandlund et al., 2016) and supposed to feature certain practical, administrative and psychometric advantages over the traditional OPI (Stansfield, 1990). In a similar fashion, latest innovations in technology have made it possible to conduct OPIs through video-conferencing applications, which helps overcome geographical and temporal obstacles (Lim, 2018; Loranc-Paszyk, 2015; Marcum & Kim, 2020; Money Penny & Aldrich, 2016; Nakatsuhara et al., 2021; Rubio, 2015; Thompson et al., 2006; Tschirner et al., 2012). The computer-delivered version of the OPI, namely *Oral Proficiency Interview-computer* (OPIc), has also been gaining wider popularity (Isbell & Winke, 2019; Isbell et al., 2019; Marcum & Kim, 2020) since it may help alleviate challenges related to cost and time to be allocated for face-to-face implementation of the OPI (Isbell & Winke, 2019; Malone & Montee, 2010). High levels of correlation between the scores of the test-takers in the OPI and the OPIc has been reported (Isbell & Winke, 2019; Thompson et al., 2006; Tschirner et al., 2012) and OPI performances of the test-takers can be recorded and watched repeatedly to produce more accurate scoring and to give more precise feedback (Jankowska & Zielińska, 2015). Despite the benefits it offers, online implementation of the OPI and similar assessment methods raises issues related to test reliability, validity and security.

Reliability & Validity of the OPI

Several concerns have been voiced as to the reliability and construct validity of the OPI (Chalhoub-Deville & Fulcher, 2003; Cho, 2004; Hırçın Çoban, 2017; Johnson, 2000; Kissau, 2014; Malone, 2003; Malone & Montee, 2010; Salaberry, 2000). To start with, it has been argued that the OPI functions as a proficiency test aiming to assess test-takers' speaking skills in comparison to language spoken by well-educated native speakers; however, definition of *well-educated* and *native speaker* is elusive and controversial (Chalhoub-Deville & Fulcher, 2003; Kasper & Ross, 2007; Liskin-Gasparro, 2003; Malone & Montee, 2010). In addition, the type of interaction occurring in the OPI hardly represents the natural and authentic conversation between speakers, which casts its validity into shadow (Malone, 2003; Malone & Montee, 2010). More specifically, interactions in real-life tend to be more symmetrical and two-way and the power relationship between the interlocutors is often less evident and influential, which is not the case in the OPI.

In line with this, two types of reliability are regarded as relevant for the assessment of speaking; *intra-rater* and *inter-rater* reliability. The former refers to the extent to which an individual rater agrees with him-/herself for scoring the same performance on different occasions whereas the latter signifies the level of correlation between two different raters in scoring the same performance (Louma, 2004). The OPI has been frequently criticized in terms of the potential problems as to the inter-rater reliability (Cho, 2004; Kissau, 2014; Liskin-Gasparro, 2003; Malone, 2003). Louma (2004) and Surface & Dierdorff (2003) maintain that well-trained raters and high-quality scoring tools and procedures are of utmost importance to achieve a high level of reliability in the assessment of speaking skills.

Pros and Cons of the OPI

As has been aforementioned, the nature of interaction in a traditional oral interview is only one-sided (from the tester to the test-taker) or asymmetric (Brooks, 2009; Hırçın Çoban, 2017; Johnson, 2000) in that the tester strictly controls the course of the interview by asking questions that the test-taker is required to answer, which

