

A Review of the IELTS Test: Focus on Validity, Reliability, and Washback

Ali Hashemi

University of Zanjan, Zanjan, Iran
e-mail: a.kian.hashemi@gmail.com

Samran Daneshfar

University of Zanjan, Zanjan, Iran
e-mail: s.daneshfar82@gmail.com

Abstract:

The International English Language Test System (IELTS) is one of the most reputable English tests that is used to assess the language proficiency of those who intend to study or work in an English speaking context. It is one of the most large-scale proficiency tests which affects the lives of many students, as well as immigrants as the results of the test, are used for making critical decisions about the test takers. Moreover, the process of designing a good test requires a clear understanding of both validity and reliability of the test format. Therefore, in the current paper, we try to offer a descriptive review of the IELTS test by concentrating on various issues such as reliability, validity and washback.

Keywords: IELTS, reliability, validity, washback

1. Introduction

English is the international language of science, commerce, trade, politics, and communication among a vast number of people in the world, and it has the most number of learners, too. It is used as a basis to employ people, to teach, and as the medium of interaction and instruction in the world's top universities as well. Thus, it

is critical to have a good command of English as a prerequisite for those who intend to study and work in an English context.

As a predefined English proficiency level is a required precondition to study abroad, it poses problems for the relevant candidates. Thus, a valid and a reliable assessment of language ability of the prospective students, who aim to get entry into the international academic situations, is an issue of concern. The IELTS test has been developed to meet these needs, further, to assess and provide information concerning the English level of the applicants, based on which the students will be labelled based on their English proficiency level according to specific band scores of the IELTS testing organization.

The structure of IELTS contains four equally weighted sub-components namely speaking, listening, reading and writing. The examination is generally held over the course of two separate days in specified test centers with trained markers and examiners. The speaking or interview section is held in a different session particularly before the paper-based session of the other three skills, listening, reading and writing. The mean of the four sub-sections makes up the candidate's overall band score. The speaking and writing modules Examiners are monitored regularly by an accredited IELTS trainer every two years; this is to make sure that the scorers mark the performance based on set standards prior to marking listening and reading papers (IELTS Homepage).

2. Background and Overview

Assessing the general English proficiency of test takers is a challenging and demanding issue to which easy answers cannot be sought as it includes various interacting components and subcomponents such as reading, speaking, listening, and writing (Hamp-Lyons, 1990). Given the reputation and prominence of IELTS in the lives of the involved people, it is becoming ever more important for the test organisation to provide evidence of quality control in the form of assessment reliability and validity to the outside world (Shaw, 2007).

With the rising demand for IELTS, the concerns for the quality of the conclusions drawn from its efficiency and what and how it actually measures, in terms of candidate's proficiency, is calling considerable attention. Thus test tasks and their authenticity, for example, is considered as crucial. According to Weir (1990), although tests are assigned to integrate different language skills, only direct tests that simulate authentic communication tasks can reflect actual authentic communicative interaction. Therefore, in designing and developing test tasks, care should be exerted to have tests that cover all areas of language use and which involve the learners in actual language use in communicative contexts (Gilfert, 1996).

3. IELTS test format

IELTS is available in two test versions: IELTS Academic and IELTS General Training. The Academic version is intended for those applying for higher education or professional registration, and the General Training aims at measuring the language ability of those migrating to Australia, Canada and the UK, or applying for

secondary education, training programs and work experience in an English-speaking environment.

Both the Academic and General Training versions provide a valid and accurate assessment of the four language skills: listening, reading, writing and speaking. Both versions are scored in the same way. The applicants take the first three parts of the test during a session in the following order: Listening, Reading, Writing with no breaks within these parts of the exam. The Speaking test is given either on the same day or 7 days before or after that, depending on local arrangements. Various native (American, British or Australian) accents might be used in the Listening part, and all standard types of English are accepted in responses.

