Problematizing Speaking Assessment

Sarra Mejri

Faculty of Letters, Arts and Humanities of Manouba, Tunisia

Abstract

The present paper addresses the intricate question of oral language performance and how it can be assessed. More specifically, the paper tackles the issue of testing speaking in an academic setting and looks into the various factors that intervene in this testing, coupled with the context of the testing process, such as cultural expectations of stakeholders, stakes, consequences…etc. It has also looked into assessing speaking in an ESP context and has ended with a few recommendations on how to improve the quality of this assessment.

Keywords: Speaking, assessment, ESP, assessment components.

1. Introduction

The English language occupies a peculiar position in global communication. It is an "international language in both a global and a local sense" (McKay 2002 p.5). Thus, this language plays a key role in both language teaching and learning fields. In learning English, as a second or a foreign language, learners are required to master the four skills efficiently; speaking, listening, reading and writing. However, speaking, as an oral productive skill, seems intuitively the most important skill to be mastered by the majority of non-native English speakers because the EFL learners' chief interest is how to communicate their ideas, opinions and views and to develop their oral fluency and accuracy inside classrooms as well as outside the learning environment (McKay 2002 p.12).

In fact, speaking is the oral channel of communication through which the learners' major concern is to interact with each other, express opinions, make arguments, transmit information and understand each other correctly. In support of this view, Florez (1999) defines speaking as "an interactive process of constructing meaning that involves producing and receiving and processing information" (p. 1). That's why learners of a foreign language must be able to produce correct chunks and sentences, and to understand when, why, and in what ways they produce language (sociolinguistic competence). So, learning to speak English requires more than knowing its grammatical and semantic rules. Due to this growing necessity of achieving effective communication and interaction in society, learners' oral language performance should be assessed, developed and improved.

This paper tackles the issue of testing speaking in an academic setting. In fact, it's the teacher who plays a role in discovering the weaknesses of his students and helps them interact successfully in the target language through assessing their productive skills. In this context, according to Skehan (1998 p.153), a test is "a systematic method of eliciting performance which is intended to be the basis for some sort of decision making".

When designing a speaking test, teachers should be aware of some criteria related to tests such as reliability, validity and practicality. In fact, Messick (1989 cited in Fulcher 1999 p.224) provides a technical definition of validity. He states that "validity is an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationale support the adequacy and the appropriateness of inferences and actions based on test scores or other modes of assessment".

In assessing the learners' mastery of speaking skills, testers should select and rely on a bundle of tests and tasks. Tasks designed to assess the candidates' speaking skills should be based on the participants' needs analysis, and should be as authentic as possible i.e., extracted from real-life situations. They should also be contextualized, complex and non-intrusive. Authentic assessment is a "form of assessment in which students are asked to perform real-world tasks to demonstrate meaningful application of essential knowledge and skills" (Mueller, 2016). After selecting the appropriate materials, teachers ask their EFL learners to perform activities, such as narrating events in a story, describing pictures, responding orally to interviews etc, then, they try to rate their learners' oral proficiency based on a well-

determined rating scale. The purpose of designing a language test is to affect positively the learning process and to improve the examinees' oral abilities (Norris, Brown, Hudson and Yoshioka 1998 pp.8-9-10).

So, the rating system is efficient in evaluating oral performance in a reliable, fair, and valid manner, based on a particular scale and appropriate criteria. The use of a variety of assessment procedures will assist in providing more valid measures of oral ability. So, evaluation should become a tool for planning and teaching.

2. Literature Review

Special attention has been paid to the speaking skill in first, second and foreign languages as it facilitates people's transmission of information, not only in the educational settings but also in our daily life. In this regard, speaking, as seen by Hornby (1995 p. 826), is using language and producing sounds and words to convey ideas, thoughts, feelings and needs in the form of speech. As a vital part of our daily life, speaking is taken for granted (Thornbury 2005 p. 1). In this respect, Ur (2000 p. 120) claims that

> of all the four language skills, speaking seems intuitively the most important: people who know a language are referred to as "speakers" of that language, as if speaking included all kinds of knowing and many if not most foreign language learners are primarily interested in learning to speak, meanwhile the classroom activities that develop learners' ability to express themselves through speech would, therefore, seem an important component of a language course.