rarely happens in an authentic conversation (Abi & Üstünel, 2018; Green, 2021; Johnson, 2000; Kasper & Ross, 2007; Liskin-Gasparro, 2003). Therefore, it can be argued that the tester plays a crucial role in the performance of the test-taker since the conversation is tightly controlled, delivered and co-constructed by the tester (Hırçın Çoban, 2017; Kasper & Ross, 2007). Put differently, *rater effects* that include consistency or inconsistency in a rater's elicitation techniques and scores, rater severity or leniency, illusory halo or horns and rater bias may lead to contamination of the assessment (Ahmadi & Sadeghi, 2016; Brooks, 2009; Brown & Abeywickrama, 2018; Chalhoub-Deville & Fulcher, 2003; Cho, 2004; Davis, 2022; Galaczi & Lim, 2022; Hırçın Çoban, 2017; Malone, 2003; O'Loughlin, 2002; Pill & Smart, 2021; Ross, 2012; Wigglesworth & Frost, 2017). As has been noted by Malone and Montee (2010), though OPIs are implemented by trained and certified testers, "...human error is inevitable and can be limited but not eliminated" (p. 978). In this regard, pairing or even grouping the test-takers and asking them to interact with each other may help remedy this weakness (Brooks, 2009; Brown & Abeywickrama, 2018; Green, 2021; Hırçın Çoban, 2017; Jankowska & Zielińska, 2015; Louma, 2004; Prasetyo, 2018; Sandlund et al., 2016) because, as has been reported by Ahmadi and Sadeghi (2016) and Brooks (2009), positive washback, more natural discourse and more various language functions can be attained and the workload of the rater in terms of time, effort and cost may be reduced. Moreover, when test-takers are paired or grouped, they tend to feel less anxious in comparison to the traditional OPI (Hırçın Çoban, 2017). Consequently, when a test-taker is paired with a friend, s/he tends to perform better (Brooks, 2009; O'Sullivan, 2002).

On the other hand, it should not go without saying that even when the interview is conducted in pairs or groups, the so-called *interlocutor effect* may emerge because each individual test-taker possesses his/her own socio-cultural identity and personal characteristics (Brooks, 2009; Brown, 2012; Nakatsuhara et al., 2022; Sandlund et al., 2016). Put differently, individuals tend to accommodate their speech in line with their interlocutors and variables such as test-takers' acquaintanceship, cultural and linguistic background, extroversion, talkativeness, gender and language proficiency may influence the course of interaction (Brooks, 2009; Galaczi & Taylor, 2021; O'Loughlin, 2002; O'Sullivan, 2002; O'Sullivan & Porter, 1996). As a result, an individual test-taker may tend to dominate the interaction and decrease other test-takers' chances of showing their real performances. In such cases, raters should closely monitor the course of interaction and present equal opportunities to the test-takers. Though paired/grouped OPIs present certain challenges, Brooks (2009) has concluded that test-takers exhibited better performance in paired format of the OPI, which resulted in more interaction, negotiation of meaning, complex output and consideration of the interlocutor.

It has been advocated that OPIs are good measures of test-takers' overall communicative competence and have some other side-benefits such as alignment of curricula, methods and classroom activities with a more focus on communication (Chalhoub-Deville & Fulcher, 2003; Kissau, 2014; Malone & Montee, 2010; Suwarno, 2017), implying that OPIs tend to yield positive washback. In this respect, Gültekin (2022) has reported that tertiary level students studying at English preparatory classes prefer OPI to traditional speaking tests since they get a chance to demonstrate their communicative skills in a realistic setting. Moreover, despite concerns over the reliability of the OPI, the reliability of OPI and its adaptability to self-assessment have been confirmed by Ma and Winke (2019). Similarly, Surface and Dierdorff (2003) demonstrated inter-rater consistency and agreement in the OPI for 19 different languages.

It would be safe to argue that assessment of speaking is already an anxiety-provoking experience for many test-takers (Marzec-Stawiarska, 2015; Sayın, 2015; Sullivan, 2011; Suwarno, 2017) and when it comes to the OPI (viewed by some as talking in a foreign language to a foreigner in a test format), the level of anxiety multiplies. It should be noted at this point that repeated practice may help individuals to overcome anxiety-provoking experiences and build their self-confidence, which implies that OPIs should also be utilized for instructional purposes in addition to assessment purposes with the aim of helping test-takers alleviate their anxiety.

Use of Rating Scales in the OPI

In order to score any oral and/or written performance, a well-designed rating scale (or *rubric*) is essential to ensure reliability (Brown & Abeywickrama, 2018; Fulcher, 2012; Galaczi & Lim, 2022; Green, 2021; Namaziandost & Ahmadi, 2019; Wigglesworth & Frost, 2017). Either *holistic* or *analytic* scales are commonly used for scoring the speaking performance of the test-takers. In a holistic scale, a single number is awarded to a specific performance (Brown & Abeywickrama, 2018; Galaczi & Lim, 2022; Gültekin, 2022; Kuiken & Vedder, 2021; Louma, 2004) and that number is meant to signify the overall speaking ability of the test-taker (Fulcher & Davidson, 2007). It should be noted that though holistic scales are user-friendly, practical and fast to use, they do not provide the test-takers with any kind of feedback (Hamp-Lyons, 1991; Kuiken & Vedder, 2021). In contrast, in an analytic scale, multiple scores across a range of categories or criteria (such as fluency, vocabulary

