ISSN 2348-3156 (Print)
International Journal of Social Science and Humanities Research ISSN 2348-3164 (online)
Vol. 14, Issue 1, pp: (337-341), Month: January - March 2026, Available at: www.researchpublish.com

Deep Insight into Interview Format vs. Paired Format in Speaking Tests

Naser Alsubaiei

DOI: https://doi.org/10.5281/zenodo.19178650
Published Date: 23-March-2026

Abstract: **This paper discusses two common formats used in speaking assessment: the interview format and the paired format. Speaking is a crucial skill in language learning, and its evaluation requires reliable and valid testing methods. The interview format is practical and easy to administer, but it may not fully represent real communication. The paired format allows more natural interaction between candidates, although it is more difficult to score and manage. The paper also examines the challenges of scoring speaking tests, including issues of reliability, validity, and examiner judgment. It concludes that combining different test formats and using clear scoring criteria can improve the accuracy of speaking assessment.**

Keywords: **Speaking assessment; Interview format; Paired format; Oral proficiency interview (OPI); Communicative competence; Language testing; Reliability; Validity; Rating scales; Speaking performance evaluation.**

1. INTRODUCTION

Speaking is widely considered one of the most essential skills in language learning. A learner cannot be regarded as proficient in a foreign language without demonstrating effective communicative ability. Due to the importance of speaking, it requires careful and systematic assessment. For this reason, many universities require international students to pass a speaking examination before being admitted. Consequently, the design and evaluation of speaking tests have become increasingly important in modern language education. However, assessing speaking ability is a complex task for both the examiner and the candidate, as numerous variables may influence the accuracy of the results.

To gain a deeper understanding of speaking assessment, this paper examines two common formats of speaking tests. First, the interview format will be discussed, including its advantages and limitations. Second, the paired format speaking test will be analyzed, with attention to its strengths and weaknesses. Finally, the paper will address the challenges involved in scoring speaking tests and explore methods for improving their reliability and validity.

2. ADVANTAGES AND DISADVANTAGES OF INTERVIEW FORMAT AND PAIRED FORMAT IN SPEAKING TESTS

2.1 The Interview Format

In the interview format of speaking assessment, often referred to as the Oral Proficiency Interview (OPI), several advantages can be identified. This type of test is relatively simple to administer, time-efficient, and appears to provide a clear indication of the candidate's ability to communicate in English. Unlike written examinations, it directly evaluates spoken performance, which is fundamentally different from reading and writing skills. In addition, the interview format allows examiners to assess multiple candidates within a limited period, making it practical in academic and institutional settings.

Despite these advantages, many scholars have criticized the interview format, particularly regarding its validity and the accuracy of the conclusions drawn from test scores. Messick (1998) argues that validity refers to the extent to which theoretical and empirical evidence supports the interpretation of test results. From this perspective, the OPI may be considered limited because it does not fully represent real-life communication.

One major limitation is that real communication occurs in different forms, including monologue, dialogue, and group interaction. However, the interview format generally focuses on dialogue between the examiner and the candidate, which does not reflect the full range of speaking situations encountered in everyday life. Brown (2003) and Bonk (2003) note that some learners perform better in discussion-based activities, while others perform better in structured dialogue. Therefore, a test that evaluates only one type of interaction cannot provide a comprehensive measure of speaking ability.

Another concern is the influence of the interviewer on the candidate's performance. As Brown (2003) explains, interviewers differ in speech style, tone, and interaction patterns, which may either facilitate or hinder the candidate's responses. As a result, the final score may reflect not only the candidate's ability but also the interaction between the examiner and the test-taker. This situation may reduce the objectivity of the test.

To minimize this problem, careful training of interviewers is required, along with regular monitoring and evaluation of their performance. For example, the same candidate may be tested by different examiners, and the results can be compared to ensure consistency. If significant differences appear, additional training may be necessary. McNamara (1997, 2002) also suggests that highly skilled interviewers may unintentionally improve the candidate's performance, which can affect the fairness of the test.

Another difficulty is that candidates often perform better when they take the test for the second time. Aarts, Flor, and Schils (1995) found that familiarity with the testing procedure can significantly influence performance. Therefore, test results may reflect experience with the test rather than actual language ability.

Hughes (1989) also points out that the interview situation may create a power relationship in which the candidate speaks to a perceived authority figure. In such cases, the candidate may hesitate to take initiative, which prevents the test from measuring natural communication skills. Consequently, the score may not accurately represent the candidate's true ability. Furthermore, Underhill (1987) argues that the interview format may not be suitable for highly proficient speakers, as it may fail to challenge them sufficiently. This limitation restricts the effectiveness of the test for advanced learners.

2.2 The Paired Format

The paired format of speaking assessment, as its name suggests, involves testing two candidates simultaneously in the presence of an interviewer. This format has become increasingly popular in modern language testing due to developments in communicative language teaching.