Accordingly, Lado (1961 cited in Fulcher 2000) argues that "the ability to speak a foreign language is, without doubt, the most prized language skill, and rightly so…" (p. 487).

With the advent of communicative language teaching, research on the development of learners' levels of oral proficiency has gained ground among various educationists, practitioners and linguists (Nakamura, 1993). The development of the speaking skill is a challenging endeavor, as it is affected by its componental and context-dependent nature. Speaking is a basic skill in language learning and teaching. It is the transmission of information by sending and receiving oral messages through the use of utterances, phrases, and expressions. In this vein, Richards (2006) argues that "natural language use" occurs "when a speaker engages in meaningful interaction and maintains comprehensible and ongoing communication despite limitations in his or her communicative competence" (p. 14). The mastery of speaking skills is considered a necessity for most EFL and ESL learners (Richards, 2008, p. 19; Florez, 1999). The learners of a foreign or a second language prioritize the speaking skill at the expense of other language skills in the learning process because they believe that if they master the oral ability, then, they will find it easy to master the other skills and thus, they will know that language and become proficient learners. This view is supported by the common question "Do you speak English?" asked foreign language learners to check their oral abilities and to what extent they master the target language.

Unlike the other language skills, the speaking skill plays a key role in the real-life context and in the educational setting. In this regard, Thornbury and Slade (2007) maintain that the spoken language "is structurally patterned and displays an orderliness that is neither chaotic nor random but, rather is tightly organized and coherent" (p.27).

In language teaching, the process of speaking assessment is viewed as very contentious and intricate (O'Sullivan, 2006)."How do teachers discover their learners' strengths and weaknesses in using the target language orally?" This question is situated at the heart of the language teaching and testing systems because it points out the necessity of tackling the issue of oral speaking assessment, which is not a new issue in the general educational measurement field, to discover the various levels that learners belong to.

Oral performance assessment has become a vital issue in language testing, as the speaking skill becomes more appealing in language teaching than before, due to the implementation of performance assessments(Hatley and Sporing 1999 pp74-75). The latter plays a key role in the educational setting as they help "document and encourage critical, creative, and self-reflective thought" (Moss 1992 pp. 229-230). This view goes in line with Underhill's (1987) argument in which he states that "the learner will probably produce more of the foreign language than he does in an hour-long written test, so the importance given to an oral test should not be reduced on the account of its brevity" (p.40).

Despite its importance, testing the speaking skill is a difficult and complex process as, sometimes, test takers can pronounce separate letters and phonemes effectively but they fail to convey ideas, opinions, and information to other people. On other occasions, test takers can transmit the intended message to others and exchange ideas with people, but their speeches can be full of syntactic and phonological mistakes. According to Bachman and Palmer (1981), "one of the areas of most persistent difficulty in language testing continues to be the measurement of oral proficiency" (p. 67). In this regard, Heaton (1988) affirms that "speaking is an extremely difficult skill to test, as it is far too complex a skill to permit any reliable analysis to be made for objective testing" (p. 88).

Similarly, Kitao and Kitao (1996) assert that

> speaking is probably the most difficult skill to test [...] A speaker can produce all the right sounds but not make any sense, or have great difficulties with phonology and grammar and yet be able to get the message across. Also, success in speaking depends to a great extent on the listener (p.2).

The presence of both speaker and listener is necessary for the spoken medium as they will ensure the communication process through the exchange of oral messages. Accordingly, Madsen (1983) affirms that "the testing of speaking is widely regarded as the most challenging of all language tests to prepare, administer and score" (p. 147). Deciding whether fluency or accuracy will be measured and choosing the criteria to be followed in the assessment process are two factors that can contribute to the difficulty of testing the speaking skills.