use, pronunciation, interaction, content, etc.) are given to a single spoken performance and these scores may be added together to obtain an overall score (Brown & Abeywickrama, 2018; Galaczi & Lim, 2022; Green, 2021; Louma, 2004; Richards, 2015). Therefore, test-takers have the chance to get specific diagnostic feedback as to their strengths and weaknesses (Kuiken & Vedder, 2021). The main advantage of analytic scales is their flexibility of use in that differential weighting can be applied across the categories in line with the population and specific purpose of the assessment (Green, 2021). As can be surmised, both types of scales have their own strengths and weaknesses; thus, Namaziandost and Ahmadi (2019) suggest that making use of both types of scales may be more suitable since they can supplement each other and produce more reliable scores.

Rater Training for the OPI

A well-designed rating scale may be regarded as the first step for a reliable assessment of speaking; however, the same scale may be interpreted differently by different raters since it is the rater that mediates between the rating scale and the performance of the specific test-taker (Davis, 2022; Galaczi & Lim, 2022; Pill & Smart, 2021). Therefore, the necessity of training the interlocutors and raters to attain high level of standardization (of the experiences of all the test-takers as much as possible) has been frequently underscored in the relevant literature (Brown, 2012; Davis, 2022; Fulcher & Davidson, 2007; Green, 2021; Hamp-Lyons, 1991; Hırçın Çoban, 2017; Kuiken & Vedder, 2021; Louma, 2004; Malone, 2000; Malone & Montee, 2010; Pill & Smart, 2021; Ross, 2012; Surface & Dierdorff, 2003; Wigglesworth & Frost, 2017). The focus of these trainings may be on the interpretation and implementation of the rating scale to be employed and, as has been underscored by Gablasova (2021), the threat of rater effect should also be covered in rater training sessions with the aim of minimizing it. Such training programs can even be conducted online thanks to web-conferencing technologies so that they become more accessible for those who are disadvantaged in terms of location and time (Brown, 2012).

Conclusion and Recommendations

This paper has underscored the need to assess the skill of speaking by referring to the fundamental role it plays in the process of communication. In other words, if we really aim at improving our learners' communicative competences, we must teach and test their oral skills effectively. One way of achieving this aim is the OPI, which can be used both for instructional and assessment purposes. The paper has briefly introduced the traditional and updated formats of the OPI with reference to the benefits it offers and drawbacks it presents. Use of rating scales and training of the raters for the OPI have also been dealt within the paper and it has been suggested that no one best scale has the potential to satisfy all goals and contexts; thus, such factors as the purpose of the assessment, the construct to be assessed, the characteristics of the population and the interpretation of the test scores should be taken into account when choosing the specific rating scale to be employed (Kuiken & Vedder, 2021).

To conclude, suitability for purpose rather than best way of testing speaking needs to be sought for and the potential of technology in assisting us for the assessment of speaking skills should be taken into consideration (Nakatsuhara et al., 2022). As an example, web-conferencing technologies, which have become regular components of our daily lives since the outbreak of the Covid-19 pandemic, may be employed in the implementation of OPI. Similarly, with the help of technologies such as eye-tracking and natural language processing, it may become viable to better understand interlocutor/rater cognition and decision-making processes (Davis, 2022). In a similar fashion, automated scoring systems promise to support and even substitute raters by scoring oral performances faster and more reliably (Brown, 2012; Galaczi & Lim, 2022; Lim, 2018; Ross, 2012). As a final note, it should not be forgotten that the successful implementation of OPIs requires financial and personnel resources as well as time and meticulous planning (Isbell & Winke, 2019; Malone & Montee, 2010; Thompson et al., 2006) and when such requirements cannot be met, it would not be feasible to use OPIs for the assessment of speaking.

Scientific Ethics Declaration

The author declares that the scientific ethical and legal responsibility of this article published in EPESS journal belongs to the author.