IELTS Academic	IELTS General Training
<p>IELTS Academic measures English language proficiency needed for an academic, higher learning environment. The test tasks are intended for all test takers in all subjects.</p>	<p>IELTS General measures English language proficiency in a practical daily context. The test tasks reflect both workplace and social situations.</p>
<p>Listening (30 minutes, plus 10 extra minutes to transfer the answers on the answer sheet.)</p> <ul style="list-style-type: none"> • Four recorded monologues and conversations 	<p>Listening (30 minutes, plus 10 extra minutes to transfer the answers on the answer sheet.)</p> <ul style="list-style-type: none"> • Four recorded monologues and conversations
<p>Reading (60 minutes)</p> <ul style="list-style-type: none"> • Three long passages with tasks • Texts range from descriptive and factual to discursive and analytical. • Includes non-verbal materials such as diagrams, graphs, and illustrations • Texts are authentic (taken from books, newspapers & journals) 	<p>Reading (60 minutes)</p> <ul style="list-style-type: none"> • Three long passages with tasks • Section I contains 2 or 3 short factual texts. • Section 2 contains two short, work-related factual texts. • Section 3 contains one long text on a general topic • Texts are authentic (taken from official documents, books and newspapers).
<p>Writing (60 minutes)</p> <ul style="list-style-type: none"> • Writing task 1 of at least 150 words where the candidates must summarize, describe, or explain a table, chart, or a diagram. • Short essay task 2 of at least 250 words. 	<p>Writing (60 minutes)</p> <ul style="list-style-type: none"> • Letter writing task of at least 150 words • Short essay task of at least 250 words.
<p>Speaking (11 to 14 minutes)</p> <ul style="list-style-type: none"> • Face-to-face interview • Includes short questions, speaking at length about a familiar topic, and a structured discussion. 	<p>Speaking (11 to 14 minutes)</p> <ul style="list-style-type: none"> • Face-to-face interview • Includes short questions, speaking at length about a familiar topic, and a structured discussion.

4. IELTS Modules

a. Speaking (11–14 minutes)

The speaking component assesses applicant's use of spoken English. Every test is recorded by the examiner.

- Part 1 - the examiner will ask you general questions about yourself and a range of familiar topics, such as home, family, work, studies and interests. This part lasts between four to five minutes.
- Part 2 – the examinee will be given a Q-card which asks the attendee to talk about a particular topic. The attendee will have one minute to prepare and is allowed to take notes on a piece of paper before speaking about that specific topic up to two minutes. At the end of this part, the examiner will ask one or two general questions on the same topic.
- Part 3 – The applicant will be asked further questions about the topic in Part 2. These will give you the opportunity to discuss more abstract ideas and issues. This section will last for about four and five minutes.

b. Listening (30 minutes, 40 questions)

The listening component has different task types:

- Multiple choice – intends to measure in detail information of specific points or an overall understanding of the main points of the listening audio.

There is a type of question followed by three possible responses, or the beginning of a sentence followed by three possible ways to fill in blank the sentence. Test takers are required to choose the correct answer.

- Matching – assesses the listening skill for detail or the test taker's understanding of information in a given conversation. It might be applied to evaluate test takers' capability to identify connections and relations among facts in the listening manuscript.
 - The applicants are expected to match a list of numbered items from the listening audio played to a set of options on the question paper.
 - Plan, map, diagram labelling, note, summary, completion – assesses the ability to understand, for example, a description of a place, and to relate this to a visual representation. This may include being able to follow language expressing spatial relationships and directions.

Test takers are required to complete labels on a plan (e.g. of a building), map (e.g. of part of a town) or diagram (e.g. of a piece of equipment). The answers are usually selected from a list on the question paper.

c. Reading (60 minutes)

The Reading section includes three different passages and 40 questions all over, designed to test a wide range of reading skills. These include reading for gist,

reading for main ideas, reading for details, skimming, understanding the logical argument and recognizing writers' opinions, attitudes and purpose. Every passage is assigned with 20 minutes and two of the passages have 13 following questions and one of the three has 14 questions.

d. Writing (60 minutes)

i. Academic writing

The topics at this part are academic research or articles, and appropriate for, test takers entering undergraduate and postgraduate studies or looking for specialized registration. There are two types of tasks:

- Writing Task 1 – A graph, table, chart or diagram will be presented to the test takers and are asked to describe, summarize or explain the information in their own words. They might be required to describe and explain data, define stages of a procedure, the working process of a machine or describe an object or an event.
- Writing Task 2 – The applicants will be required to write an essay in response to a point of view, argument or problem. Responses to both tasks must be in a formal format.

ii. General Training writing

Topics are of general interest. There are two kinds of tasks:

- Task 1 –The test takers will be given a situation and required to write a letter inquiring information, or describing that situation. The style of the letter might be personal, semi-formal or formal.
- Task 2 – The test taker will be asked to write an essay in response to a point of view, argument or problem. The essay can be fairly personal in style.

Improving IELTS score

We can make use of a variety of real world tasks to boost each of these skills for a successful of achievement in the IELTS exam. For every skill, there are profound sources of materials to help develop these skills to a desired level. But it could be concluded that most skills are dependent on vocabulary. Vocabulary expansion is of great significance in IELTS writing, reading, listening, and speaking. Of course, this does not mean that vocabulary alone is enough. But that vocabulary is the core of skills and vocabulary development is a constituent of any exam. In some, for example, reading, vocabulary is of utmost importance. Thus, in reading, for example, we can invest on vocabulary using special techniques and strategies like pre-teaching of vocabulary in context, extensive reading, and intensive reading and lead to enhanced reading and vocabulary skill. Hashemi, Mobini, and Karimkhanlooie (2016) found that, among other techniques, the KWL (what I know, want to know, and learnt) and the Brainstorming techniques are very effective techniques in this concern.

5. Reliability

Generally, the reliability of a test is described as the extent to which the results or that can be recognized consistent or stable. For example, if examiners administer a placement test to their learners on one occasion, they would like the scores to be very much the same if they were to administer the same test again in another later session (Brown, 1996). Reliability to Bachman and Palmer (1996) is the consistency of measurement. To be considered as reliable, multiple administrations need to produce consistently similar results from an identical or near-identical test (Bachman, 1990; Weir, 1990). In line with these theorists, Hughes (1989) emphasizes that, although it is difficult to have 100% reliable tests, test writers commonly attempt to design tests yielding similar results in different administrations.

Another issue concerning reliability is that throughout developing tests, care should be exerted to make sure that the tests assess the actual skill of the test takers, systematic errors, and the unsystematic errors have no effect on test performances (Alderson et.al., 2005), to reduce or eliminate measurement error. There will, however, be a certain amount of flexibility in these figures as it is nearly impossible to ignore the vast variation in different factors involved with a test taker's performances (Kluitmann, 2008).

Bachman (1990) illustrates further concepts related to test reliability, including test method facets, random factors and personal attributes. Cover areas such as the testing situation, the test rubric, input the expected response and at the end, the relationship between input and response are covered by test method facets. Personal attributes encompass age, gender, cognitive style and background; later on, he lists a variety of random factors like tiredness, emotional condition and even some random differences in the testing context. Regarding all these notions, it is obvious that the exact estimation of test reliability is often seen to be a difficult task.

6. Validity

Focusing solely on reliability would not mention that a language test is a complete measure of language proficiency. However, validity is a separate but equally important issue. For instance, if the IELTS were administered to a group of foreign students as a test of their abilities in mathematics, the reliability would be high because the test would spread the students out rather consistently along with a continuum of scores. However, the IELTS is clearly not valid for the purpose of testing mathematical ability (Brown, 1996). The general notion of validity is defined as "the degree to which a test measures what it requires, or intents, to be measuring" (Brown, 1996, p. 231).

Kaplan (1964) points out that "The validity of a measurement consists in what it is able to accomplish, or more accurately, in what we are able to do with it. The basic question is always whether the measures have so arrived at that they can serve effectively as means to the given end" (p.198). A significant list of validities is present in the academic area of language assessment. Owen et.al. (1997) state that this list is often demanding due to the number of approved validities. However, the

most significant kind of validity could be highlighted such as construct, content, face and criterion-related validities.