According to Taylor (2004), the introduction of paired and group speaking tests reflects changes that occurred during the 1980s in the teaching of English as a Foreign Language. Research in applied linguistics during the 1970s led to a greater understanding of the communicative function of language. As a result, language teaching shifted from focusing on knowledge about language to emphasizing the ability to use language effectively in real communication. This change influenced the design of speaking tests, encouraging the use of interactive formats such as paired tasks.

One of the main advantages of the paired format is that it allows examiners to observe a more natural and complex form of communication compared with the traditional interview format. Taylor (2000) suggests that interaction between two candidates creates a communicative environment that more closely resembles real-life situations. In this format, communication does not occur only between the examiner and the candidate, but also between the candidates themselves, which may provide a more accurate representation of speaking ability.

In many paired speaking tests, two examiners are involved. One examiner, often called the interlocutor, participates directly in the interaction, while the second examiner observes the performance without taking part in the conversation. This arrangement allows the test to be evaluated from two different perspectives. The interlocutor can assess how effectively the candidates communicate during the interaction, while the observer can provide a more objective judgment of their language performance.

From this perspective, the paired format may offer a more reliable evaluation of communicative competence because it measures the ability to participate in real conversation rather than simply answering questions. Since language is primarily used for interaction, testing candidates in a communicative situation may provide more valid results.

However, despite these advantages, the paired format also has several limitations. Fulcher (2003) notes that this type of test is more complex to administer and score than the traditional interview format. The presence of two examiners may lead to differences in judgment, and disagreements between the interlocutor and the assessor may make scoring more difficult. In addition, the behavior of one candidate may influence the performance of the other, which can reduce the fairness of the test.

Another potential problem is the influence of the interlocutor on the conversation. If the interlocutor provides too much guidance, the candidate's performance may be artificially improved. On the other hand, if the interlocutor provides too little support, the candidate may not have enough opportunity to demonstrate his or her ability. Therefore, the effectiveness of the paired format depends greatly on the training and consistency of the examiners.

3. THE SCORING OF SPEAKING TESTS

The challenges associated with scoring speaking tests are closely related to the nature of speaking assessment itself. Unlike many other types of examinations, speaking tests are primarily qualitative rather than quantitative. As a result, they rely heavily on human judgment, and the criteria used for scoring are often subjective. This subjectivity makes it difficult to establish completely objective measurement standards.

In situations where a single examiner evaluates a candidate, the examiner becomes part of the testing process, which may influence the final score. Communication is a highly complex activity that can be affected by many factors unrelated to language ability, such as personality, appearance, confidence, or the relationship between the examiner and the candidate. These factors may unintentionally influence the examiner's judgment and reduce the accuracy of the assessment.

Another important issue concerns the question of what exactly should be measured in a speaking test. Earlier testing approaches focused mainly on linguistic knowledge, such as vocabulary size, grammatical accuracy, and the ability to produce correct sentence structures. Candidates were often placed on a scale according to these measurable elements. However, more recent approaches emphasize communicative competence, which refers to the ability to use language effectively in real-life situations.

Both approaches have advantages and disadvantages. Traditional tests that focus on grammar and vocabulary may be easier to score objectively, but they do not always reflect how language is actually used in everyday communication. On the other hand, communicative tests attempt to measure real interaction, but they provide fewer objective data, making scoring more difficult. Underhill (1987) explains that communicative speaking tests are often expensive to administer because they require authentic materials, trained examiners, and carefully designed scoring procedures.

Another difficulty in scoring speaking tests relates to the design of rating scales. According to Underhill, rating scales are usually based on the description of a typical learner, which makes it difficult to evaluate candidates with very different levels of ability. In addition, rating scales that contain too many details may confuse the assessor, while scales that are too general may not provide enough guidance. If the descriptions of different levels overlap or contradict each other, the examiner may find it difficult to decide which level best represents the candidate's performance.

For example, rating scales for spoken English often include several levels that describe different degrees of vocabulary knowledge, grammatical accuracy, and communicative ability.

However, a candidate may show characteristics from more than one level, which makes the scoring process uncertain. In such cases, the reliability of the test may be reduced.

Upsher and Turner (1995) explain that the development of communicative language testing has expanded the criteria used in language assessment. Instead of answering simple questions, students are now expected to perform tasks that require extended speaking and interaction. As a result, rating scales must evaluate not only grammatical accuracy but also appropriateness, effectiveness, and fluency in communication. This development requires examiners to make more complex qualitative judgments.

They also suggest that one useful method for improving scoring procedures is to involve several teachers or examiners in the evaluation process. By comparing their scores and discussing the differences, it becomes possible to develop clearer and more consistent scoring guidelines. However, agreement among examiners does not necessarily mean that the test itself is accurate. If the test is not based on clear and well-designed criteria, the results may still lack validity even when the scorers agree.

