

## An Assessment of IELTS Speaking Test

Shahzad Karim, Naushaba Haq

Department of English, The Islamia University of Bahawalpur, Pakistan

---

### Article Info

#### Article history:

Received April 04, 2014

Revised Jul 15, 2014

Accepted Aug 24, 2014

---

#### Keyword:

Assessment

Evaluation

IELTS

Speaking

---

### ABSTRACT

The present study focused on assessing the speaking test of IELTS. The assessment discussed both positive aspects and weaknesses in IELTS speaking module. The researchers had also suggested some possible measures for the improvement in IELTS speaking test and increasing its validity and reliability. The researchers had analysed and assessed IELTS speaking test in the light of both theoretical and practical perspectives presented by experienced researchers in the field of language testing and evaluation. The researchers' major concern in the assessment of IELTS speaking test was to do utmost effort to avoid the element of subjectivity and to present some logical and practical suggestions for improving IELTS speaking test.

Copyright © 2014 Institute of Advanced Engineering and Science.  
All rights reserved.

---

### Corresponding Author:

Naushaba Haq,

Department of English,

The Islamia University of Bahawalpur,

Pakistan.

Email: lsntl@ccu.edu.tw

---

## 1. INTRODUCTION

Speaking is a productive skill. From testing point of view, it is special because it is interactive in nature and has to be measured directly in live interaction. The basic purpose of developing speaking skill is to interact successfully in that particular language and it involves comprehension as well as production. Speaking test has been a part and parcel of world wide large scale language proficiency tests like IELTS, TOEFL, and Cambridge exams like FCE and CAE. However, the present study aims at assessing IELTS Speaking Test only.

Speaking test is the last of the four tests in IELTS. It consists of a face to face interview between the candidate and an IELTS trained examiner. The interview lasts for 11 to 15 minutes and is recorded on an audio-cassette. The test is divided into three phases.

- **Phase 1** is introduction which is carried out in a series of short questions and answers in order to make the candidate comfortable and to develop some familiarity with the candidate. The examiner asks very simple questions about candidate's own self like his/her home, family, country, work, study, interests etc. For example: "Why did you decide to study Engineering?" "What are some of the most popular drinks in your country?"
- **Phase 2** is an individual long turn where the candidate has to speak on a selected topic for 2 to 3 minutes. Each candidate is given a topic and he/she has to talk about it in the form of a monologue in limited time i.e. 2 to 3 minutes. The object or topic to be described is general in nature like a river, beach or a film etc.
- **Phase 3** comprises of a two way discussion or dialogue between the candidate and the interviewer. It is thematically linked to the topic of the long turn i.e. phase 2.

## 2. METHODOLOGY

The focus of the present study is on assessing the speaking test of IELTS. Many researchers have proposed various aspects and ways of assessing the oral ability. However, Hughes' (2003) criteria for assessing oral ability seem appropriate in assessing the oral ability. Hughes (2003) emphasises following steps in assessing the oral ability [1].

- To set an appropriate task to elicit representative sample of the population.
- To ensure validity and reliability of elicited sample and its scoring.

Hence, keeping in view the comprehensive approach of Hughes' (2003) prescribed criterion, the researchers have decided to follow his steps with some variation in order to assess IELTS speaking module [1]. As in IELTS interview is used as a tool for eliciting sample of speaking, so firstly the appropriateness of interview as a tool for eliciting representative sample will be assessed and then it will be followed by the assessment of validity, reliability and practicality of the IELTS speaking test.

## 3. RESULTS AND ANALYSIS

### 3.1. Assessment of the Appropriateness of Interview as a Sample Eliciting Tool

Though interview is the most widely used task for testing speaking skill, yet it has some drawbacks as well. Here we will discuss interview in the context of IELTS speaking test. In IELTS, interview is used in its traditional form which has one serious drawback i.e. in such interviews the interviewer remains dominant because he is responsible for taking all the initiatives, while the candidate or interviewee has just to respond to the questions asked to him. Thus, in this way only one style of speech is elicited and many aspects of speaking like asking question and taking initiatives to start a discussion remain hidden. Hughes (2003:119) discussed this idea in the following words [1].