The term, construct validity, is often central to theoretical testing literature, encompassing whether or not the test is actually testing the criteria it claims to test (Bachman, 1990; Hughes, 1989; Weir 1990). Hughes (1989) explains this further by stating that “the word construct refers to any underlying ability which is hypothesized in a theory of language ability”. Based on Bachman (1990) through applying construct validity empirical testing of hypothesized relationships between test scores and actual abilities takes place.

The concept of criterion-related validity is basically a subset of the ideas discussed under construct validity. Richards and Schmidt (2013) state that it is based on the extent to which a new test is compared or correlated with an established external criterion measure. There are two commonly recognized types of criterion-related validity: concurrent validity and predictive validity. Hughes (1989) argues that to establish concurrent validity the test and the criterion are administered at the same time. On the other hand, the notion of predictive validity in a test regards to whether or not the test consistently and accurately predicts the candidates’ future performance and behaviour.

The other type of validity, content validity, requires a test to be well described, including references to its theoretical and empirical foundations, in addition, the test makers are expected to describe the purposes and characteristics of the items (Turkstra et.al. 2005). Hughes (1989) claims that a test to be content valid should consist of a representative sample of language structures and skills which it is supposed to be concerned with. Oller (1979) explains that content validity supports the test takers to perform genuinely the same tasks or basically similar to tasks one normally performs in displaying the skill or ability the tests indicate to assess.

The third primary type of validity, face validity, is achieved when the test looks as if it is measuring what it is supposed to and should appear as such to the test-taker (Turkstra et.al. 2005). Brown (1994) explains face validity as for whether the test, on the face of it, truly resembles to test what it is intended to assess, from the learner’s perspective. If test-takers accept that the gained results are accurate, then face validity can be associated more with acceptance than actual validity (Alderson, Clapham and Wall, 2005)

7. Washback

Language testing is an uncertain and approximate business at the best of times, even if, to the outsider, this may be camouflaged by its impressive, even daunting, technical (and technological) trappings, not to mention the authority of the institutions whose goals tests serve. Every test is vulnerable to good questions (McNamara, 2000) Tests are inevitably political since what they do is to sort and select to meet society’s needs. Testers cannot expect that their work will not have a political dimension. The proper reaction to such concern is sure to act with professional skill and rectitude within the contexts in which they work. (Davies 1990)

The ELT scholars have addressed the power of examinations which hold in the schools and their critical importance of promoting the educational process. Pearson (1988) comments that, "It is generally accepted that public examinations influence the attitudes, behavior, and motivation of teachers, learners, and parents" (cited in Fulcher and Davidson, 2007, p.222).

Therefore, backwash is also an important factor considering language assessment. Owen et.al. (1997) demonstrate that changing either a test or the marking system of that test can have consequences on how the test subject is taught and how students might approach their learning.

According to Buck (1988), washback is a natural tendency for both teachers and learners in order to modify their classroom activities to the requirements of the test, especially when the test is of more significance to students' future, and pass rates are used as a measure of teacher success. This impact of the exam on the classroom or backwash effect is highly important and can be either helpful or damaging.

8. Reliability in IELTS

To keep reliability in marking in both the objectively and the subjectively marked modules of IELTS has been a significant interest. As reported on the IELTS homepage a rigorous process of test production has produced Reading and Listening versions with an average Cronbach Alpha of 0.88, estimated from the performance of over 90,000 applicants on thirteen reading and listening versions acceptable as a measurement of the consistency and reliability of a test (UCLES, 2007).