Acknowledgements

* This article was presented as an oral presentation at the International Conference on Science and Education (www.iconse.net) held in Antalya/Turkey on November 15-18, 2022

References

- Abi, M. & Üstünel, E. (2018). Multiple questions in different sections of oral proficiency interviews. *ELT Research Journal*, 7(2), 91-103.
- ACTFL (American Council on the Teaching of Foreign Languages). (2012). *ACTFL proficiency guidelines 2012* [Electronic version]. Retrieved from <https://www.actfl.org/sites/default/files/guidelines/ACTFLProficiencyGuidelines2012.pdf>
- Ahmadi, A. & Sadeghi, E. (2016). Assessing English language learners' oral performance: A comparison of monologue, interview, and group oral test. *Language Assessment Quarterly*, 13(4), 341-358. <https://doi.org/10.1080/15434303.2016.1236797>
- Boeru, M. & Cizer, L. (2021). Tell, say, talk—speaking strategies and techniques to growing confident communicators for oral proficiency interviews in English. In Boldea, I., Sigmirean, C. & Buda, D.-M., *Reading Multiculturalism: Human and Social Perspectives*, 111-117. Arhipelag XXI Press.
- Brooks, L. (2009). Interacting in pairs in a test of oral proficiency: Co-constructing a better performance. *Language Testing*, 26(3), 341-366. <https://doi.org/10.1177/0265532209104666>
- Brown, H. D. & Abeywickrama, P. (2018). *Language assessment: Principles and classroom practices (3rd Ed.)*. Pearson Education.
- Chalhoub-Deville, M. & Fulcher, G. (2003). The oral proficiency interview: A research agenda. *Foreign Language Annals*, 36(4), 498-506.
- Cho, S. (2004). Oral proficiency interview: Pros and cons. *The Korean Language in America*, 9, 144-151.
- Davis, L. (2022). Rater and interlocutor training. In Fulcher, G. & Harding, L. (Eds.), *The Routledge handbook of language testing (2nd Ed.)*, 322-338, Routledge.
- Fulcher, G. (2012). Scoring performance tests. In Fulcher, G. & Davidson, F. (Eds.), *The Routledge handbook of language testing*, 378-392, Routledge.
- Fulcher, G. & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. Routledge.
- Gablasova, D. (2021). Corpora for second language assessments. In Winke, P. & Brunfaut, T. (Eds.), *The Routledge handbook of second language acquisition and language testing*, 45-53, Routledge.
- Galaczi, E. D. & Lim, G. S. (2022). Scoring performance tests. In Fulcher, G. & Harding, L. (Eds.), *The Routledge handbook of language testing (2nd Ed.)*, 495-510, Routledge.
- Galaczi, E. D. & Taylor, L.B. (2021). Measuring interactional competence. In Winke, P. & Brunfaut, T. (Eds.), *The Routledge handbook of second language acquisition and language testing*, 338-348, Routledge.
- Glisan, E. W., Swender, E. & Surface, E. A. (2013). Oral proficiency standards and foreign language teacher candidates: Current findings and future research directions. *Foreign Language Annals*, 46(2), 264-289. <https://doi.org/10.1111/flan.12030>
- Green, A. (2021). *Exploring language assessment and testing (2nd Ed.)*. Routledge.
- Gültekin, E. (2022). *The impact of American council on the teaching of foreign languages (ACTFL) oral proficiency interview on assessing speaking skill in English preparatory schools at higher education in Turkey*. Master's thesis, Maltepe University, Institute of Postgraduate Studies.
- Hamp-Lyons, L. (1991). Scoring procedures for ESL contexts. In Hamp-Lyons, L. (Ed.). *Assessing second language writing in academic contexts*, 241-276, Ablex.
- Hırçın Çoban, M. (2017). *Resolving interactional troubles in paired oral proficiency assessment in an English as a foreign language context*. Master's thesis, Hacettepe University, Graduate School of Educational Sciences.
- Isbell, D. & Winke, P. (2019). ACTFL Oral proficiency interview—computer (OPIc). *Language Testing*, 36(3), 467-477. <https://doi.org/10.1177/0265532219828253>
- Isbell, D. R., Winke, P. & Gass, S. M. (2019). Using the ACTFL OPIc to assess proficiency and monitor progress in a tertiary foreign languages program. *Language Testing*, 36(3), 439-465. <https://doi.org/10.1177/0265532218798139>
- Jankowska, A. & Zielińska, U. (2015). Designing a self-assessment instrument for developing the speaking skill at the advanced level. In Pawlak, M. & Waniek-Klimczak, E. (Eds.), *Issues in teaching, learning and testing speaking in a second language*, 251-265, Springer.
- Johnson, M. (2000). Interaction in the oral proficiency interview: Problems of validity. *Pragmatics*, 10(2), 215-231.