“The relationship between the tester and the candidate is usually such that the candidate speaks as to a superior and is unwilling to take the initiative. As a result, only one style of speech is elicited, and many functions (such as asking for information) are not represented in candidate's performance.”

So, in each phase, the candidate should be given the opportunity to ask questions. It will not only help the candidate in building up his confidence, put him into ease, but will also help the interviewer in assessing candidate's questioning skills. Moreover, it will also help the candidate to get clarification to avoid going astray during the course of the interview and be more focussed.

Another drawback of the IELTS interview is its formal context. In real life situation, mostly, we have to speak in informal context. As the requirements of speaking skill vary in both formal and informal context, the formal context of IELTS interview may not elicit and analyze speaking skill in its true sense. Moreover, the controlled conditions during the interview do not allow interviewee to speak as freely as one speaks in real life. Thus, the information elicited cannot be true representative of real life speaking skills.

In real life we have to speak in different situations and contexts and our language varies according to those different contexts. In interview, the use of language in those different contexts cannot be assessed as it can be assessed through role play tasks. Hughes (2003) conforms to this idea by saying: “In my experience, however, where the aim is to elicit 'natural' language and attempt has been made to get the candidates to forget, to some extent at least, that they are being tested, role play can destroy this illusion.” (p. 120) [1]. So instead of asking the candidate to speak in the form of a monologue, it is better to let him/her speak through some role play activity which is more relevant to real-life situations.

Moreover, in real life, ideas are not well formed in mind. They have to be generated immediately and quick responses are required. Whereas in IELTS, especially in the second part i.e. of individual long turn the candidate is given some time to formulate ideas, even spare paper and pencil are provided to jot down the ideas which, normally does not happen in real life. These aspects of IELTS speaking test seem a bit unnatural. Hence, it is suggested that it should be made more natural and close to real life situations.

### 3.2. Assessment of Validity

The validity of a test can be judged by considering “does the test test what it is supposed to test?” [2]. In order to have a better idea of the validity of the IELTS speaking module, we may investigate it under its sub-categories like content validity, face validity and criterion validity. But before discussing it in accordance with the above mentioned categories, we may have a brief description of what validity is. According to Hughes (2003) a test is said to be valid if it measures accurately what it is supposed to measure [1]. However, Hennig (1987), Bachman (1990) and Messick (1995) are of the opinion that validity is relative and it depends upon the purpose of the test. A test cannot be completely valid. It may be valid for one purpose but not for another [3]-[5]. Messick (1996) considers validity as an integral and unified concept [6].

But for the sake of convenience and to analyze it thoroughly that no major source of validity should remain hidden, the validity of the IELTS speaking test is being assessed here through type by type.

#### a) Content Validity

A test is said to have content validity if its contents consist of items which can elicit the representative sample of that particular skill. The importance of content validity lies in the fact that the accuracy of measurement of a certain skill depends upon the accuracy of the content validity. Hughes (2003) elaborates that the contents of a test should not be based on what is easy to test rather what is important to test [1]. For example a test for postgraduate level learners should not contain the same set of items and structures as for undergraduate level learners. IELTS speaking test has same structure and content for the learners of all levels and no consideration is paid to their educational background or age. Thus, the content validity of IELTS speaking test may be questioned.

Another basic consideration of content validity is that the language sample collected in a short period of time of the test should be representative of the language used in real-life situation as Hasselgreen (2004: 12) says [7]:

“The sample of language collected in the short space of test-time is somehow representative of the language of real-life communication, and relevant to the specified domain. This representativeness is evaluated in the process of content validation, with respect not only to linguistic forms but also to the functions and conditions of speaking.”

The IELTS speaking test does not fulfil this criterion of content validity as the interview cannot represent the use of spoken language in real-life situations. The interview usually tends to be more formal and unnatural.