Several factors contribute to the mark assigned to an examinee's response to a test designed to measure his/her ability to produce an oral performance and communicate effectively. Figure 1 displays the interaction between the different components of speaking assessments.
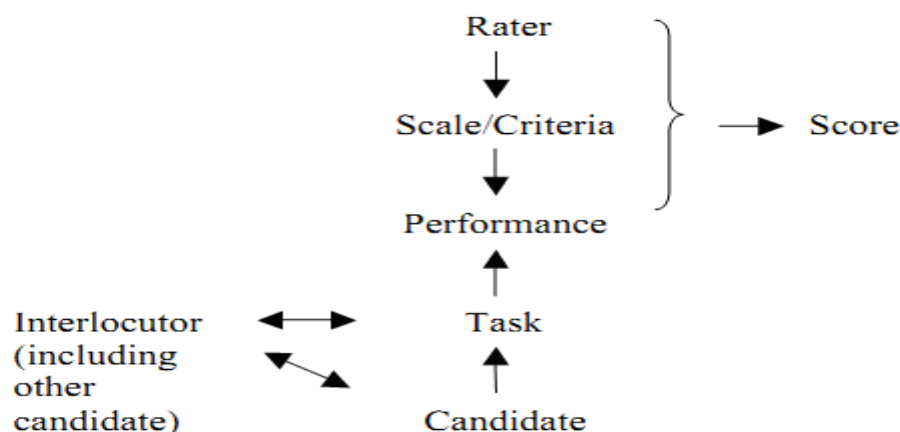
Hence, different factors can influence the score assignment task while assessing candidates' speaking proficiency. Prompt types can have an impact on test takers' oral performances. Task familiarity can be an advantage for examinees. For instance, some examinees may perform better with a group or paired discussion because they are used to interact with others in such authentic situations. Interlocutors' characteristics can also affect not only oral performances (being very assertive, the members of the test takers' group can have an impact on the test taker's oral ability in a group oral test) but also the way raters assess test takers' speaking abilities (in a one-on-one interview, for example, the interviewer can affect scores by being more or less supportive with his interviewees). Rating scales play a pivotal role in evaluating learners' language skills, by linking the assigned marks to the construct being measured (Ockey and Li 2015, pp.3-4). To minimize such variances, rating rubrics should clearly reflect the construct. Moreover, raters play a crucial role in the oral communication assessment process. They are considered as a source of error measurement as they can affect marks, depending on the way they are trained to interpret the scoring rubrics and assess oral production. In this respect, Upshur and Turner (1999) elucidate that "the rater is not only an additional source of measurement error but, as a methods facet, may also exert systematic- although unwanted- effects upon scores" (p.87).

These different factors coupled with the context of the testing process, such as cultural expectations of stakeholders, stakes, consequences…etc., can be viewed as sources of construct–irrelevant variance in assessing learners' oral skills. As one of the major threats to validity, construct-irrelevant variance (CIV) is defined as "variance in test-taker scores

attributable to extraneous factors that distort the meaning of the scores, thereby decreasing the validity of the proposed interpretation" (AERA et al., 2014, p. 217). This means that EFL teachers are bound to face numerous challenges.

Hence, as Luoma (2004) stressed, being orally proficient in a non-native language is challenging and exacting. It is thus hard not only to master the speaking skill of a foreign language but also difficult to test this skill. In testing students' oral performances, "the assessors need to make an instantaneous judgment on a range of aspects of what is being said, as it is being said" (p.x). That is to say, rating learners' oral proficiency should be based on a given aspect during a short time. This process requires a lot of knowledge, skills, experience, training and practice which is difficult even for experienced raters. The complexity of the speaking assessment process is further supported by O'Sullivan's (2012)claim that "[i]t is customarily believed that the most troublesome tests to expand and execute are tests of spoken language ability" (p.234). This goes in line with Chuang (2009) who affirms that since there are diverse internal and external factors that influence raters, evaluating oral productions appears to be one of the uttermost difficult tasks to administer.