- Kasper, G. & Ross, S. J. (2007). Multiple questions in oral proficiency interviews. *Journal of Pragmatics*, 39(11), 2045-2070.
- Kissau, S. (2014). The impact of the oral proficiency interview on one foreign language teacher education program. *Foreign Language Annals*, 47(3), 527-545. <https://doi.org/10.1111/flan.12092>
- Kuiken, F. & Vedder, I. (2021). Scoring approaches: Scales/rubrics. In Winke, P. & Brunfaut, T. (Eds.), *The Routledge handbook of second language acquisition and language testing*, 125-134, Routledge.
- Lim, G. S. (2018). Conceptualizing and operationalizing second language speaking assessment: Updating the construct for a new century, *Language Assessment Quarterly*, 15(3), 215-218, <https://doi.org/10.1080/15434303.2018.1482493>
- Liskin-Gasparro, J. E. (2003). The ACTFL proficiency guidelines and the oral proficiency interview: A brief history and analysis of their survival. *Foreign Language Annals*, 36(4), 483-490.
- Loranc-Paszylk, B. (2015). Videoconferencing as a tool for developing speaking skills. In Pawlak, M. & Waniek-Klimczak, E. (Eds.), *Issues in teaching, learning and testing speaking in a second language*, 189-203, Springer.
- Louma, S. (2004). *Assessing speaking*. Cambridge University Press.
- Ma, W. & Winke, P. (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>
- Malone, M. (2000). Simulated oral proficiency interviews: Recent developments. *ERIC Digest*, 1-7.
- Malone, M. E. (2003). Research on the oral proficiency interview: Analysis, synthesis, and future directions. *Foreign Language Annals*, 36(4), 491-497.
- Malone, M. E. & Montee, M. J. (2010). Oral proficiency assessment: Current approaches and applications for post-secondary foreign language programs. *Language and Linguistics Compass*, 4(10), 972-986. <https://doi.org/10.1111/j.1749-818x.2010.00246.x>
- Marcum, J. & Kim, Y. (2020). Oral language proficiency in distance English-language learning. *Computer Assisted Language Instruction Consortium (CALICO Journal)*, 37(2), 148-168. <https://doi.org/10.1558/cj.37788>
- Marzec-Stawiarska, M. (2015). Investigating foreign language speaking anxiety among advanced learners of English. In Pawlak, M. & Waniek-Klimczak, E. (Eds.), *Issues in teaching, learning and testing speaking in a second language*, 103-119, Springer.
- Mikhailova, J. (2007). Rethinking description in the Russian SOPI: Shortcomings of the simulated oral proficiency interview. *Foreign Language Annals*, 40(4), 584-603.
- Money Penny, D. B. & Aldrich, R. S. (2016). Online and face-to-face language learning: A comparative analysis of oral proficiency in introductory Spanish. *Journal of Educators Online*, 13(2), 105-133.
- Nakatsuhara, F., Inoue, C. & Khabbazzashi, N. (2021). Measuring L2 speaking. In Winke, P. & Brunfaut, T. (Eds.), *The Routledge handbook of second language acquisition and language testing*, 285-294, Routledge.
- Nakatsuhara, F., Khabbazzashi, N. & Inoue, C. (2022). Assessing speaking. In Fulcher, G. & Harding, L. (Eds.), *The Routledge handbook of language testing* (2nd Ed.), 209-222, Routledge.
- Namaziandost, E. & Ahmadi, S. (2019). The assessment of oral proficiency through holistic and analytic techniques of scoring: A comparative study. *Applied Linguistics Research Journal*, 3(2), 70-82. <https://doi.org/10.14744/alrj.2019.83792>
- O'Loughlin, K. (2002). The impact of gender in oral proficiency testing. *Language Testing*, 19(2), 169-192.
- O'Sullivan, B. (2002). Learner acquaintanceship and oral proficiency test pair-task performance. *Language Testing*, 19(3), 277-295.
- O'Sullivan, B. & Porter, D. (1996). Speech style, gender, and oral proficiency interview performance. *Paper presented at the Annual Meeting of the Southeast Asian Ministers of Education Organization Regional Language Center Seminar*, Singapore.
- Önalın, O. (2020). Automated scoring of productive skills in language assessment. In Çelik, S. (Ed.), *Preparing teachers for a changing world: Contemporary issues in EFL education*, 103-114, Vizetek.
- Pill, J. & Smart, C. (2021). Raters: Behavior and training. In Winke, P. & Brunfaut, T. (Eds.), *The Routledge handbook of second language acquisition and language testing*, 135-144, Routledge.
- Prasetyo, H. (2018). Paired oral tests: a literature review. *LLT Journal: A Journal on Language and Language Teaching*, 21(Suppl), 105-110. <https://doi.org/10.24071/lt.2018.Suppl2110>
- Richards, J. C. (2015). *Key issues in language teaching*. Cambridge University Press.
- Ross, S. J. (2012). Claims, evidence, and inference in performance assessment. In Fulcher, G. & Davidson, F. (Eds.), *The Routledge handbook of language testing*, 223-233, Routledge.
- Rubio, F. (2015). Assessment of oral proficiency in online language courses: Beyond reinventing the wheel. *The Modern Language Journal*, 99(2), 405-408.
- Salaberry, R. (2000). Revising the revised format of the ACTFL Oral Proficiency Interview. *Language Testing*, 17(3), 289-310.

