



Specifying Criteria for the Assessment of Speaking Skill: A Library Based Review

Muhammad Ahmad^{1#}, Hafiz Muhammad Qasim^{1*} and Qamar Khushi²

¹*Department of Applied Linguistics, Government College University, Faisalabad, Pakistan*
*E-mail: #<ahmad453@yandex.com>, *<m_qasim@hotmail.com>*

²*Department of English, Bahauddin Zakariya University, Multan, Pakistan*
E-mail: qamarkhushi@gmail.com

KEYWORDS Speaking Skill. Speaking Test. Speaking Test Construction. Speaking Assessment Criteria. Speaking Test Principles. Speaking Test Scoring

ABSTRACT This study aims to specify criteria for speaking skill assessment. Mainly, it focuses on answering questions about speaking skill, speaking skill assessment, different levels and types of speaking, suitability of tasks, scoring procedure and current trends in speaking skill testing. It is a library based descriptive study. Its material has been retrieved from different web sources. It concludes that the key to assess speaking lies in understanding of the continuum of spoken language, task types and scoring rubric. Thus, the construction of a sound speaking test requires time and effort in specifying the criteria of speaking, selecting tasks to elicit particular behaviour and in developing practical scoring rubric. To have a knowledge of the in current trends speaking tests, one can consult IELTS, TOEFL or TOEIC. Moreover, modern world has started testing speaking skills through technology. In this regard, CBLTs and WBLTs are popular these days.

INTRODUCTION

Language testing is the subfield of applied linguistics (Grabe 2010). It mainly focuses on the assessment of first, second or other languages in educational institutions to rate the learning of a particular language. Language assessment is also utilised for citizenship, workplace, immigration, and asylum purposes (Clapham and Corson 1997; Inbar-Lourie 2008). The idea of testing languages, for the longest time, has always been revolving around assessing the knowledge of languages themselves. But nowadays, the notion of testing for communicative competence is becoming more popular (Bailey and Nunan 2004; Canale 2014; Harding 2014; Morrow 2018; Nation and Newton 2008). In the process of testing communicative competence, listening and speaking most commonly used tasks include role play and gap completion (Bellack et al. 1979; Kitao and Kitao 1996; Lynch and McNamara 1998). For ESL teachers, it is very important to improve the learn-

ers' delivery skills, enhance their confidence, and develop their critical thinking and organisation skills. For this purpose, such a reliable and valid approach of assessment is needed as can help decide whether the goals are achieved or not. Similarly, oral communication needs such a scheme of assessment as can originate from discrete language skill classes such as the comprehension of listening (Dandonoli and Henning 1990; Nakamura and Valens 2001; Riggio 1986). Therefore, language teachers as well as language testers require such a scheme as can take a subjective-qualitative observation and change it into an objective-quantitative measure (Butler and Stevens 1997; Nakamura and Valens 2004; Sercu 2004). An issue, in the evaluation process, is the choice of criteria for the evaluation of learners' performance. The said criteria must be one of the first steps in devising a speaking test. Learners must be made aware before time of what is likely to be expected from them. It can help them decide on what ground the performance will be assessed (Butler and Stevens 1997).

Speaking is, doubtlessly, one of the most complicated skills to assess (Chou 2018; Correia 2016; Guettal 2008; Levelt 1994; Morrow 2018; Rychtarik 2014; Schmidt 1992). It involves such

Address for correspondence:

Muhammad Ahmad
Department of Applied Linguistics,
Government College University,
Faisalabad, Pakistan
Cell: +92300-8135645
E-mail: ahmad453@yandex.com

skills as may have no or little connection with one another. Resultantly, they affect on objective testing. A major complexity is the true realization of speaking skills in testing. For, it is not easy to assess a large number of students in a short time. Therefore, the assessor is put under great pressure (Heaton 1988). In the same way, Kitao and Kitao (1996) and Knight (1992) argue that there is yet dearth of appropriate criteria for assessing speaking skills. In addition to it, the evaluation and scoring of the speaking skills are still considered as major problems. In O'Malley and Pierce's (1996) view, selection of appropriate assessment tasks and determination of evaluation criteria is another major challenge. According to Celce-Murcia et al. (1996) and Pennington (1999), the testing of learners' ability to speak lacks solid grounding on theories, reliable test designs and pedagogy. Similarly, Brown (2004) describes two major challenges in assessing speaking: (1) the interaction of listening and speaking (for example, the use of much clarification) can make it difficult to treat speaking apart, (2) the speaker's strategy to dodge certain forms to convey meaning may make it difficult for test makers to design a solid elicitation technique (one that can result in the expected target form).