3.1 Enhancing the Reliability and Validity of Speaking Test Scores

Before discussing ways to improve speaking test assessment, it is necessary to clarify the meanings of validity and reliability.

Validity refers to the extent to which a test measures what it is intended to measure. Henning (1987), cited in Alderson, Clapham, and Wall (1996), defines validity as the appropriateness of a test or any of its components as a measure of the

intended ability. A test can only be considered valid if the evidence supports the interpretation of its results for a specific purpose.

Reliability, on the other hand, refers to the consistency of test results. Underhill (1987) defines reliability as the degree to which a test produces the same results when the same learners are tested on different occasions. If scores vary significantly without a clear reason, the test cannot be considered reliable.

Test designers can improve validity and reliability in several ways. One important step is the careful construction of rating scales. Underhill (1987) suggests that rating scales should be improved through continuous revision. If a description is unclear, it should be modified, and if it cannot be clarified, it should be removed. Hughes (1989) states that scoring becomes more reliable when the descriptors are clear, well defined, and familiar to the examiners, and when the examiners are properly trained to use them.

Hughes (1989) also proposes several methods for increasing reliability:

Using multiple independent scorers

Speaking performances should be evaluated by at least two examiners working independently. Their scores should then be compared by a third examiner who investigates any major differences.

Providing clear and explicit instructions

Test instructions must be clear and unambiguous so that candidates understand exactly what is required. Misunderstanding the task may affect performance and reduce reliability.

Training examiners carefully

Examiners should not be allowed to score speaking tests without proper training. Their scoring patterns should be monitored, and examiners who show inconsistent judgments should receive additional training or should not be used again.

By applying these procedures, speaking tests can become more consistent and more accurate, although complete objectivity may never be fully achieved.

4. CONCLUSION

This paper has examined the advantages and disadvantages of two common formats used in speaking assessment: the interview format and the paired format. It has also discussed the main difficulties involved in scoring speaking tests and explored several methods for improving the validity and reliability of such assessments.

In modern educational and professional contexts, the evaluation of speaking ability has become increasingly important. Globalization has increased the need for effective communication in English, as many individuals study, work, and conduct business in international environments. For this reason, speaking tests must be designed carefully to ensure fairness, accuracy, and consistency.

A reliable speaking test should combine objective criteria with the evaluation of real communicative performance. While traditional interview tests are practical and easy to administer, they may not always reflect natural language use. On the other hand, paired or interactive formats provide a more realistic communicative situation, but they are more complex to administer and score. Therefore, neither format alone can provide a complete measure of speaking ability.

In addition, scoring speaking tests presents particular challenges because it depends on human judgment. Factors such as examiner differences, rating scale design, and test conditions may influence the results. For this reason, improving reliability and validity requires clear scoring criteria, careful examiner training, and the use of multiple evaluators whenever possible.

In conclusion, speaking assessment should not rely on a single test or a single format. A combination of different testing methods, including interviews, paired tasks, and communicative activities, is more likely to provide an accurate and meaningful evaluation of a learner's speaking proficiency. Although the assessment of spoken language is complex and sometimes subjective, it remains an essential part of language testing and must continue to be developed through research and careful practice.

REFERENCES

[1] Aarts, Flor, & Schils, Erik. 1995. Relative Clauses, the Accessibility Hierarchy, and the Contrastive Analysis Hypothesis. *IRAL*, 33, 1, 47-63.

[2] Bonk, W. J. & Ockey, G. J. 2003. A Many-Facet Rasch Analysis of the Second Language Group Oral Discussion Task. *Language Testing* 20.1.89-110

[3] Brown, A.2003."Interviewer Variation and the Co-construction of Speaking Proficiency." *Language Testing*, 20,1, 1-25.

[4] Fulcher, G. 1997. *Testing Second Language Speaking*. London: Longman

[5] Hughes, A. 1989. *Testing for Language Teachers*, Cambridge University Press, Cambridge.

[6] McNamara, T. 1997. "'Interaction' in second language performance assessment: whose performance?" *Applied Linguistics*, 22, 221-242.

[7] McNamara, T., Hill, K., and May, L.2002. "Discourse and Assessment" *Annual Review of Applied Linguistics*., 22, 221-242.

[8] Messick, SJ. 1998. *Assessment in Higher Education: Issues of Access, Quality, Student Development and Public Policy*. Erlbaum, New York:

[9] Taylor, L. 2000. "Investigating the Paired Speaking Test Format". UCLES *Research Notes* 2,14-15.

[10] Available on line at: http:// Cambridge-efl.org.rsnotes/0002/rn2.pdf

[11] Alderson, C., Clapham, C, and Wall, D. 1996. *Language Test Construction and evaluation.* Cambridge: CUP.

[12] Upsher,J. Turner, C. (2005). "Constructing rating scales for second language tests." *ELT Journal*. 49, 1, 3-12