Global Journal of English Language Teaching
Vol. 2, No. 1, 9-15, 2022
ISSN(E) 2835-9992 / ISSN(P) 2836-0001
DOI: 10.20448/gjelt.v2i1.4128

Investigating the Validity of the IELTS Listening Test

Peter Peltekov

University of Calgary, Canada.
Email: pnpeltek@ucalgary.ca

Abstract

The International English Language Testing System (IELTS) has gained popularity in recent years and it is now accepted by many educational institutions worldwide. While IELTS offers a distinct academic version of the reading and writing test components, it uses the same listening module for the General Training and the Academic exam. The present article explores to what extent the listening subtest in the Academic IELTS test is a useful measure of test-takers listening ability and a predictor of their academic success. This study focusses on the test's construct validity. A critical review of the major strengths and potential drawbacks of the listening test is followed by a comparison of the test scores with another traditionally accepted test of academic English. Conclusions about the validity of the IELTS listening test are drawn along with some suggestions for design improvement. Future research directions are also proposed.

Keywords: IELTS, Language assessment, Listening test, Construct validity.

Citation | Peter Peltekov (2022). Investigating the Validity of the IELTS Listening Test. Global Journal of English Language Teaching, 2(1): 9-15.
History:
Received: 13 June 2022
Revised: 22 July 2022
Accepted: 5 August 2022
Published: 29 August 2022

Publisher: Asian Online Journal Publishing Group

Funding: This study received no specific financial support.
Competing Interests: The author declares that there are no conflicts of interests regarding the publication of this paper.
Transparency: The author confirms that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained.
Ethical: This study followed all ethical practices during writing.

Contents

Contribution of this paper to the literature

This critical analysis of the IELTS test contributes to our understanding of the construct validity of the IELTS listening module. Existing empirical evidence is used as a basis to analyze to what extent the listening subtest in its current form is a valid measure of L2 English learners' listening skills.

1. Introduction

Language proficiency tests are often used as gate-keeping measures and can have important consequences for both the assessees and the assessment users (Green, 2014). Such high-stakes tests are often criticized for the impact they can have on test takers. As Shohamy (2001) points out, a single test can automatically trigger important decisions such as admission or graduation. Therefore, Green (2005) warns that gate-keeping decisions should be based on "multiple sources of evidence of learners' abilities". Of course, this is not always practicable, and normally a single test is used. In this case it is important that the test is well validated. The problem is that "[m]any public tests with a significant burden of responsibility in important decision making about individuals have been too little validated" (McNamara, 2000). Therefore, test validation is an important part of psychometrics research, and it deserves adequate attention.

While the Test of English as a Foreign Language (TOEFL), described by Shohamy (2001) as "the symbol of a [language] test", still remains one of the most widely used tests of English, IELTS has also gained popularity and it is now accepted by thousands of educational institutions worldwide, including Norwegian universities (Samordna Opptak, 2020), as an alternative to TOEFL. Although the popularity of IELTS may speak to its capacity to provide useful inferences of candidates' English language proficiency, test validation is an ongoing process and assessment instruments should continuously be improved (Green, 2014). The present study is a further step towards this ongoing validation process.

The listening module of the IELTS test deserves special attention, as the academic version of IELTS uses the same listening test that is used in its general training version. It is worthwhile investigating whether this might have any effects on the construct validity of the academic IELTS test. Moreover, listening is generally neglected and considered "the least researched of all four language skills" (Vandergrift, 2007), although it is one of the skills posing great difficulties for L2 students in English-speaking universities (Huang & Finn, 2009). Therefore, the present article will aim to analyze the listening portion of the IELTS test.