Still another challenge in testing is that it is likely to find such individuals as can make the diverse sounds of a second language suitably. Hence, there may still be the chance that they can fail to convey their thoughts appropriately. This is one of the major problems that speaking skill assessors come across while assessing the verbal production of testees. On the other hand, the opposing situation can also take place. Such as some of the individuals may be able to express their thoughts clearly, but at the same time they may fail to utter all of the sounds appropriately. Still another problem is that listening and speaking skills are interrelated (Idrissova et al. 2015). Therefore, sometimes it becomes difficult to keep them mutually-exclusive. In most of the cases, speaking depends on the comprehension of spoken inputs. Consequently, it impacts badly on the testing of speaking skill because the tester is unable to know whether he is testing listening and speaking together or purely speaking (ESSAYS UK 2016). All of these challenges can affect the meaningfulness of a speaking test if constructed inappropriately. Therefore, this

essay aims to enlist a number of such criteria or principles as can help devise a suitable speaking test. To limit the focus to the topic, a number of questions have been raised. Such as:

1. What is meant by the speaking skill?
2. What are different levels of speaking?
3. What are different types of speaking?
4. Which tasks are suitable for the assessment of speaking skills?
5. How should a speaking test be scored with minimum subjectivity?
6. What are some of the general principles for the development of a speaking skill test?
7. How speaking skill is being tested in high-stake language tests?
8. What are the current practices of speaking skill assessment?

METHODOLOGY

This is, indeed, a library based research and is qualitative as well as descriptive in nature. The material for this essay has been retrieved from online sources. A brief description of these sources is given here:

Web Search Engines

Web search engines are the system software prepared to search for the information on World Wide Web. These engines present information in a line of results popularly known as Search Engine Results Pages (SERPs). There are many search engines but this essay utilises the following ones;

Google Search

It is also called Google or Google web Search. It is one of the most widely used web search engines on World Wide Web with more than three billion searches per day. For this essay, it has been utilized to search for latest and relevant information in the form of blog posts, essays, articles, books etc.

Google Scholar

Google Scholar is another web search engine which indexes metadata or full texts of schol-

arly literature across disciplines. It includes journals, books, thesis, dissertations, conference proceedings, abstract etc. It is highly esteemed among world researchers for an easy access to relevant and credible materials. Therefore, it was also utilized to seek material related with speaking comprehension test.

Databases

Databases are the collections of data in an organised form. Databases are stored and accessed with the help of a computer in an electronic form. These days the use of databases is increasing particularly the research databases have succeeded to get the attention of researchers. Realizing the significance of the databases, following database were consulted to obtain material for this study:

Foreign Language Assessment Directory (FLAD)

It is managed by Center for Applied Linguistics (CAL). It provides free and open access to the information about more than two hundred assessments carried over more than 90 languages of the world other than English. It publishes currently used information about assessment at different schools levels.

MLA International Bibliography

It is a subject index for books, book chapters, research articles, conference papers, and dissertations on drama, criticism, comparative as well as general literature, languages and linguistics. For its relevance to the field of languages and linguistics, this database has been utilised for this study.

Journal Storage

Journal Storage, popularly known as JSTOR, is another digital library. It is famous for providing back and current issues of journals, books and other primary sources. It gives access to the full texts of about 2000 world famous journals. Its use is increasing day by day in academic and research circles.

Academic Research Complete

It is the changed form of 'Academic Search Premier'. It is a tremendous source of access to a large number of research journals from a multi-disciplinary database. Such as EbscoHost which particularly covers the fields of arts and literature, biology, chemistry, education, engineering, history, physics, psychology, religion, language and linguistics.

Research Journals

To have topic-specific information from research journals, websites of Clarivate Analytics (Master-Journals List) and SJR (Scimago) were visited to find suitable journals about speaking skills assessment. But both of these databases provided not even a single journal related with speaking assessment, speaking testing, speaking evaluation etc. However, a number of journals were found which were named with 'Speech' related keywords but, they did not provide any relevant information. Therefore, some journals from the fields of general linguistics and ELT were consulted. Among them include Applied Linguistics, TESOL Quarterly, Computer Assisted Language Learning, The Cambridge Guide to Second Language Assessment, Language Testing etc. The reason for selecting journals from these two sources is that they are recognised being most credible and authentic sources of indexing the academic research journals.