To estimate the reliability of Writing and Speaking Modules in IELTS will take different methods but their quality is confirmed through a comprehensive program of training, certification, and monitoring of examiners. The voice recording method is applied during the interview of speaking module, and scripts from the writing tasks and are kept for two months at least. Further, the results of IELTS are checked prior to release, and a formal procedure allows candidates to query their results within two weeks. (Fazel and Ahmadi, 2011)

While there is a significant emphasis held on certification, retraining and standardization of examiners for the writing and speaking parts, one of the drawbacks is that the IELTS homepage doesn't actually offer any data analysis for the reliability of the modules. However, they offer an overall reliability estimate based on Fldt and Brennan's (1989) theory over the four modules which gives a high coefficient of 0.95, in turn producing a low SEM of 0.21. Nonetheless, producing a true coefficient figure which hasn't potentially been managed for the purpose of propaganda and marketing is challenging and until there is an actual method of objectively estimating the reliability coefficient for the writing and modules speaking.

Hughes (2003) claims that for the validity a test must confirm consistently accurate measurements, therefore, it must be reliable. Regarding the 2014 IELTS Annual Review, the reliability of the reading test is measured using Cronbach's alpha for

internal consistency of the 40-items used in the test, this highlights that the grades accomplish adequate levels of reliability at about 0.90.18.

Some researchers have implied that different types of language exams should imply varying reliability coefficients. Lado (1961), for instance, argues that reading tests claim a higher reliability extent than auditory comprehension tests, which in turn are more reliable than oral production tests. This will confirm that it is within the reading subtest which significant correlation between IELTS scores and academic results appeared. Dooley and Oliver (2002) illustrate that the significant correlation for business students might be for the reason that the discipline is generally regarded as the 'most linguistically demanding'. Though the correlations amount was low, to expect a correlation of above 0.3 (about 10 percent) in predictive validity studies of such types is not required. (Criper and Davies 1988)

Alshammari (2016) states that the case of strictness of IELTS scoring system could be a major factor affecting the reliability of candidates' performances and final scores. However, a body of more research on these issues is required to improve the test scoring system in order to be more consistent with the purpose of the test which must be focused on assessing how well candidates use their reading skills to comprehend what they read regardless of any other considerations.

The IELTS exam contains four components upon which an overall band score is awarded. Thus an estimate of composite reliability offers a useful measure of overall test reliability. Fazel and Ahmadi (2011) applied a method of estimating the reliability of a composite taken from Feldt and Brennan (1989). In order to generate a careful estimate, minimum alpha values were used for the objectively scored papers; and g-coefficients for the single rater condition on subjectively marked papers. The findings of the study proved a positive reliability amount of Iranian learners estimated reliability for the Academic module through a pilot study by Salmani-Nodoushan (2002).

9. Validity in IELTS

A number of predictive validity studies have attempted to distinguish the connection between IELTS scores and academic performance, with variable results. Some studies (Fiocco 1992; Gibson & Rusek, 1992) have discovered no connection between them, while others (Bellingham 1995; Cotton & Conrow, 1998; Feast, 2002) have discovered generally positive correlations between IELTS entry levels and GPAs. Applying a homogenous sample, Yen and Kuzma (2009) showed significant positive correlations between IELTS scores and GPAs and therefore confirmed IELTS' predictive power in students' academic performance. The study proves the legitimacy of using students' IELTS grades as an entrance criterion.

The significant body of work on the predictive validity of IELTS goes back to research conducted prior to the revision of the IELTS test in 1995. Fiocco (1992) observed a group of 61 non-English-speaking-background students at Curtin University in Western Australia, some of whom did not meet the university's IELTS requirement, but were nevertheless admitted to their mainstream course. The study included two different phases which in the first one the students' IELTS grades were

correlated with their semester weighted averages (PPI) attempting to gain some insight of whether it is possible to regard IELTS as a predictor of success. Within the second part, students were interviewed to reveal their concerns about the language difficulties they were experiencing, and whether these difficulties were reflected in their IELTS scores.