#### b) Face validity

According to Hughes (2003: 33) “a test is said to have face validity if it looks as if it measures what it is supposed to measure” [1]. Hasselgreen (2004: 14) mentions two important factors which may affect the face validity of a test [7]. The two factors are:

- Unfamiliarity of format
- Lack of authenticity in test task

If we evaluate IELTS speaking test for face validity, it can be said that IELTS fulfils the criterion of face validity as its format is quite clear and well established. Besides, many sources like books, research reports and websites are available which provide not only suitable guidelines about format but also provide helping materials to the candidates.

#### c) Criterion-related validity

There are two kinds of criterion-related validity.

- Concurrent validity
- Predictive validity

IELTS speaking test may not fulfil concurrent validity as it consists of just a short 11 to 14 minutes interview in which all aspects of speaking skill may not be assessed as they can be assessed in role-play tasks, oral presentations or picture cued tasks. Thus, the speaking skill elicited from interview may not be the representative of overall speaking ability in all contexts of real life.

Predictive validity “concerns the degree to which a test can predict candidates’ future performance” [1]. In both tests of IELTS i.e. IELTS general and IELTS for academic purpose speaking test is the same. There is no change in the speaking test with reference to the change of context of the two. IELTS speaking test may have better predictive validity in general context as the way it is administered, it may assess general speaking ability in a better way compared to the speaking in academic context because the requirements of speaking in general are quite different from that of different subject specific academic context.

### 3.3. Assessment of Reliability

The reliability of a test is determined by the consistency of its marks as remarked by Hughes (2003: 36) “The more similar the scores would have been, the more reliable the test is said to be.” This similarity and consistency of scores depends upon two factors [1].

- Raters’ grading
- Test conditions

We may discuss these two factors one by one with reference to the IELTS speaking module.

a) Raters' grading

Reliability based on raters' grading is of two types, inter-rater reliability and intra-rater reliability. These two types are discussed by Hasselgreen (2004: 21) in the following words [7]:

“Inter-rater reliability is the extent to which different raters are able to agree on the same performances, while intra-rater reliability is the extent to which the same rater would (hypothetically) be consistent if applying the same criteria to the same performance repeatedly.”

In IELTS speaking module, inter-rater reliability may be affected because the oral ability of the candidate is assessed by a single rater. Moreover, the grading is done on the basis of vague, holistic band scale in which there is general division of bands on the basis of categories like fluency, grammatical accuracy, coherence and pronunciation, but no specific marks are allocated to each category which may result into marking according to the preference of the rater. Thus, in order to remove the doubt of subjectivity test should be scored by two independent raters who should not know how each one of them has scored the test.

The impact of the interviewer differences on the result and final score of the test should be seriously taken into account in a rating process because a candidate's reported proficiency level is not only his/her inherent ability but also depends upon interviewer's variability and subjectivity. For example, some raters treat 'fillers' as positive because of its native-like speech style; whereas others may consider it as a reflection of limited vocabulary. Similarly, some assessors consider 'disfluency' as a native-like speech style because many times in real-life situations the native speakers tend to pause in their speech especially when they speak while deeply thinking. On the other hand, some assessors may think of 'disfluency' as a drawback. Brown and Hill (2007: 55) also say that there are generally two types of interviewers: 'the difficult interviewers and the easy interviewers' [8]. The former ones even induce complex skills of speculating and justifying opinions while assessing the candidates' speaking skill. They sometimes tend to argue and interrupt candidates with another question even before they complete their response to the previous question. In contrast, the latter ones normally use simple and economical questions and do not bother the candidates with argumentative questions. They normally ask open-ended questions, show scaffolding behaviour and make questions understandable [9]. Hence, some element of unfairness is evident in the latter ones even though they seem cooperative with the candidates because the candidates with assistance tend to perform better than the ones without assistance. So different type of interviewers cause different problems for the candidates through which the candidates can be either advantaged or disadvantaged by the 'luck of draw' in interview allocation. Therefore, in my opinion, both types of interviewers should be present as examiners for each candidate.

b) Test Conditions

The test conditions like partner compatibility, physical environment and test procedure also play a vital role to ensure test-reliability.

