



3 Crucial Elements to Consider for Speaking Assessments

by [Chris Huang](#)

Speaking is considered by many language educators to be one of the most difficult skills to assess. When designing speaking assessments, an educator needs to consider some important rudiments. First, the assessment of speaking requires either a real-time or recorded oral performance from the student. Second, the educator needs to develop the grading method and criteria. (I.e., not only does an assessment need to be designed, but the method for assessment needs to be created along with an appropriate grading rubric.) Third, the evaluator will need to provide a high-quality evaluation. In this article, we analyze these crucial elements and examine how they can impact the quality of speaking assessments.

1. Speaking Assessment Methods

Direct Methods

This type of speaking assessment is commonly defined as face-to-face interaction with at least one human interlocutor (Qian, 2009). The most prominent direct assessment method of speaking is an interview in which learners engage in a structured or semi-structured interaction with an interviewer or an interlocutor. Speaking assessments that mainly involve interviews are known as oral proficiency interviews (OPI; Qian, 2009).

Pros

- Speaking is performed in a manner that almost duplicates real-life communicative situations.
- The assessment promotes a natural speaking environment that can lead to higher evaluation accuracy.

Con

The assessment may lack authenticity because learners are completely aware that they are interacting with language assessors rather than real interlocutors, inhibiting their abilities to speak naturally (Yoffe, 1997).

Note: One of the most renowned speaking assessments is the [ACTFL OPI](#). This OPI starts with a series of warm-ups followed by a series of questions with a gradual increase in difficulty, eliciting learners to respond with increasing levels of complexity. The performance of learners

for this assessment is typically evaluated by the interviewer or additional evaluators through observation of the interaction in real-time or after the interview is completed (Yoffe, 1997).

Semidirect Speaking Assessments

Semidirect speaking assessments are popular for large-scale testing situations without interlocutors. In semidirect methods, prerecorded questions are prepared for the learners, and they take the test under laboratory-like conditions. Instead of being rated in real-time, answers are recorded and evaluated later (Qian, 2009).

Pros

- Without interlocutors, this method promotes higher reliability (more consistency and stability) and efficiency (less time and lower cost of administration).
- The construct irrelevant variance (extraneous, uncontrolled variables during testing) may be reduced due to the absence of interviewer influence (Ginther, 2013). This creates a more equitable environment for learners to perform up to their potential.

Cons

- Learners tend to display higher levels of formality and cohesion in their responses, causing hesitations and longer pauses when speaking, which can result in lower speech fluency.
- Reports from surveys have revealed that most learners have found communication with a recorder to be unnatural (Ginther, 2013).

In addition to direct and semidirect assessments, there are many other assessment methods that can be used to evaluate English speaking. Some of the more common of these methods are indirect assessments, self-assessments, peer assessments, and portfolio assessments.

I have decided to focus on direct and semidirect assessments because these two assessments are often considered to be more reliable and valid than other assessment methods; they provide not only opportunities for learners to communicate in realistic real-time situations, but also immediate feedback for learners to improve their speaking abilities over time. If you'd like to read more about the other assessment types, following are some resources:

- [Indirect assessment](#)
- [Self- and peer assessment](#)
- [Portfolio assessment](#)

2. Assessment Scales

Once you have decided on the type of assessment to use, the evaluation scale must be carefully developed. Two of the most globally accepted evaluation scales for speaking assessments involve holistic and analytic scales. The effectiveness of a speaking evaluation will heavily depend on the appropriateness of the assessment scale you chose. A suitable evaluation scale must be carefully chosen to align with the purpose of the evaluation. Following, we look at some

guidelines for creating your own scales as well as two of the most popular scales used for speaking assessments.

Holistic Scale

The holistic scale is one of the most basic evaluation scales, involving a single scale that measures all criteria together as a whole. To create a holistic scale, your first step is to establish in-depth descriptors and benchmark performance indicators so the levels on the scale are distinguished. Each level on the scale should include clear descriptions of a few important categories in the learner's speaking performance. For instance, in a speaking assessment, the categories for a holistic scale could include

- the ability to speak clearly and fluently,
- the level of grammatical structures used, and
- the level of vocabulary used.

Due to the simplicity of a holistic scale, it can make sense to narrow down the number of categories. Including too many categories for each level will make the assessment too complex, making it difficult to pinpoint the exact level of the learner's performance. If you wish to assess more categories of a learner's speaking performance, you should consider using an analytic scale. Here is a link to an example of what a holistic rubric could look like for a speaking assessment: [Holistic Rubric Example](#).

Pros

- Increased practicality; the assessors can score the student's performance in an extremely efficient manner (Metruk, 2018).
- The descriptors and benchmark performance indicators provide some washback effect for students because they can understand where their speaking level lies. These indicators can also help the students understand the requirements for the next level (Brown, 2017).