Data Search

The data for this study were obtained through above web search engines, databases and research journals. For this purpose, following procedure was adopted:

Keyword Search

To search relevant information from different databases, web search engines and journals, keyword technique was utilised. For this purpose, some keywords covering the topic of the essay were determined which included some general and some specific words or terms. Among general terms included speaking, test-

ing, evaluation, assessment, tasks, scoring, rubric, assessment, testing, principles, criteria etc. Then, these terms were joined together with one another to make them specific to the topic of the study like testing speaking, speaking skill test(s), testing speaking skills, principles for devising a speaking skill test, assessing speaking skills, criteria for the assessment of speaking skills, speaking test tasks, scoring rubrics etc.

These terms and keywords were typed in the search bars of different web search engines, journals and databases in two different ways. First of all, they were typed word by word in search bars and then they were retyped adding (+) like (speaking+skill+tests+principles+criteria+ tasks+ rubrics+scoring)

Research Question Search

Other than keywords, the information was also gathered by simply typing research questions in the search bars of the mentioned databases and search engines like ‘what are the principles for a speaking skill test?’ or ‘what should be the criteria for a speaking test?’ etc.

Phrase Search

After keyword and research question search, the information was sought through typing phrases in the search bars like ‘tasks for speaking test’, ‘criteria/principles for the construction of a speaking test’ etc.

Data Processing

The said data, which comprised of a number of online books and research articles, at first, underwent the process of abstract analysis which helped filter irrelevant data. The rest of data, which was relevant to the topic of this study, was read and relevant information was extracted in a separate word file. In this process, Zotero was also utilised to access most relevant information from different articles with the help of keywords.

OBSERVATIONS AND DISCUSSION

Speaking Skill

Speaking is a real-time, aural, oral and productive skill (Bailey 2003). It is real-time because

the other interlocutor has to wait for a speaker to speak right then and the speaker cannot revise his response as he might do in writing. It is productive because the language is directed outward. It is aural because the response is interrelated with the input often received aurally and it is oral because the speech is produced orally.

Levels of Speaking

From the highest to its lowest level, speaking can be dissected into text, utterance, clause, phrase, word, morpheme and phoneme. Among these, text is the highest whereas phoneme is the lowest level of speaking (van Lier 1996). Success in speaking means being able to communicate, presenting the message accurately, and making the acceptable use of language throughout these levels. Knowing about these levels helps test maker understand what to expect from test taker’s performance.

Types of Speaking

Spoken language can be in the form of monologues or dialogues. A monologue can be planned or impromptu while a dialogue is almost always unplanned interpersonal or transactional. Each can be either familiar or unfamiliar. Brown (2004: 142-143) suggests a list of micro and macro skills of speaking to help determine test maker as what to assess (whether to assess on smaller chunks of language or on larger ones).

Micro Skills

These skills: (i) create differences between allophonic and phonemic variants; (ii) produce language chunks of varied lengths; (iii) produce stress patterns, intonation contours and rhythmic structure; (iv) produce reduced forms of phrases or words; (v) use appropriate lexical units to achieve pragmatic purposes; (vi) generate fluent speech at diverse delivery rates; (vii) monitor an individual’s own speech and use different strategic devices for example, fillers, self-corrections, pauses, and backtracking—to improve the intelligibility level of the message; (viii) use grammatical classes of words (such as verbs or nouns etc.), systems (such as pluralisation and tense agreement etc.) and word order, rules, patterns, or forms etc.; (ix) generate speech in natural con-

stituents that is, in correct pause groups, phrases, breath groups, and sentences etc.; (x) convey a specific meaning in varied forms of grammar and (xi) make use of cohesive devices in speech.

Macro Skills

These skills: (i) accomplish communicative functions appropriately in accordance with participants, goals and situations; (ii) use suitable conversation rules, implicatures, styles, redundancies, registers, pragmatic conventions, etc. in face to face conversations; (iii) communicate connections and links between different events and convey such relations as peripheral and central ideas, given information, new information, exemplification and generalization and (iv) express kinesics, body language, facial features, and other non-verbal cues with verbal language (Brown 2004).

Assessment Tasks

Brown (2004) provides five types of assessment tasks. The headings that follow are Brown's proposed categories but the tasks in each category come also from the descriptions by Heaton (1988), Hughes (1989) and O'Malley and Pierce (1996). In the past, it was agreed that speaking left no tangible product to be assessed (unlike writing). However, today technology has made it possible to record the speech in every type of the task. A challenge of this sort has little relevance to today's practice. Therefore, the following types of task may involve recording the test taker's speech.