From the data analysis in the first phase, no significant difference was reported 'no meaningful difference' in the correlation between PPIs and IELTS scores between those who had met the university's English language requirements, and those who had not. No significant difference was found in the correlation between those who had entered courses that were considered 'linguistically demanding' and 'less linguistically demanding' (Fiocco 1992). However, the findings of the second phase proposed that language did have a role to play in academic success. Some students reported on having other problems including teaching methods, the teachers and feelings of alienation and isolation. (Dooley and Oliver 2002)

Ferguson and White (1994) indicate that the predictive validity of IELTS is higher when IELTS scores are lower, i.e. the correlation should be higher for students whose band scores fall below 6.0. Eventually, in the context of predictive validity of four modules of the IELTS test, Cotton and Conrow (1998) noticed that it is the IELTS reading subtest which had a moderate positive correlation with academic results. Moreover, they reported a negative correlation with global academic results regarding speaking subtest, inferring that the reading subtest scores, in particular, were best able to predict subsequent academic performance (p. 109).

Dooley and Oliver (2002) considered the students' performance of three main language groups, Indonesian, Cantonese and other Chinese languages to investigate whether other factors, such as language background, mediate the predictive validity of IELTS. The results of their study revealed no important difference in the comparison between IELTS scores and academic success across disciplines. Accordingly, the results of their study proved that the predictive validity of IELTS does not differ according to the linguistic demand of the courses.

10. Washback in IELTS

The increasingly recognized impact of examination boards on education process and therefore generally on the society is confirmed. That is due to the examinations widespread recognition and cash-in-value. Such impact is often perceived in two levels, i.e. at the macro or social institutions or the micro level or what is dealing with the individual (Green, 2007). The effects which could be positive or unintended negative effects are referred to as the washback, in which the effects are considered related to teaching practice, learning outcomes and also to the courses. Since the population of English users is considerably soaring, the application of international tests such as IELTS and its consequences request a high scholar consideration like the above-discussed issues like reliability and validity. On the other hand, the significant influence of such high-stakes testing on teaching and learning derives researchers to investigate the washback of IELTS.

Green (2007) investigated the influence of test preparation courses on improving the students' IELTS academic writing module score. The findings illustrated that the test-driven instruction did not raise students' scores but in order to improve their score, the material covered on the test was to be integrated with regular teaching and prior preparation. Moreover, he stated that different variables, i.e. learners' background, course length and the amount of exposure outside the classroom context, may affect washback.

Rashidi and Javanmardi (2011) clarified on the possible washback effect of IELTS preparation courses on the learning and teaching consequences or the students' success through the examination. They applied the Hughes' 1994 trichotomy washback model. The three phases included washback to participants, to the process and to the products. While IELTS preparation courses had positive impacts on the students' learning processes and their achievement in the examination, the conclusion is not absolute regarding the students' both optimistic and pessimistic expectations toward some aspects of these courses through answers to different questions prior and after the IELTS course. Therefore, the results of their study brought about no definite advantage of the preparation courses.

Hayes and Read (2004) IELTS academic module assesses four macro-skills through a variety of tasks which are designed to stimulate authentic study tasks. Therefore, they argue that it is intended to have a positive washback effect aiming at progressing applicants' language proficiency in a way that assists their academic study through English. They showed a clear evidence of washback in IELTS preparation course at school; however, while the teacher and students were narrowly focused on the practice of the test tasks rather than the improvement in language proficiency the results didn't seem to be the type of positive effects expected.

On the study of IELTS washback on textbooks, Yue (1997) suggested that some of the textbooks clearly pay attention to practicing skills and subskills that are requested by IELTS, providing adequate information about the exam and increase students testing abilities. In this case, IELTS is improving positive washback on preparation materials. But it is preferred that the analysis of textbook materials revealed both direct relationships between textbook and test system like same formats and task types, in addition, indirect ones such as "opportunities to intensify the performance of English-speaking culture-relevant micro-skills, functions, activities, in relevant settings, media modes." If the test has been developed to meet their real communication needs, in this case, both direct and indirect test-related materials could assist learners to prepare for the test as well as enhance their learning and their language performance.