As McNamara (2000) explains, language tests are used to make inferences about "how a person is likely to manage the language and communicative demands of the subsequent non-test or criterion situation, for example, listening to lectures". Are scores from the Listening subtest in the IELTS Academic test related to the amount of language-related academic difficulties students experience in their subsequent studies? In other words, does the Listening test have any predictive value for test-takers? There is a scarcity of research focussed on this question, and the few studies available yield contradictory results. While some studies suggest that that there is a positive correlation between the IELTS listening scores and students' academic performance and self-perceived listening abilities (e.g., Breeze and Miller (2011)), other studies indicate no such relationship (e.g., (Ingram & Bayliss, 2007; Kerstjens & Nery, 2000)). As Cooper (2013) suggests, however, listening ability is just one of many factors involved in students' academic performance. Therefore, the relationship between scores obtained on the IELTS Listening subtest and students' academic performance might not always be so evident. Breeze and Miller (2011) also suggest that the strength of the correlations between the IELTS Listening subtest scores and students' final grades might as well depend on the local context. As the authors explain, in some contexts (e.g., formal, teacher-centered lectures), listening competence is much more central to students' academic achievement than in other situations. Despite these contextual variations, the IELTS academic listening test can still be validated by analyzing its structure and comparing it to other tests that are intended to measure the same construct.

2. Definition of Construct

In order to be able to judge the validity of an assessment method, it is important to first of all clearly define what the assessment is supposed to measure (Green, 2014). The IELTS Academic test is meant to test academic English language proficiency (IELTS, 2020). Academic language proficiency is central to scholastic success and can be empirically distinguished from interpersonal communicative skills (Cummins, 1980). Academic English listening skills, in particular, refer to the "advanced level listening skills for academic purposes" (Taylor & Geranpayeh, 2011). Richards (1983) defines academic listening as the skills required to listen to lectures, and provides a series of micro-skills involved in academic listening, such as the ability to identify purpose and scope of the lecture, the ability to identify the topic of the lecture and follow the topic development, the ability to identify relationships among units within discourse, etc. This construct definition should be considered in the design and evaluation of an academic listening test.

3. Description and Analysis of the IELTS Listening Test

As McNamara (2000) explains, the process of test validation involves not only examining the empirical evidence (i.e., the hard facts) but also "thinking about the logic of the test, particularly its design and its intentions" (p. 48). It is important to analytically reflect upon some of the main threads to the test validity such as: the test *content* (i.e., what the test contains), the test *method* (i.e., how test takers engage with the test), and the test *construct* ("the underlying ability captured by the test", McNamara (2000). Therefore, a qualitative analysis of these aspects of the IELTS listening test would be a first step towards the investigation of the test validity. This also involves a systematic description of the characteristics of the test tasks and the way they may or may not reflect the target language use (TLU) contexts. Different frameworks for analysing test task characteristics (e.g., (Bachman & Palmer, 1996; Douglas, 2000)) have been proposed, but they all begin with an important element—the test rubric.

4. The Rubric

The test rubric includes information about the structure of the test, the time allotment for each task, and the procedures that are followed. The Listening module in the IELTS Academic test lasts approximately 30 minutes. Test takers are instructed to write their answers on the question paper as they listen to the recordings, and at the end they are given additional 10 minutes to transfer their answers to the official answer sheet. It is important to note that for the listening subtest there is no distinction between academic and general versions, meaning that all test takers listen to recordings on both general and academic topics. This is a characteristic that may affect the extent to which the test reflects the TLU domain, which I explain in more detail below. There are 40 questions, and the question types include: multiple choice, matching, plan/map/diagram labelling, form/note/table/flow-chart/summary completion, and sentence completion. Evaluation is straightforward—each correct answer is awarded one point (IELTS, 2020). It is noticeable that the instructions provided to test takers and test administrators are very detailed and clear, which ensures the consistency in the procedures, and the reliability of the test.

5. The Input and the Expected Response

In order to examine the test tasks characteristics in more detail, it is important to analyze the input that test takers receive, the type of response they are required to provide, and the way the input and response interact with each other (Bachman & Palmer, 1996; Douglas, 2000).

BISC and CALP. The input includes some topics that are more academic in nature (e.g., air and water pollution) but also topics that are conversational in genre, such as finding a restaurant and accommodation (Douglas, 2000). Therefore, in the response students are expected to demonstrate both their cognitive academic language proficiency (CALP) and their basic interpersonal communication skills (BICS), as defined by Cummins (1980).