Imitative Tasks

These tasks involve repeating a small stretch of language with a main focus on pronunciation. The test maker considers using this type of assessment if he is not interested in test taker's competence in understanding and conveying meaning or in getting involved in the interactive conversation. The competence assessed through these tasks is that of purely phonetic, prosodic, lexical and grammatical etc.

Intensive Tasks

Such type of tasks includes:

Reading Aloud

Heaton (1988) and Hughes (1989) maintain that the use of reading aloud may not be appropriate because of the difference in processing written input from that of spoken one. However, a check on stress-pattern, rhythm and pronunciation alone may be conducted using reading aloud. Brown (2004) suggests that the test taker can use reading aloud as a companion for other more communicative tasks.

Directed Response Tasks (Such as response to a Recorded Speech)

One of the most popular tasks of speaking for its practicality and mass lab use, despite its mechanical and non-communicative nature is directed response. Directed Response Task (DRT) is beneficial to draw out a specific form of grammar or transform a sentence which requires minimal processing (micro skills 1-5, 8 and 10) (Brown 2004).

Sentence or Dialogue Completion

Heaton (1988) warns that this type of task may provide illogical flow of conversation given that the sentence or dialogue completion is done in lab (which is what normally administered). Therefore, this type will probably be beneficial only for assessing test taker's micro skill of providing the right chunks of language and other pronunciation features. However, as Brown (2004) exemplifies a more responsive-type of sentence or dialogue completion (SDC) which may actually be free of the said limitation and keep us away from the risk of judging a test taker's competence as insufficient caused by aural misunderstanding in processing the input. SDC helps measure speaking competence apart from its interrelatedness to listening.

Translation up to Simple Sentence Level (Interpreting-game)

Interpreting, as Hughes (1989) describes, may involve the test conductor acting as a native speaker of the test taker's first language and the test taker interpreting the utterance into English. It is believed that because speaking is

negotiation of intended meaning (O'Malley and Pierce 1996), interpreting-game can be used to measure test taker's competence in conveying his message into the target language (Brown 2004).

Picture-cued Task (Including Simple Sequence)

Pictures are mostly convenient to elicit description (Hughes 1989). In addition to describing comparison, order of events, positions and location, a more detailed picture may be used to elicit test taker's competence in telling a plan, directions and even opinions (Brown 2004).

Responsive Tasks

Responsive tasks involve small dialogues as well as responses to spoken prompts like simple greetings, requests and comments etc. Other popular responsive tasks are described here:

Questioning and Answering

Questions, at responsive level, appear most likely to be referential (as opposed to intensive, display question). Referential question requires test takers to produce meaningful language in response. Such questions may require an open-ended response or a counter-question directed to the interviewer (Brown 2004).

Paraphrasing

Oral paraphrasing can have written or aural input with the later being more preferable. A paraphrase, like speaking assessment, should be conducted with caution because test taker's competence may be mistakenly judged by their short-term memory and listening comprehension instead of their speaking production.

Interactive Tasks

These tasks are based on larger dialogues, on transactional and interactional conversation. Such type of tasks includes:

Interview

Interview can be face-to-face, one-on-one or two-on-one each with its advantages and dis-

advantages. A two-on-one interview may save time and scheduling and provide authentic interaction between two test takers, although it can pose a risk of one test taker's domination over the other. Canale (1984) proposes four main steps to follow to conduct an oral proficiency test: (1) Warm Up: small talk about identity, origin and the like; (2) Level-Check: description without pause, *wh*-questions, reading passages aloud, briefly guided role-plays, and telling how to do or make something etc.; (3) Probe: field-related questions and (4) Wind-down: easier questions pertaining to test taker's feeling about the interview

In addition to Canale's (1984) proposal, Hughes (1989) proposes 11 rules to conduct an interview: (1) prepare spoken tests as long as feasible, (2) incorporate as large a sample of specified substance as is possible in available time, (3) map the test with great care, (4) provide the testees with possible maximum 'fresh starts', (5) choose interviewers with care and train them as well, (6) get help from a second tester, (7) select such topics and tasks as pose no difficulty for the candidates in their L1, (8) arrange interview in a peaceful room, (9) collect as much information as possible and (10) do not speak too much (the interviewer).

The challenge with an interview is how the open-ended response is scored. Creating a consistent, workable scoring system to ensure reliability has been one of the major challenges in designing an interview as means to assess speaking (Brown 2004). There are at least two solutions to this problem: one is using an analytical scoring rubric and the other is a holistic one. Rescoring the performance later from the tape can be an alternative, too (O'Malley and Pierce 1996).