Lewthwaite (2007) focused on the perceptions of teachers and students involved in the IELTS writing test. The results of the study found a significant overlap between the IELTS writing tasks needed and what students and staff thought was required in a writing course. Some of the participants felt a tension between what was possible in an exam course and what the ideal pedagogic practice. Considering the differences commented between the IELTS and university writing tasks in his study, the IELTS tasks showed a suitable teaching emphases guide to help students' writing

output applying the necessary skills. All the observed feelings were the positive reported attitudes towards the exam in contrast with what some researchers found washback negatively. (Davidson & Mandalios, 2005)

11. Conclusion

The International English Language Testing System (IELTS) is famous worldwide with speakers of English as a second language as a demonstration of their language proficiency. It is a highly recognized language certificate by authorities both for academic and immigration purposes. Therefore, the popularity of the IELTS test implores and gets the language assessment scholars attention to study its aspects, limitations and strengths. The present study brought a critical review of the IELTS test following some aspects and notable issues concerning language assessment generally. Critical works on the IELTS reliability, validity and washback highlighted in this study as the main concepts of every language test. Research highlights that IELTS can be applied as a measure of English as an international language needs to be made available. The commitment of IELTS to research and its responsiveness to findings is proved in the literature. One of the drawbacks of IELTS reliability is that the IELTS homepage doesn't actually propose any data analysis for the reliability of the modules; however, they offer an overall reliability estimate. While some researchers approved the IELTS validity attention to a series of issues suggests that IELTS should embark on much more research to improve test validity as well as reliability. Eventually, regarding the consequences and impacts of an examination on the learning and teaching process (students, educators and further on the materials), there must be an important attention toward the IELTS washback in order to improve in its authenticity and usefulness.

References

- Alderson, J., Clapham, C., & Wall, D. (2005). Item writing and moderation. *Language test construction and evaluation*, 40-72.
- Alshammari, M. M. (2016). IELTS ACADEMIC READING MODULE TEST: VALIDITY AND RELIABILITY. *British Journal of English Linguistics*, 4(2), 26-37.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests* (Vol. 1). Oxford: Oxford University Press.
- Bellingham, L. (1995). The relationship of language proficiency to academic success for international students. *New Zealand Journal of Educational Studies*, 30(2), 229-232.
- Brown, H.D. (1994). *Principles of Language Learning*. 3rd ed. White Plains, New York: Longman.
- Brown, J. D. (1996). *Testing in language programs*. Upper Saddle River, NJ: Prentice Hall Regents.

- Buck, G. (1988). Testing listening comprehension in Japanese university entrance examinations. *JALT Journal* (10), 12-42.
- Cotton, F., & Conrow, F. (1998). An investigation of the predictive validity of IELTS amongst a group of international students studying at the University of Tasmania. *International English Language Testing System (IELTS) Research Reports 1998: Volume 1*, 72.
- Criper, C., & Davies, A. (1988). ELTS validation project report. University of Cambridge, Local Examinations Syndicate.
- Davidson, P. & Mandalios, J. (2005). Using Assessment to Facilitate Learning. Presentation given at the 9th CTELT Conference, Dubai, United Arab Emirates.
- Davies, A. (1990). *Principles of language testing*. Oxford: Basil Blackwell.
- Dooley, P. and Oliver, R. (2002). An investigation into the predictive validity of the IELTS test as an indicator of future academic success. *Prospect*, 17(1), 36-54.
- Fazel, I., & Ahmadi, A. (2011). On the relationship between writing proficiency and instrumental/integrative motivation among Iranian IELTS candidates. *Theory and Practice in Language Studies*, 1(7), 747-757.
- Feast, V. (2002). The impact of IELTS scores on performance at university (Doctoral dissertation, Flinders University).
- Feldt L.S & Brennan R. L. (1989). Reliability. In Linn (Ed): *Educational Measurement*, 3rd Edition. American Council on Education: Macmillan.
- Ferguson, G., White, E. (1994). *A Predictive Validity Study of IELTS*. University of Edinburgh, Institute for Applied Language Studies.
- Fiocco, M. (1992). English proficiency levels of students from a non-English speaking background: A study of IELTS as an indicator of tertiary success. Unpublished research report, Curtin University of Technology, Perth, Australia.
- Fulcher, G. & Davidson, F. (2007). *Language Testing and Assessment* (1st.ed.). London: Routledge Press.
- Gibson, C., & Rusek, W. (1992). The validity of an overall band score of 6.0 on the IELTS test as a predictor of adequate English language level appropriate for successful academic study. Unpublished Masters of Arts thesis: Macquarie University, New South Wales, Australia.
- Gilfert, S. (1996) A Review of TOEIC. In *The Internet TESL Journal*. <http://iteslj.org/Articles/Gilfert-TOIEC.html> Accessed 6th January 2017
- Green, A. (2007). *IELTS washback in context: Preparation for academic writing in higher education* (Vol. 25). Cambridge: Cambridge University Press.
- Hamp-Lyons, L. (1990). Second language writing: Assessment issues. *Second language writing: Research insights for the classroom*, 69-87.
- Hashemi, Ali, Mobini, Fariba, and Karimkhanlooie, Guiti (2016). The Impact of Content-based Pre-reading Activities on Iranian High School EFL Learners' Reading Comprehension. *Journal of Language Teaching and Research*, Vol. 7, No. 1, pp. 137-145.

