THE VALIDITY OF SPEAKING TESTS

¹Kalmuratova Inkar, ²Muradkasimova Kamola

¹a master student of NukSPI, ²a scientific advisor, PhD Of Uzbekistan State World Languages university

Abstract. Speaking skill is considered as one of the crucial skills this is due to the fact that it is needed in our everyday interaction. Therefore, the article aims to investigate speaking skill and the ways to measure it. Moreover, the research deals with the validity of speaking tests. For this purpose, the research analyzed the concepts of scholars on the validation of speaking skill. The findings of the research depicts that it is important to assess speaking skill with the help of tests.

Keywords: validity, speaking skill, a test, a method, to assess, types of validity.

Introduction. Speaking is a productive skill where speakers creatively combine grammatical, lexical and discursive structures. Therefore, to elicit only the structures that test takers are interested in, the stimulus they design should be carefully crafted and should prohibit test-takers from using other structures and avoiding or paraphrasing the target structures.

Speaking tasks and their evaluation criteria should be designed based on the analysis of the students' needs and the test's aims. To choose the most appropriate tasks for their tests, exam developers should clearly understand what the test scores will be used for and what type of information the test takers need. The speaking-assessment tasks must also be authentic (i.e., they should involve realistic and genuine communicative interactions) and contextualized (i.e., as "normal" conversations do not occur in a vacuum, the exam tasks should describe the conversational contexts in as much detail as possible).

With more open-ended tasks, test takers have the liberty to respond with a wider variety of words and structures not anticipated by the test writers. To avoid problems that might damage both the reliability and validity of the exam, test writers should prepare detailed analytical rubrics where every type and piece of information is allocated an individual score. Speaking rubrics should be as detailed as possible and, depending on their aims of the exam, should allocate points for pronunciation, fluency, grammar, vocabulary, discoursal elements, and level of pragmatic appropriacy.

When designing tasks for testing students' oral ability, Hughes [4.113] suggests that the following three rules are followed:

(i)   Set tasks that form a representative sample of the population of oral tasks that we expect candidates to be able to perform.

(ii)   Create tasks that elicit behavior that truly represent the candidates' ability.

(iii) Design the tasks in such a manner that the collected sample of behavior can and will be scored validly and reliably.

Ur supports including oral proficiency tests in language exams: In principle, a language test should include all aspects of language skills-including speaking. Speaking is not just "any skill"- it is arguably the most important, and therefore, should take priority in any language test. If you have an oral proficiency test at the end of a course, then this will have a "backwash effect": teachers and students will spend more time on developing skills during the course itself. Conversely, if you do not have such a test they will tend to neglect them. Students who speak well but write badly will be discriminated against if all or most of the test is based on writing [5.134].

However, assessing oral ability is problematic due to its being evaluated by human raters and the number of the raters as well. It is claimed by Alderson, Clapham and Wall that scoring of oral ability is highly subjective and this is one of its characteristics[1]. Heaton also expresses the importance of the rater and the difficulty of making objective judgments:

- success in communication often depends as much on the listener as on the speaker
- a particular listener may have a better ability to decode the foreign speaker's message or may share a common nexus of ideas with him or her, thereby making communication simpler. Two native speakers will not always, therefore, experience the same degree of difficulty in understanding the foreign speaker [3.88].

One of the most essential elements in speaking tests is stated as the validity of speaking tests. Validity has recently emerged as a most important consideration in developing and evaluating language tests. The validity of a test can only be established through a process of validation, and this must ideally be done before the results can be used for any particular purpose. In order to carry out such validation, a validation study has to be undertaken, on the basis of which one can arrive at a conclusion as to whether the interpretations and uses of the test results are valid.