Drama Tasks

O'Malley and Pierce (1996) divides drama-like tasks into three sub-types, such as; (i) improvisations, (ii) role play and (iii) simulation. The difference of each is respectively the preparation and scripting. Improvisation provides very little opportunity to the test taker to prepare the situation and incite creativity in using the language. Role play provides slightly longer

time and the test taker can prepare what to say. Meanwhile, simulation (including debate) requires planning and decision making. Simulation may involve real-world socio-drama which is the pinnacle of speaking competence. Like interview, drama-like task may cause and unpredictable response. Therefore, same care taken to tackle interview may be useful for this type of task as well.

Discussion and Conversation

Discussions and conversations (Brown 2004) provide somewhat similar difficulties in terms of predictability of the response hence, consistency of the scoring to that of interview and drama-like tasks. Test makers seem to choose this type of task as an informal assessment to elicit and observe test taker's performance in: (1) starting, maintaining and ending a topic, (2) getting attention, interrupting and controlling, (3) clarifying, questioning and paraphrasing, (4) signaling for comprehension (such as nodding), (5) using appropriate intonation patterns, (6) using kinesics, eye contact and body language and (7) being polite, being formal and other sociolinguistic situation.

Games

It is almost impossible to list all games, but virtually all games that can elicit spoken language objectively can be used as informal assessment tasks for speaking. Brown (2004) warns us that using games may go beyond assessment and adds that a certain perspective needs to be maintained in order to keep it in line with assessment principles. Some examples of games which Brown (2004) mentions (tinker toy, crossword puzzle, information gap, predetermined direction map) can all fall in the umbrella of information-gap activities by O'Malley and Pierce's (1996) standpoint as he explains that information gaps are the activities where a student is provided with information that another (such as his pair) does not know but needs to know. An information gap activity involves collecting complete information to restructure a building, sequence a picture into order or simply find the differences between two pictures. To score an information gap activity, O'Malley and Pierce

(1996: 83) suggests the test makers to consider the speaker's "accuracy and clarity of the description as well as on the reconstruction."

Extensive Tasks (Monologue)

The following are monologues which take longer stretches of the language and require extensive (multi-skills) preparations. The terms are self-explanatory and some may actually possess some characteristics with some types previously explained only with longer and broader scope of language use.

Speech (Oral Report/Presentation)

It is a common practice to present papers, reports, or models in educational settings. Therefore, oral presentations, reports can be utilised to evaluate the speaking skill analytically or holistically. But it must be kept in mind that it is best for the advanced or intermediate levels focusing on delivery and contents (Brown 2004).

Picture-cued Story Telling

Similar to the limited version, at this level the main concern of using pictures or the series of pictures is to make it into a stimulus for the longer stories. For this purpose, a six-picture series with an adequate amount of details about the character and settings will be adequate to test (Brown 2004).

Retelling a Story, News Event

Different from paraphrasing, retelling a story takes a longer stretch of discourse with different, preferably narrative genre. The focus is usually on meaningfulness of the relationship of events within the story, fluency and interaction to the audience (Brown 2004).

Translation (Extended Prose)

In this type of task, a longer text preferably in written form, which is presented in the test taker's native language, is to be studied prior to interpreting the text with ease in the actual testing. The text can cover a dialogue, procedure, complex directions, synopsis or a play script.

Caution should be made concerning with this type of task because this particular type requires a skill not intended for every speaker of a language. Therefore, if this type is to be used, a degree of confidence should be made sure (Brown 2004).

Theory of Speaking Assessment

All of above discussed points somehow or the other can help the construction of speaking assessment tasks for particular speaking skills. However, so far as the preparation of a speaking test is concerned, speaking assessment theory by Canale and Swain (1980) can prove more helpful. This theory concentrates on four competencies regarding the ability of speaking. Among these competencies include: (1) Discourse competence: it involves coherence and cohesion, (2) Grammatical competence: it involves the knowledge of vocabulary, grammar, or the mechanics of language, (3) Strategic competence: it focuses on the ability to use the language strategies properly and, (4) Socio-linguistic competence: it involves the social as well as cultural knowledge of the target language users.