Testing speaking skills is also a complicated process for EFL learners whose major objective is to prove their English proficiency in authentic contexts. Their tasks consist in applying what they have learned, like grammatical rules and vocabulary, and using it to communicate with others and convey their opinions. Students at the tertiary level for instance should be prepared to perform their speech in front of the public in different spoken forms, such as oral presentations to be able to generate fluent performances. As a subcomponent of L2 proficiency, L2 fluency is defined as the "speedy and smooth delivery of speech without (filled) pauses, repetitions, and repairs" (p.224). It is measured based on different criteria and aspects, namely silent pauses, non-lexical filled pauses, repetitions, and corrections, and length of silent pauses, and mean duration of syllables (p.226). The intricacy of speaking is associated with learners' ability "to fill time with talk… to talk in coherent, reasoned and semantically dense sentences… to have appropriate things to say in a wide range of contexts"; and to "be creative and imaginative…in language use" (Richards, 1990, p. 75). However, Luoma (2004 pp.17-18) finds that the speakers of the target language usually utilize "vague words", "fillers", "hesitation makers", and "fixed phrases" in everyday conversations. In the same context, Chomsky (1965 p.4) states that "a record of natural speech will show numerous false starts, deviations from rules, changes in mid-course, and so on".

Some common factors are interrelated to come up with a final mark in oral assessments. Raters are not aware of what the test takers think and feel about the assessment task. (Davis, 2009 p. 367). Examinees may experience some stress, anxiety, or distrust when they become the focus of their teachers' or their friends' attention during an oral exam. In this regard, speaking in public, such as making oral presentations, can undermine learners' trustfulness and is ineffective in developing students" oral competences (King 2002 p.403). These problems may hinder students from performing their speaking competences.

3. Implications

Several implications were suggested in this research in which measuring learners' speaking skills were investigated in the English for specific purposes (ESP) field. Oral assessment is a multi-faceted tool that helps students boost their speaking skills, acquire the target language, and enhance the learning process. Teachers can apply various testing methods to encourage students' positive involvement in the learning process depending on the context in which language is used. ESP performance assessment, for instance, requires students' mastery of a higher degree of technical language in addition to their good communicative competences to perform the target language successfully in a particular academic, professional or vocational field. In this respect, Knoch and Macqueen (2019) define assessment in ESP as "any assessment process, carried out by and for invested parties, which is used to determine a person's ability to understand and/or use the language of a professionally-oriented domain to a specified or necessary level" (p. 2).

Testing learners' technical knowledge of using English for real communication in context-relevant testing situations implies the selection and creation of authentic tasks and suitable materials. According to Douglas (2013), "ESP assessment is clearly a definable subfield of language assessment, with its focus on assessing ability to use language precisely to perform relevant tasks in authentic contexts while integrating appropriate aspects of field-specific background knowledge" (pp.378- 379).

Alternative assessment activities, such as portfolios, journals, observations, exhibitions, oral presentations, experiments, interviews, and projects can be also suggested. They enable learners to make plans according to their own learning needs and interests, boost their motivation and enhance the learning process. These authentic testing strategies reflect students' performance in different aspects and offer continuous feedback that helps judge their oral skills.

A further implication can be based on the use of case studies as a formative assessment technique in an ESP course. They improve students' motivation to learn a language, develop problem-solving skills, analytical and critical thinking skills as well as language skills in real-world contexts, promote interaction and active learning, and induct students' into their professional world. In this vein, Rodgers (1993) reiterates that "realistic cases provide an opportunity for hands-on experience which allows students to apply 39 their theoretical knowledge and develop managerial skills" (p. 3).

References

American Educational Research Association. (2014). *Standards for educational and psychological testing.* American Educational Research Association American Psychological Association National Council on Measurement in Education.

Bachman, L. F., & Palmer, A. S. (1981b). The construct validation of the FSI Oral Interview. *Language Learning*, vol.31, pp. 67-86.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Massachusetts Institute of Technology: Cambridge, Massachusetts.

Chuang, Y. Y. (2009). A study of college EFL students' affective reactions and attitudes toward two types of performance-based oral tests. *Journal of Educational Research*, vol. 43, pp. 55-80.