Taylor and Geranpayeh (2011) explain that the reason why academic listening tests might include tasks focussed on interactional and transactional exchanges is to acknowledge "the student's need to be able to cope socially" in the TLU domain. It is difficult to argue against this proposition, but if the test target population consists of "those entering or undertaking academic studies" (p. 93), then the focus of an academic listening test should be on measuring academic listening skills. Instead, one half of the listening subtest in the IELTS Academic test is based on general listening skills (IELTS, 2020), which could be a potential source of construct-irrelevant variance and, as a result, it can affect the test validity.

When both BICS and CALP are measured, separate scores should be provided for academic listening (Huang & Finn, 2009). The reason why it is so important to make a precise inference of test takers' abilities to listen to lectures is not just because this ability would be crucial for them in the TLU domain but also because academic listening is special in many ways. For example, it involves "long stretches of talk, and the listeners don't have the opportunity of engaging in the facilitative functions of interactive discourse" (Huang & Finn, 2009). It is also normal that input in academic listening contains a higher number of lexical items can be found in the Academic Word List (Coxhead, 2000). Last but not least, lecture comprehension does not depend so much on the meaning of individual words and sentences as on their inter-relatedness and students' ability to use explicit discourse markers to understand the structure of the lecture (Buck, 2001).

Reactivity. The relationship between input and response can vary in levels of reactivity (i.e., the degree of reciprocity/interaction, Green (2014). The non-reciprocal relationship between the input and the responses (i.e., the input in the recordings cannot be altered by the responses) in a listening test could affect the situational authenticity of the tasks. Although non-reciprocal test tasks are a necessary component of language tests (Douglas, 2000), the lack of reciprocal tasks may lead to under-representation of the listening construct, as in many TLU situations speakers can influence the input through interpersonal interaction. For example, lectures can become interactive when students ask questions. Research shows that in interactive listening, speakers can negotiate meaning and receive modified input which is more comprehensible (Long, 1996). However, it would be unrealistic for students to expect to interrupt a lecturer each time they experience comprehension difficulties, especially in large university classes. Therefore, it is unlikely that the low level of reactivity observed in the IELTS Academic listening test would have major effects on the situational authenticity of the test. On the contrary, we could expect that the higher demands of the non-interactive listening on the test would reflect more closely the test takers' needs in the TLU domain.

Scope. Green (2014) defines scope as "the amount and range of input to be processed in order to respond". Most of the recordings in the IELTS listening test are short and responses normally refer to a specific part of the audio input, meaning that the scope of the relationship between the input and the response is rather narrow. Moreover, for most tasks, test takers are given time to read the questions before they listen to the recording, which draws their attention to specific details only. As Green (2014) suggests, however, listening tests should assess not only listening for details but also other types of listening (e.g., listening for gist and general ideas). On the other hand, allowing students to preview the questions can be regarded as beneficial, as it may reduce some memory-related issues, and, as a result, it can allow them to improve their listening comprehension performance (Hemmati & Ghaderi, 2014).

Directness. No background knowledge is required by test takers, as responses are entirely based on the input (IELTS, 2020). Despite the direct relationship between input and response, topic knowledge is still among the factors that can influence the comprehension process (Brunfaut, 2016). While we can expect positive associations between prior knowledge and listening comprehension, it is sometimes possible that prior knowledge can also lead to inaccurate comprehension if the listener does not listen for possible contradicting information (Macaro, Vanderplank, & Graham, 2005). Therefore, in order to neutralize the effects of background knowledge, it is necessary to ensure that the test is not limited to a single topic. A review of the sample tasks in the IELTS listening test shows that there is a great range of topics included, which is a proof of effective content sampling.