- In IELTS the condition of partner compatibility is not fulfilled because, in it, the interviewer remains dominant and is responsible for taking all the initiatives.
- Considering the aspect of test procedure, it has also been noticed that the use of just one format i.e. interview to assess the speaking skill of the candidate may not work well as someone may not feel comfortable in formal and somewhat restricted context of the interview and may not perform well. While the same candidate may perform well in some other item like role-play or other tasks used to elicit language sample. So some additional task should be used to elicit reliable data as Hughes (2003: 44) suggests that “the addition of further items will make a test more reliable” [1]. Moreover, he suggests that the other item should be different from the previous one so that more information should be gained. This additional information makes results more reliable.

### 3.4. Practicality

Another important aspect in testing is the practicality and efficiency of the test. If a test is not practical, it will be of no use even though it is reliable and valid. Weir (1993) mentions that practicality involves questions of economy, ease of administration, scoring and interpretation of results [10]. Considering all these aspects IELTS seems to be highly practical as it does not take much time and is easy to administer. Moreover, it also reduces the fatigue factor on the candidate.

## 4. DISCUSSION

The study focused on assessing the IELTS speaking test. Hughes' (2003) criteria were followed to assess and evaluate the IELTS speaking test [1]. The IELTS speaking was assessed in two steps. Firstly, the

appropriateness of interview as a tool for eliciting representative sample was assessed. The second step consisted of assessing the validity, reliability and practicality of the IELTS speaking test.

While assessing the interview as a speaking data eliciting tool in IELTS, it was found that the role of the interviewer remains dominant and the interviewee has to respond only to the questions asked by the interviewer. Hence, it elicits only one aspect of speaking. Other aspects of speaking like asking questions and taking initiatives to start discussion remain dormant. This is in accordance with what Hughes (2003) points out as a weakness in a speaking test [1]. Another weakness of the IELTS speaking test is its formal context only. In daily life we have to speak mostly in informal context, but the IELTS speaking test does not test the speaking skills in informal contexts. Moreover, in real life, ideas are not well formed in mind. They have to be generated immediately and quick responses are required. But in IELTS, especially in its second part, the candidate is given enough time to formulate his/her ideas. This is not in accordance with the real life speaking skills. Hence, the assessment of speaking skills in IELTS can be said a bit unnatural.

Alderson (1995:170) says that the validity of a test is judged by considering “does the test test what it is supposed to test?” In order to have a better idea about the validity of the IELTS speaking test, it was assessed by dividing it into sub-categories like content validity, face validity and criterion-related validity [2]. The content validity of the IELTS speaking test may be questioned because it has the same content for the learners of all levels without bringing into consideration their educational background and age. The content validity of the IELTS speaking test can also be questioned on the grounds that the IELTS interview cannot represent the use of spoken language in real-life situations. This is what Hasselgreen (2004) says that the language sample collected in short period of time of the test should be representative of the language used in real-life situations [7]. So far as the face validity is concerned, Hasselgreen (2004) mentions two important factors which may affect the face validity of a test [7]. The two factors are unfamiliarity of format and lack of authenticity in task. The IELTS speaking test fulfils the criterion of face validity. Its format is quite clear and well-established. Besides, many source like books, research reports and sample tests are available which provide enough guideline about the format of the IELTS speaking test. Criterion-related validity has two aspects: concurrent validity and predictive validity. The IELTS speaking test may not fulfil the concurrent validity because it consists of just 11 to 14 minutes interview in which all aspects of speaking skills may not be assessed. IELTS predictive validity may also be questioned because it may assess general speaking ability in a better way compared to speaking in an academic context because the requirements of speaking in general are quite different from that of different subject specific academic contexts.