General Principles

COE (2001) and O'Sullivan and Nakatsuhara (2011) propose to pay attention on following general principles for the construction of speaking skills test: (a) what is to be tested (content and construct principle); (b) provide the testees with such a range of tasks as could help them perform their best (test method); (c) indicate how to rate/score the performance (rating criteria); (d) explain how the results will be interpreted (score interpretation criteria); (e) report scores for each task separately (test results reporting) and (f) develop locally sensitive instruments (practicality)

Scoring Rubrics

O'Malley and Pierce (1996) propose a number of steps for devising the rubric: (1) set criteria for the success of the task, (2) set language dimensions for assessment (such as fluency, grammar, pronunciation, vocabulary etc.), (3) give proper weight to all dimensions (if possi-

ble, provide omissions) and (4) focus on what the test takers *can* do, instead of what they *cannot* do. Which rubric is better? Whichever is used, if high accuracy is the goal, multiple scoring is required (Hughes 1989). Since a test taker's speech can now be recorded for second-time scoring by a different rater, a balance between holistic and analytical rubric (that is, use two types of rubric for the same task whenever possible) is recommended.

An effective assessment, in the view of Brown (2004), should follow this rule: (1) particular criteria, (2) suitable task, (3) obtaining best possible output and (4) reliable and practical scoring measures. Scoring remains the major challenge in assessment. There are two different types of known scoring rubric for speaking: (1) holistic and (2) analytical. A holistic rubric range, for example, from 1 to 6 each reflecting unique capacity of the speaker with 6 being normally native-like traits and 1 a total misuse of language which incite misunderstanding. But the analytical rubrics score testees' performance in various subcategories such as comprehension, fluency, pronunciation, grammar and task completion vocabulary. There are two familiar practices concerning the later: (1) full amount of score is summed in average to show an overall score or (2) each category is given a different value sometimes without summing up the total score.

Ensuring Objectivity in Scoring

For, scoring process involves the judgment of a person therefore, the assessment of speaking, no doubt, can be subjective (Rychtarik 2014). A difference between subjective and objective assessment lies in the fact that, "objective assessment is assessment from which subjectivity is removed" (Council of Europe 2001: 188). Therefore, possible measures should be taken to minimise the element of subjectivity. In this regard, following measures can be more helpful if taken properly: (a) there is a general principle that the more the criteria, the greater the chances of achieving objectivity in scoring. Therefore, the tester should establish a set of criteria (Rychtarik 2014); (b) if the test deviser cannot develop or establish a set of criteria then he can follow the criteria proposed by some experts. There are many such criteria. For example, 14 qualitative categories' criteria proposed by the

Council of Europe (2001) can be more helpful. The 14 qualitative categories' criteria include turn taking strategies, fluency, precision, thematic development, vocabulary range and control, grammatical accuracy, phonological accuracy etc. however, it is not possible for an assessor to apply all criteria at a time. Applying more than 7 criteria can cause cognitive overload on him. Therefore, the Council of Europe (2001) suggests that the application of 4-5 criteria is enough for the feasibility as well as reliability of the assessment. Brown (2007: 352) has merged some of the criteria to be followed easily such as pronunciation, fluency, vocabulary, grammar, discourse features (socio-linguistic appropriateness, cohesion etc.) and task accomplishment; (c) another way to minimise objectivity in the scoring process, according to (Rychtarik 2014), can be achieved by breaking the speaking skills into sub-skills. In this regard, micro and macro-skills by Brown (2004), as given above, may well be followed.

Current Practices of Testing Speaking Skills

Speaking test makers should be well aware of the current practices in speaking skill assessment regarding time distribution, tasks, scoring procedure, etc. For this purpose, they can see the practices in standard tests like IELTS, TOEFL and TOEIC (see for example, Cullen et al. 2014; ETS 2017; Lougheed 2012).

Current Trends

Computer Assisted Language Learning (CALL) is recognised as a valuable source for learning a foreign language. The assessment of speaking skill through CALL procedure has also emerged as a current trend. A number of studies by Davidson and Coombe (2012), Luoma (2004), Neri et al. (2000), Pennington (1999) have also highlighted the significance of the assessment of speaking skill through CALL procedure.

Davidson and Coombe (2012) particularly emphasise that the significance of web-based and computerised testing should not be ignored in this age of information technology. They suggest computer-based language testing (CBLT). For, such type of testing can help prepare reliable as well as valid tests that can measure the speaking skills more accurately. The same is true

for Web-based language testing (WBLT). According to Shin (2012), WBLTs are even more beneficial than CBLTs regarding practice and quality for example, WBLTs have the ability to enhance various aspects of authenticity including interactional as well as situational authenticity plus, they provide consistent scores regarding common construct features.

CONCLUSION

The key of assessing speaking skill is an understanding of the continuum of (1) spoken language, (2) task types and (3) scoring rubric. This non-rigid separation between one level of competence and another requires time and effort in specifying the criteria of speaking, task to elicit particular behavior and in developing practical and representative scoring rubric. The variety of task types will help test maker to decide which one is appropriate for the wide range of the continuum of this particular skill.