Authenticity. The major drawback of the IELTS listening test concerns the level of authenticity (Huang & Finn, 2009). As Green (2014) suggests, recordings should ideally be obtained from real world sources. The input data in the IELTS listening subtest, however, consists entirely of scripted (i.e., non-authentic) materials (IELTS, 2020). This feature reduces the level of authenticity, which is one of the most important characteristics of effective assessment tools (Bachman & Palmer, 1996). The lack of significant correlation between scores on the listening module and

subsequent academic performance (Aryadoust, 2012; Ingram & Bayliss, 2007; Kerstjens & Nery, 2000) can therefore be partially attributed to the limited situational authenticity of the test. It is expectable that if the input does not reflect the TLU domain closely enough, the predicative validity of the assessment will be low, which would also reduce the beneficial consequences of the test.

Construct-irrelevant variance. A major problem in the assessment of receptive skills is that they cannot be directly observed (Green., 2014). Therefore, another potential issue in the listening module's design is the need for test takers to use construct-irrelevant skills to demonstrate listening comprehension. As Flowerdew and Miller (2012) explain, "the success or failure of learners on the test may rest not so much with their listening ability but with their reading, writing and speaking proficiencies". A review of the test items in the IELTS listening test reveals that test takers are required to read all questions and sometimes also the possible answers to these questions. Instead of asking test-takers to read them, they should be allowed to listen to them, as in a listening comprehension test there should be "no reading involved either in the question prompts or in the answer choices" (Vandergrift, 2006). Sometimes test takers also need to provide a written response, while poor spelling and grammar are penalized (IELTS, 2020). While the TLU domain often requires integration of different skills, the effects of construct-irrelevant skills should be minimized in order to improve the validity of the test. Just like the writing test does not involve any listening, the listening test should also not require any writing and spelling skills. When it is impossible to completely avoid the use construct-irrelevant skills, it is important that these skills do not affect the scoring (Brunfaut, 2016). Therefore, poor spelling and grammar in the answer sheet should not be penalized as long as the responses are intelligible.

Multiple-choice questions. Although multiple-choice questions (MCQs) may be a good way to assess receptive skills, providing just a few possible answers can encourage guessing and distort the results. Each of the MCQs in the IELTS listening module offers only three answer choices, meaning that test takers have over 33% chance of guessing the correct answer. As Aryadoust (2012) demonstrates, "low-ability people who have received training in test-taking strategies appear to be taking advantage of this fact, leading to flawed test results". Therefore, we can expect that these test items would, to a large extent, test assessees' luck and/or test-taking skills.

Visual input. One important question on which researchers seem to be divided regards the use of visual input in tests of listening comprehension. Buck (2001) claims that adding visual information can unnecessarily increase the cognitive load of test-takers, which may interfere with the testing process (p. 254). Buck adds that visual information also "has the potential to influence or change the listener's interpretation of the speaker's words in a significant way" (p. 48). Therefore, he advises that when visual information is present, it is important to ensure that it "complements the audio text, rather than conflicts with it" (p. 123). Moreover, test takers differ in their ability to interpret visual stimuli, and the presence of visual information in the listening test can be regarded as construct-irrelevant, as it may test assessees' ability to make sense of visual cues, rather than their listening abilities (p. 172). All of these concerns are irrelevant to the IELTS listening test, as it includes only audio input. Therefore, in this line of thought, the validity of the test is not threatened. According to Wagner (2007), however, "[t]o preclude non-verbal information on listening tests could be seen as a threat to the validity of the inferences made about a person's L2 listening ability". Indeed, listening has a pragmatic side which involves understanding the speaker's intended meaning and includes listeners' ability to interpret non-verbal cues such as facial and hand gestures (Brunfaut, 2016). Moreover, attending university lectures always involves listening to *and looking at* the lecturer (which is the case of most online lectures, too), and the inclusion of only audio materials in the IELTS listening test decreases the situational authenticity of the test, sinse the tests tasks are not truly representative of the TLU context. Not surprisingly, Ockey (2007) argues that the academic listening construct might have to be reconceptualized and suggests that video materials should be utilized in tests of academic listening, which would "make it more possible to generalize the results of the test to lecture comprehension in a typical classroom environment".

6. A Comparison with the TOEFL Listening Test

If Shohamy (2001) defines TOEFL as "the symbol of a [language] test", then it would be reasonable to compare the IELTS listening test with the TOEFL listening test. Since the TOEFL test is the most widely recognized English language test in the world (ETS, 2010), it is expected that the validity of the IELTS test might be predicted by similarities in the design and structure of the two tests as well as by any observable correlations between the test scores.