Con

Even though the scale is supported with descriptors for each level, it may not be sufficient for instructional and placement decisions to be effectively made (Ginther, 2013).

- **ACTFL OPI:** The scale for ACTFL OPI is illustrated as an inverted holistic scale ranking the levels from lowest to highest. Each level on the scale is accompanied by level descriptors, which are used to represent the qualitative summary of the observations by the raters. During the speaking performance measurements, benchmark performances are decided for the exemplification of the scale's level and its descriptors. The main components in the descriptors of each level are typically pronunciation, grammar, vocabulary, phonological control, and organization (Brown, 2017).

Analytic Scale

Analytic scales provide an in-depth assessment of speaking by breaking down the speaker's performance into several categories. To create an analytic scale, your first step is to identify categories (e.g., pronunciation, fluency, grammar structures, vocabulary, content) and determine the performance indicators for the oral assessment. Each category should reflect a different aspect of the learner's language skills.

Once the categories have been set, the next step would be to create descriptions for different levels of performance in each category. The descriptions should serve as a benchmark to help evaluators rate the learner's level of performance in each category. Here is a link to an example of what an analytic rubric could look like for a speaking assessment: [Analytic Rubric Example](#).

Pro

Detailed feedback on areas of strengths and weakness can be provided to both the evaluator and learner. This can be useful for instructional and placement decisions to be effectively made.

Con

Analytic rubrics are more time-consuming because they take much more time to create and apply, compared to a holistic rubric.

- **CEFR Scale:** One of the more renowned analytic scale assessments is Pearson's Versant Test, which utilizes the Common European Framework of Reference (CEFR) scale. This semidirect artificial intelligence speaking assessment organizes the learner's performance measurement into four categories: sentence mastery, vocabulary, fluency, and pronunciation. Each category receives a score that contributes to one overall score in the evaluation. The overall score falls under one of the levels in the CEFR scale to determine the learner's level.

3. Evaluator Preparation and Training

The last element to consider for speaking assessments involves the evaluators. Evaluators may have different tendencies in scoring speaking assessments, which can negatively affect the consistency and quality of the evaluation. To combat this, you can conduct evaluation training with a single rating criteria. Aside from the training, evaluators can also prepare by familiarizing themselves with the evaluation's sequence of operation to strengthen evaluation consistency.

Pro

Higher inter-rater reliability as evaluators would share one consistent rating criteria, avoiding different tendency scoring.

Con

Can be time- and cost-consuming to plan, prepare, and carry out the evaluation training for the assessment.

Conclusion

Direct and semidirect methods involving an OPI format are the best options for assessing a student's natural communicative ability. Once you have decided on the type of assessment to use, the evaluation scale must be carefully developed.

Finally, once your scale has been constructed, it is highly recommended that you review the scale by piloting it and checking if there are any practicality or consistency issues before the actual assessment. In addition, conducting a post-use review of the scale about what went well and what didn't go well would also be beneficial for improving the scale for future use. By using reliable and valid assessments, educators can ensure that English learners receive appropriate feedback to develop their language skills and achieve higher levels of speaking proficiency.

References

- Brown, D. B. (2017). Developing and using rubrics: Analytic or holistic? *Shiken*, 21(2), 20–26. https://hosted.jalt.org/sites/jalt.org.teval/files/21_02_20_Brown_Statistics_Corner.pdf
- Ginther, A. (2013). Assessment of speaking. *Research Gate*. https://www.researchgate.net/publication/277707664_Assessment_of_Speaking
- Metruk, R. (2018). Comparing holistic and analytic ways of scoring in the assessment of speaking skills. *Research Gate*. https://www.researchgate.net/publication/323629161_Comparing_Holistic_and_Analytic_Ways_of_Scoring_in_the_Assessment_of_Speaking_Skills
- Qian, D. (2009). Comparing direct and semi direct methods for speaking assessment: Affective effects on test takers. *Language Assessment Quarterly*, 6(2), 113–125. <https://doi.org/10.1080/15434300902800059>
- Yoffe, L. (1997). An overview of the ACTFL Proficiency Interview: A test of speaking ability. *Shiken*, 1(2), 2–13. <https://hosted.jalt.org/test/PDF/Yoffe1.pdf>

Chris Huang, from Vancouver, Canada, has taught English for more than 7 years in Japan. He has taught all ages of students, from small children to adults. Currently, he is an English instructor at Nagoya University of Foreign Studies, Sugiyama Jogakuen University, and Nagoya University of the Arts, located in Aichi, Japan. Chris has a BA in business administration from Simon Fraser University and a master's in TESOL from Horizons University.