From this point, Hughes views a test as a valid "if it measures accurately what it is intended to measure" [4.22], which is a rather general statement that can be approached in more specific way. For example, Weir perceives validity as "the extent to which a test can be shown to produce data, i.e., test scores, which are an accurate representation of a candidate's level of language knowledge or skills. In this revision, validity resides in the scores on a particular administration of a test rather than in the test *per se*" [6.12] However, Heaton argues that, "this validity is obtained as a result of comparing the results of the test with the results of some criterion measure such as: the teacher's

ratings or any other such form of independent assessment given at the same time. Result as obtained are measures of the test concurrent validity" [3. 161].

Regarding testing speaking, Weir argues that, "there is a strong case for testing spoken language performance directly, in realistic situations, rather than testing hypothetical knowledge of what might be said. If we wish to make statements about capacity for spoken interaction we are no longer interested in multiple choice, pencil-and-paper tests, that is, indirect tests of speaking where spoken language is conspicuously absent. To test speaking ability, we should require candidates to demonstrate their ability to use language in ways which are characteristic of interactive speech, i.e., to process the language in the way" [6.103].

Depending on the aims of the exam and the variables test administrators choose to evaluate, the final score related to the students' performance can be based on three evaluation techniques:

(i) ***Error-Based Method:*** For this method to be successfully implemented, there should be three examiners in the room. The first one is the test taker's interlocutor (i.e., the person giving instructions, asking questions, providing prompts and answering questions posed by the test taker). The second examiner counts all of the test taker's utterances (e.g., words, phrases and sentences; if the focus is on a particular word category: how many physical appearance adjectives are utilised). The last examiner counts the number of mistakes made by the test taker The final score is the ratio of the utterances vs mistakes (e.g., 100 utterances / 20 mistakes = 50 Overall score;). With the error-based method, it is easy to objectify and calculate students' final scores. However, when it comes to speaking, such mechanical analyses do not always provide the most valid and reliable evaluation methods.

(ii) ***Analytic Method:*** This scoring begins by designing detailed rubrics about the micro and macro skills considered important with the group assessed in the exam. A few examples could be:

Beginner level: correct pronunciation, word stress, and intonation contours

Intermediate level: produce language chunks, generate fluent speech, respond with relevant phrases

Advanced level: generate fluent and intelligible speech, use grammatically correct sentences, follow pragmatic conventions. The speaking exams are audio or video-recorded. Examiners watch the recording a number of times, every time scoring the test taker's performance related to just one of the listed criteria (e.g., word stress). Madsen (1983) argues that both teachers with no or little specialised training as well as highly trained examiners can use the Analytical Method of scoring as it is "consistent and easy to use" [3. 167].

(iii) ***Holistic method/Goal Oriented Method*** A method where examiners evaluate several criteria simultaneously and assign an overall performance score is called a holistic method of scoring. With this scoring method, individual criteria such as pronunciation, fluency, grammar, vocabulary etc. are still considered, but the more important factor affecting the scoring is whether or not test takers are able to achieve their goals (e.g., ask for directions, describe a person, book a room). Language errors that impede successful communication are more heavily penalized, while the ones that do not are penalized more lightly.

Conclusion. Taking all into account, it can be concluded that testing speaking skill is considered as challenging and problematic due to its subjective character. However, the research presented some valuable methods to evaluate test takers speaking skill. Moreover, the validity of speaking test is also investigated in the research. The result of the research depicted that validity in speaking tests is regarded as one of the crucial elements in assessment. Therefore, it is recommended to design speaking tests as valid as possible.

References:
1. Alderson, J. C., Claphamn, C.,& Wall, D. (1995). Language test construction and evaluation. Cambridge: Cambridge University Press.
2. Davies, A. and C. Elder. 2005. Validity and validation in language testing. In E. Hinkel (ed.) Handbook of research in second language teaching and learning. Mahwah, New Jersey: Lawrence Erlbaum.
3. Heaton, J. B. (1990). Writing English Language Tests. Longman
4. Hughes, A. (2003). Testing for Language Teachers (Second Edition). Cambridge University Press
5. Ur,P.(1996). A Course in language teaching. Cambridge: Cambridge University Press
6. Weir, C. (1988). Communicative Language Testing. Exeter University Press.