The sample test items and the instructions provided on the official TOEFL website (TOEFL, 2020) show that the TOEFL listening test shares many similarities with the IELTS listening test. For example, both tests measure test takers' abilities to understand both lectures and conversations, without providing separate scores for assessees' CALP and BICS listening skills. Therefore, in both cases, the scores might not be indicative of test takers' academic listening performance (Huang & Finn, 2009). Since the TOEFL test is so commonly accepted by higher education institutions, it appears that this feature of the listening test is not of great concern to test developers and test users. Therefore, in this aspect, the similarity between the TOELF and IELTS listening tests might not be a proof of the construct validity of the IELTS test, but it certainly suggests that both tests enjoy face validity.

Another similarity between the two listening tests is that both of them involve construct-irrelevant skills. The following question from the TOEFL listening sample test illustrates how test takers' may be able to provide a correct answer entirely based on their reading skills (TOEFL, 2020):

Read part of the conversation again. Then answer the question.
(Female student) I'm sorry I had to miss practice, though. I feel bad about that.
(Male coach) Family's very important.
What does the man mean when he says: "Family's very important."
a. He hopes the woman's family is doing well.
b. He would like to meet the woman's family.
c. The woman should spend more time with her family.
d. The woman had a good reason for missing practice.

Although the test taker is supposed to make inferences from different parts of the audio text to answer the question, the correct answer (i.e., 'd') can also be inferred based on the test taker' good reading and reasoning skills.

The example above shows that the TOEFL listening test also involves reading, which is a construct-irrelevant skill. Unlike the IELTS listening test, however, it does not involve any writing. While the IELTS listening test includes a variety of response types (many of which are written responses of more than one word), the TOEFL listening test uses only multiple-choice questions (MCQ). Because MCQ are such a common testing format, it is also possible that limiting the response types to MCQ in the TOEFL listening test also reduces the effects of another construct-irrelevant variable—assessees' test taking skills. In other words, candidates' performance on the IELTS listening test may be more dependent on the degree of their (un)familiarity with the task types (e.g., diagram labelling, flow-chart completion) in comparison to the TOEFL listening task. Most importantly, MCQ do not require any active language production[1]. Finally, providing four answer choices (as opposed to three options in IELTS) significantly reduces the possibility that assessees' correct responses would be due to chance.

Another advantage of the TOEFL listening test might be the presence of visual input. As Buck (2001) suggests, the provision of a picture accompanying the audio recording can activate relevant knowledge schemata and help test takers to focus their attention. Moreover, there is less variability in listeners' engagement with still photographs than with videos (Ockey, 2007), which suggest that the inclusion of pictures on a listening test is not likely to have a significant impact on test takers' relative scores.

The content of the TOEFL listening test includes largely American cultural references (Geranpayeh & Taylor, 2013), and while the IELTS listening test includes a range of international accents, the recordings in the TOEFL test use only the North American English accent (Huang & Finn, 2009). Since both tests are used internationally and candidates are likely to encounter speakers from many different L2 backgrounds among their future colleagues and professors (even at a single American university), it is reasonable to include more than one accent on the listening test. Therefore, we could argue that the IELTS listening test is better designed in this aspect.

One more major noticeable difference is that the TOEFL listening test is considerably longer—it may last for up to 90 minutes, and it includes up to 51 questions (TOEFL, 2020). A longer test can often provide more precise evidence of test takers' abilities. Some of the questions in the listening section, however, do not count toward the test taker's score, as these questions are part of the trialling process of new test items. This appears to be a common piloting procedure used by test developers (Green, 2014), and it undoubtedly contributes to the overall validity of the test. However, it is not clear how much of the extra length of the test is due to these test items and whether test takers experience any additional fatigue as a result of it, which could affect their performance on the test items that are actually scored.