In fact, success in speaking depends upon the communication of message accurately through different levels ranging from highest to lowest levels (such as text, utterance, clause, phrase, word, morpheme and phoneme). Therefore, it is imperative for a speaking test deviser to have a good knowledge of these levels. Furthermore, speaking is of two types that is, (1) monologue and (2) dialogue. The former can be planned or unplanned whereas the later is always planned. Any of these two types can be utilised in a speaking test but in the light of basic principles. Similarly, speaking is divided into micro as well as macro skills. Such type of division can be useful for a speaking test in a number of ways that is, (1) they help the test makers determine about what to assess (for example, whether to assess on smaller chunks of language or on the larger ones). In addition, there are five types of assessment tasks including imitative, intensive, responsive, interactive and extensive. All of these tasks can also be utilised to in a speaking skill test. But it should be kept in view that all of these tasks involve different items which are applicable differently at different levels. Therefore, it is essential for a speaking skill test deviser to have a basic knowledge of different levels, tasks, sub-tasks and the application criteria of these tasks.

So far as the scoring of a speaking test is concerned, it is advisable for a test deviser/evaluator to have an established set of rubrics. Experts have proposed different sets of rubrics which a speaking test examiner can easily use for his help. However, he can develop his own rubrics if he deems suitable.

To observe the current fashion in speaking skill tests, a test deviser can see IELTS, TOEFL and TOEIC. From these tests, he can seek guidance regarding selection of tasks and items, time allocation, scoring procedure, etc. Some experts have started testing speaking skill with the help of technology. In this regard, CBLTs and WBLTs are emerging popularly in the field of language testing. These tests are more valid and reliable as compared to the typical tests.

RECOMMENDATIONS

The study suggests to follow the specific criteria for the assessment of speaking skill. For this purpose, the test practices of IELTS, TOEFL and TOEIC tests can be taken as a model. Moreover, CBLTs and WBLTs are latest as well as technology oriented and can also be utilised for better results.

REFERENCES

- Bailey KM 2003. Speaking. In: D Nunan (Ed.): *Practical English Language Teaching*. New York: McGraw-Hill Companies, pp. 47-66.
- Bailey KM, Nunan D 2004. *Practical English Language Teaching: Speaking*. New York: McGraw-Hill Companies.
- Bellack AS, Hersen M, Lamparski D 1979. Role-play tests for assessing social skills: Are they valid? Are they useful? *Journal of Consulting and Clinical Psychology*, 47(2): 335-342.
- Brown HD 2004. *Language Assessment: Principles and Classroom Practices*. White Plains, NY: Pearson Education.
- Brown HD 2007. *Teaching by Principles: An Interactive Approach to Language Pedagogy*. 3rd Edition. White Plains, NY: Pearson Publications.
- Butler FA, Stevens R 1997. Oral languages assessment in the classroom. *Theory into Practice*, 36(4): 214-219.
- Canale M, Swain M 1980. Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1): 1-47.
- Canale M 2014. From communicative competence to communicative language pedagogy. In: JC Richards, RW Schmidt (Eds.): *Language and Communication*. New York: Routledge, pp. 14-40.
- Celce-Murcia M, Brinton DM, Goodwin JM 1996. *Teaching Pronunciation: A Reference for Teachers of English to Speakers of Other Languages*. Cambridge: Cambridge University Press.
- Chou MH 2018. Speaking anxiety and strategy use for learning English as a foreign language in full and partial English medium instruction contexts. *TESOL Quarterly*, 52(3): 611-633.
- Clapham C, Corson D 1997. *Encyclopedia of Language and Education: Volume 7: Language Testing and Assessment*. Dordrecht, Netherlands: Kluwer Academic.
- Correia RC 2016. Assessing speaking proficiency: A challenge for the Portuguese EFL teacher. *e-TEALS*, 7(1): 87-107.
- Council of Europe 2001. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.
- Cullen P, French A, Jakeman V 2014. *The official Cambridge guide to IELTS*. Cambridge: Cambridge University Press. From <<https://quicklylearningenglish.files.wordpress.com/2016/05/cambridgeielts10.pdf>> (Retrieved on 15 June 2018).
- Dandonoli P, Henning G 1990. An investigation of the construct validity of the ACTFL proficiency guidelines and oral interview procedure. *Foreign Language Annals*, 23(1): 11-22.
- Davidson P, Coombe C 2012. Computerized language assessment. In: C Coombe, P Davidson, B O'Sullivan, S Stoyhoff (Eds.): *The Cambridge Guide to Second Language Assessment*. New York, NY: Cambridge University Press, pp. 267-273.
- Essays UK 2016. Testing of speaking skills issues and solution English language essay. From <<https://www.ukessays.com/essays/english-language/testing-of-speaking-skills-issues-and-solution-english-language-essay.php?vref=1>> (Retrieved on 29 December 2018).
- ETS 2017. *The Official Guide to the TOEFL*. New York: McGraw Hill Education.
- Grabe W 2002. Applied linguistics: An emerging discipline for the twenty-first century. In: RB Kaplan (Ed.): *The Oxford Handbook of Applied Linguistics*. Oxford: Oxford University Press, pp. 3-12.
- Guettal I 2008. *Introducing Storytelling Activities as a Basic Technique to Develop Students' Oral Performance an Experimental Approach Case Study: Second Year Students of English at Batna University*. Doctoral Dissertation. Algeria: Universite Mohamed Khider.
- Harding L 2014. Communicative language testing: Current issues and future research. *Language Assessment Quarterly*, 11(2): 186-197.
- Heaton JB 1988. *Writing English Language Tests*. London: Longman.
- Hughes A 1989. *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- Idrissova M, Smagulova B, Tussupbekova M 2015. Improving listening and speaking skills in mixed level groups (on the material of New English File). *Procedia-Social and Behavioral Sciences*, 199: 276-284.