Davis, L. (2009). The influence of interlocutor proficiency in a paired oral assessment. *Language Testing*, vol. 26(3), pp. 367-396.

Douglas, D. (2013). ESP and assessment. In B. Paltridge, & S. Starfield (Eds.), *The handbook of English for Specific Purposes* (pp. 367-383). Wiley Blackwell.

Florez, M. A. C. (1999). Improving Adult English Language Learners' Speaking Skills. National Center for ESL Literacy Education.

Fulcher, G. (1999). Assessment in English for Academic Purposes: Putting Content Validity in Its Place. *Applied Linguistics*, vol.20, (2), pp.221-236. Oxford University Press.

Fulcher, G. (2000). The communicative legacy in language testing. *System*, vol. 28, pp.483-497. Pergamon.

Hartley, L., & Spöring, M. (1999). Teaching communicatively: assessing communicatively?. *Language Learning Journal*, vol. 19(1), pp. 73-79.

Heaton, J. B. (1988). Writing English language Tests. *Longman Handbooks for Language Teachers.*

Hornby, A. S. (1995). *Oxford Advanced Learners' Dictionary of Current Language*. Oxford: Oxford University Press.

King, J. (2002). Preparing EFL learners for oral presentations. *Dong Hwa Journal of Humanistic Studies*, 4, pp. 401–418.

Kitao, S.K and Kitao, K. (1996). Testing Speaking. Doshishka University. Kyoto, Japan.

Knoch, U., & Macqueen, S. (2019). *Assessing English for professional purposes*. Routledge.

Luoma, S. (2004). *Assessing speaking*. Cambridge University Press.

Madsen, H. S. (1983). *Techniques in testing.* Cambridge University Press.

McKay, S. L. (2002). *Teaching English as an international language*. Oxford Handbooks for Language Teachers.

McNamara, T. F. (1996). Chapter two: Second language performance assessment. *Measuring Second Language Performance.* London and New York: Longman.

Moss, P. A. (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. *Review of Educational Research*, vol. 62(3), pp.229-258.

Mueller, J. (2016). Authentic Assessment Toolbox. Home Page.

Nakamura, Y. (1993). Measurement of Japanese college students' English Speaking ability in a classroom setting. *Unpublished doctoral dissertation*, International Christian University, Tokyo.

Norris, J. M, Brown, J. D, Hudson, T, & Yoshioka, J. (1998). Designing second language Performance.

O' Sullivan, B. (2006). Modelling Performance in Oral Language Tests: *Language Testing and Evaluation*.

O' Sullivan, B. (2012). Assessing speaking. In C. Coombe, P. Davidson, B. O' Sullivan, and S. Stoynoff (Eds.) *the Cambridge guide to second language assessment*, pp. 234- 246. New York: Cambridge University Press.

Ockey, G. J., & Li, Z. (2015). New and not so new methods for assessing oral communication.

performance-based oral tests. *Journal of Educational Research*, vol. 43(2), pp. 55-80.

Richards, J. C. (2006). *Communicative Language Teaching Today*. New York: Cambridge University Press.

Richards, J. C. (2008). *Teaching listening and speaking: From theory to practice*. Cambridge: Cambridge.

Richards, Jack C. 1990. *The Language Teaching Matrix*. New York: Cambridge University Press.

Rodgers, D. (1993). Teaching ESP-B and Business Communications with the Case Study Method.

Skehan, P. (1998). *A cognitive Approach to Language Learning*. Washington, DC: AERA.

Thornbury, S. & Slade, D. (2007). *Conversation: from description to pedagog*y. Cambridge Language Teaching Library

Thornbury, S. (2005). *How to teach speaking*. Cambridge University Press.

Underhill, Nic. (1987). *Testing the Spoken Language*. New York: Cambridge University Press.

Upshur, J. A & Turner, C. E. (1999). Systematic effects in the rating of second-language speaking ability: Test method and learner discourse. *Language Testing*, volume 16, 1, pp. 82-111.

Ur, P. (2000). *A course in language teaching*: *Practice and theory*.