- Hayes, B., & Read, J. (2004). IELTS test preparation in New Zealand: Preparing students for the IELTS academic module. *Washback in language testing: Research contexts and methods*, 97-111.
- Hughes, A. (1989). 1989: Testing for language teachers. Cambridge: Cambridge University Press.
- Hughes, A. (2003). Testing for language teachers. Cambridge: Cambridge University Press.
- Kaplan, A. (1964). *The Conduct of Inquiry: Methodology for Behavioral Science*. San Francisco: Chandler.
- Kluitmann, S. (2008). Testing English as a Foreign Language. Two EFL-Tests used in Germany. Unpublished Doctoral Thesis (ELT), University of Albert-Ludwigs.
- Lado, R. (1961). *Language Testing: The Construction and Use of Foreign Language Tests. A Teacher's Book*. Pennsylvania State University: Longmans.
- Lewthwaite, M. (2007). Teacher and student attitudes to IELTS writing tasks: positive or negative washback. *UGRU Journal*, 5, 1-16.
- McNamara, T. (2000). *The social dimension of language testing*. Malden, MA: Blackwell.
- Oller, J. W. (1979). *Language tests at school*. London: Longman.
- Owen, C. et al (1997) *Testing*. Centre for English Language Studies, Birmingham University.
- Rashidi, N., & Javanmardi, F. (2011). The IELTS preparation washback on learning and teaching outcomes. *Cross-Cultural Communication*, 7(3), 132-144.
- Richards, J. C., & Schmidt, R. W. (2013). *Longman dictionary of language teaching and applied linguistics*. New York: Routledge.
- Salmani-Nodoushan, M. A. (2002). Text Familiarity, Reading Task, and ESP Text Performance: A Study on Iranian LEP AND NON-LEP University Students. Unpublished Ph.D. Dissertation, University of Tehran, Tehran.
- Shaw, S. D. (2007). Modelling facets of the assessment of writing within an ESM environment. *Research notes*, 27(1), 14-19.
- Turkstra, L. S., Coelho, C., & Ylvisaker, M. (2005, November). The use of standardized tests for individuals with cognitive-communication disorders. In *Seminars in Speech and Language* (Vol. 26, No. 04, pp. 215-222). Copyright© 2004 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA.
- UCLES. (2007). IELTS handbook 2007. Retrieved from the IELTS Web site: http://www.ielts.org/pdf/IELTS_Handbook_2007.pdf
- Weir, C. (1990) *Communicative Language Testing*. New York: Prentice Hall
- Yen, D., & Kuzma, J. (2009). Higher IELTS score, higher academic performance? The validity of IELTS in predicting the academic performance of Chinese students. *Worcester Journal of Learning and Teaching*, (3), 1-7.
- Yue, W. W. (1997). An investigation of textbook materials designed to prepare students for the IELTS test: A study of washback. Unpublished master's thesis, Lancaster University, England, 226.