Reliability means consistency in scores and the consistency of scores depends upon two factors: raters’ grading and test conditions [1]. Further, the raters’ grading is of two types: inter-rater reliability and intra-rater reliability. In IELTS speaking module, inter-rater reliability may be affected because the speaking skill of a candidate is assessed by only one rater. Moreover, the rating is done on the basis of a holistic band scale. In matter of test conditions, the IELTS speaking test does not fulfil the condition of partner compatibility because in it the interviewer remains dominant and is responsible for taking all initiatives. Considering the aspect of test procedure, the use of only interview to assess the speaking skills of the candidate may not work well as some candidates may not feel comfortable in formal and somewhat restricted context of the interview. Hughes (2003: 44) rightly suggests that “the addition of further items will make the test more reliable” [1]. In matter of practicality, the IELTS speaking test can be said highly practical because it seems to fulfil the principles of economy, ease of administration, scoring and interpretation of results.

## 5. SUGGESTIONS

Keeping in view the above mentioned discussion, the following suggestions are presented for bringing improvements in the IELTS speaking test and making it more reliable and valid.

- Time frame (11 to 15 minutes) is less to assess the oral ability of a non-native speaker. If the candidate wants to expand the topic and asks supplementary questions he/she should be encouraged. It will not only be helpful to elicit more authentic data but will also provide opportunity to the rater to assess a candidate’s questioning skill which is an important aspect of speaking skill.
- A single task i.e. interview is not sufficient to elicit the required data. At least one more task like role play or picture cued task should also be introduced.
- There should be more than one examiner. It will not only increase the reliability of assessment but will also reduce entire responsibility from a single rater. Further, it will also help to make the discussion more informal and will reduce pressure on the candidate.

There should also be some variation in grading scale considering the age factor and educational background of the candidate.

**REFERENCES**

- [1] Hughes, A., "Testing for Language Teachers", Cambridge: Cambridge University Press, 2003.
- [2] Alderson, J. C. Clapham, C. & Wall, D., "Language Test Construction and Evaluation", Cambridge: Cambridge University Press, 1995.
- [3] Henning, G., "A Guide to Language Testing: Development, Evaluation, Research", Rowley: Newbury House, 1987.
- [4] Bachman, L.F., "Fundamental Considerations in Language Testing", Oxford: Oxford University Press, 1990.
- [5] Messick, S., "Validity of Psychological Assessment", *American Psychologist*, vol/issue: 50(9), pp. 741-749, 1995.
- [6] Messick, S., "Validity and Washback in Language Testing", *Language Testing*, vol/issue: 13(3), pp. 241-256, 1996.
- [7] Hesselgreen, A., "Studies in Language Testing", Cambridge: Cambridge University Press, 2004.
- [8] Brown, A. & Hill, K., "Interviewer Style and Candidates' Performance in the IELTS Oral Interview", In I. Taylor & P. Falvey (eds.), *IELTS Collected Papers: Research in Speaking and Writing Assessment*. New York: Cambridge University Press, 2007.
- [9] Van Patten, B. & Williams, J., "Theories of Second Language Acquisition", Mahwah, NJ: Lawrence Erlbaum, 2007.
- [10] Weir, C., "Understanding and Developing Language Tests", London: Princeton-Hall, 1993.

**BIOGRAPHIES OF AUTHORS**

Shahzad Karim is an Assistant Professor in the Department of English, The Islamia University of Bahawalpur, Pakistan. His areas of research interest are second language acquisition, assessment and evaluation, language shift and language maintenance, culture and language learning and textbook evaluation. His research papers are published in national and international research journals. He has also authored a book titled 'Implicit and Explicit knowledge and medium of instruction in Pakistan' published by Lambert Academic Publishing, Germany (2011).

Naushaba Haq is a Lecturer in the Department of English, The Islamia University of Bahawalpur, Pakistan. Her areas of research interest are assessment and evaluation and second language learning. Her research papers are published in national and international research journals.