- Sandlund, E., Sundqvist, P. & Nyroos, L. (2016). Testing L2 talk: A review of empirical studies on second-language oral proficiency testing. *Language and Linguistics Compass*, 10(1), 14-29. <https://doi.org/10.1111/lnc3.12174>
- Sayın, B. A. (2015). Exploring anxiety in speaking exams and how it affects students' performance. *International Journal of Education and Social Science*, 2(12), 112-118.
- Stansfield, C. W. (1990). A comparative analysis of simulated and direct oral proficiency interviews. *Paper presented at the Annual Meeting of the Regional Language Centre Conference*, Singapore.
- Staples, S., Laflair, G. T. & Egbert, J. (2017). Comparing language use in oral proficiency interviews to target domains: Conversational, academic, and professional discourse. *The Modern Language Journal*, 101(1), 194-213. <https://doi.org/10.1111/modl.12385>
- Sullivan, J. H. (2011). Taking charge: Teacher candidates' preparation for the oral proficiency interview. *Foreign Language Annals*, 44(2), 241-257.
- Surface, E. A. & Dierdorff, E. C. (2003). Reliability and the ACTFL Oral Proficiency Interview: Reporting indices of interrater consistency and agreement for 19 languages. *Foreign Language Annals*, 36(4), 507-519.
- Suwarno, P. (2017). The importance of oral performance: what makes a successful proficiency-based foreign language program. *ISLLAC: Journal of Intensive Studies on Language, Literature, Art, and Culture*, 1(1), 1-18.
- Swender, E. (2003). Oral proficiency testing in the real world: Answers to frequently asked questions. *Foreign Language Annals*, 36(4), 520-526.
- Thompson, G. L., Cox, T. L. & Knapp, N. (2016). Comparing the OPI and the OPIc: The effect of test method on oral proficiency scores and student preference. *Foreign Language Annals*, 49(1), 75-92. <https://doi.org/10.1111/flan.12178>
- Tschirner, E., Bärenfänger, O. & Wanner, I. (2012). *Assessing evidence of validity of assigning CEFR ratings to the ACTFL oral proficiency interview (OPI) and the oral proficiency interview by computer (OPIc)*. Technical Report 2012-US-PUB-1. Retrieved from <http://www.languagetesting.com/wpcontent/uploads/2014/02/OPIc-CEFR-Study-Final-Report.pdf>
- Wigglesworth, G. & Frost, K. (2017). Task and performance-based assessment. In Shohamy, E., Or, I. G. & May, S. (Eds.), *Language testing and assessment (3rd Ed.)*, 121-134, Springer International Publishing.

Author Information

Ahmet Önal

Suleyman Demirel University

Isparta, TURKEY

Contact e-mail: ahmetonal@sdu.edu.tr

To cite this article:

Önal, A. (2022). Oral proficiency interview (OPI) for the assessment of speaking. *The Eurasia Proceedings of Educational & Social Sciences (EPESS)*, 28, 11-18.