- Inbar-Lourie O 2008. Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25(3): 385-402.
- Kitao SK, Kitao K 1996. Testing speaking. *Report No. TM025215*. (ERIC Document Reproduction Service No. ED398261).
- Knight B 1992. Assessing speaking skills: A workshop for teacher development. *ELT Journal*, 46(3): 294-302.
- Levitt WJ 1994. The skill of speaking. *International Perspectives on Psychological Science*, 1: 89-103.
- Lougheed L 2012. *Longman Preparation Series for the New TOEIC Test*. New York: Longman.
- Luoma S 2004. *Assessing Speaking*. New York: Cambridge University Press.
- Lynch BK, McNamara TF 1998. Using G-theory and many-facet Rasch measurement in the development of performance assessments of the ESL speaking skills of immigrants. *Language Testing*, 15(2): 158-180.
- Morrow CK 2018. Communicative language testing. In: JI Liantas (Ed.): *The TESOL Encyclopedia of English Language Teaching*. United States, NJ: Wiley, pp. 342-350.
- Nakamura Y, Valens M 2001. Teaching and testing oral communication skills. *Journal of Humanities and Natural Sciences*, 3: 43-53.
- Nation IS, Newton J 2008. *Teaching ESL/EFL Listening and Speaking*. New York: Routledge.
- Neri A, Cucchiari C, Strik H, Boves L 2002. The pedagogy-technology interface in computer assisted pronunciation training. *Computer Assisted Language Learning*, 15(5): 441-467.
- O'Malley JM, Pierce LV 1996. *Authentic Assessment for English Language Learner: Practical Approaches for Teachers*. White Plains, NY: Addison Wesley.
- O'Sullivan B, Nakatsuhara F 2011. *Quantifying Conversational Styles in Group Oral Test Discourse*. Basingstoke: Palgrave MacMillan.
- Pennington MC 1999. Computer-aided pronunciation pedagogy: Promise, limitations, directions. *Computer Assisted Language Learning*, 12(5): 427-440.
- Riggio RE 1986. Assessment of basic social skills. *Journal of Personality and Social Psychology*, 51(3): 649-660.
- Rychtarik S 2014. Designing criteria to assess speaking skills. Doctoral Dissertation. From <https://www.academia.edu/7719276/Designing_Criteria_to_Assess_Speaking_Skills> (Retrieved on 28 December 2018).
- Schmidt R 1992. Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition*, 14(4): 357-385.
- Sercu L 2004. Assessing intercultural competence: A framework for systematic test development in foreign language education and beyond. *Intercultural Education*, 15(1): 73-89.
- Shin SY 2012. Web-based language testing. In: C Coombe, P Davidson, B O'Sullivan, S Stoyhoff (Eds.): *The Cambridge Guide to Second Language Assessment*. Cambridge: Cambridge University Press, pp. 274-279.
- Ur P 2008. *A Course in Language Teaching*. Stuttgart/Germany: Ernst Klett Sprachen.
- van Lier L 1996. *Interaction in the Language Curriculum: Awareness, Autonomy and Authenticity*. London: Longman.

Paper received for publication in January, 2019
Paper accepted for publication in February, 2019