7. The Empirical Evidence

Most empirical studies investigating the validity of the IELTS listening test have focussed on the predictive validity of the test. As mentioned earlier, the results of these studies are inconclusive and, in some cases, contradictory. While some studies found a positive correlation between the scores from the Listening subtest and students' academic performance (e.g., Breeze and Miller (2011)), others found that only those subtests that have an academic version (i.e., Reading and Writing) were significantly correlated with students' performance in their first semester of studies (e.g., Kerstjens and Nery (2000)). Finally, there are also studies that did not find any correlations between the overall IELTS scores and students' academic achievement (e.g., Cotton and Conrow (1998)).

Studies that aimed to compare the predictive validity of IELTS and TOEFL, also produced inconsistent results. Johnson and Tweedie (2017) found that TOEFL is a better predictor of academic success than IELTS, but Hill, Storch, and Lynch (1999) found the opposite. If we consider that the first study was carried at a Canadian university and the latter one at an Australian university, we can conclude that the predictive validity of the tests is context-dependent (i.e., TOEFL has greater predictive validity in the Canadian context and IELTS in the Australian context). However, the reason why there might be a variation in, or even lack of, correlations between language tests scores and students' academic outcomes is that there are many other variables that can affect students' performance. Some of these variables are linguistic (e.g., the amount of ESL support students receive during their academic program, Cotton and Conrow (1998), while many others are non-linguistic (e.g., students' learning aptitude, motivation, financial status, time-management skills, and abilities to adjust to a new culture, (Cotton & Conrow, 1998; Hill et al., 1999).

While the predictive validity of a test is an important aspect of its criterion-related validity, another important aspect is the test's concurrent validity, which involves "comparisons with alternative estimates of assessees' knowledge, skills or abilities (such as teacher judgements.

or results of other assessments) that are obtained at around the same time" (Green, 2014)[2]. Therefore, it is necessary to compare the scores from IELTS with the scores from a test that has a longer history and is more widely accepted as a valid measure of students' English language proficiency (e.g., TOEFL). If we can make the same inferences about test takers' academic listening abilities based on each of the two tests, then this would be evidence of the validity of the IELTS listening subtest.

Geranpayeh (1994) compared the IELTS and TOEFL scores of two groups of Iranian graduate students. Relatively high correlations were reported for the total test scores of the first group and moderate correlations for the second group. Scores from some subsections were available for the subjects in the latter group, and the statistical analyses showed moderate correlations between the listening subtests.

More recently, a large scale study also compared the scores of students who took both IELTS and TOEFL (ETS, 2010). According to this research report, there are moderately high correlations between the overall IELTS and TOEFL test scores. Although the correlations for subsections were weaker than the correlations for the total test scores, the data analyses in this study also showed moderate correlations between the listening subtests see Table 1.

[1] A study with L1 Mandarin speakers showed that the majority of them (97%) preferred MCQ over cloze tests and open-ended questions in tests of listening comprehension, because they felt less pressured and nervous if they did not have to reconstruct the absorbed auditory message under time limitations (Cheng, 2004).

[2] The same concept ("the relationship of the test results to some other form of measurement such as other valid test scores") has been defined as *empirical validity* by Flowerdew and Miller (2012)

Table 1. Correlations between IELTS and TOEFL scores.

Study	N	L1	Listening	Speaking	Reading	Writing	Total
Geranpayeh (1994), Group A	113	Farsi	N/A	N/A	N/A	N/A	0.83
Geranpayeh (1994), Group B	103	Farsi	0.56	N/A	0.61	N/A	0.67
ETS (2010)	1153	various	0.63	0.57	0.68	0.44	0.73

In summary, previous research suggests that there are moderate correlations between scores on the IELTS and TOEFL listening subtests, which is a (moderate) indicator of the concurrent validity of the IELTS academic listening test.

8. Conclusion

IELTS is perhaps the most widely used test of English language proficiency after TOEFL, and this fact alone speaks of its face validity and potentially also of its criterion-related validity. As suggested by the present literature review, however, the design of the IELTS Academic listening subsection might not be flawless. Despite the moderate correlations between IELTS and TOEFL scores, it is important to keep in mind that concurrent validation suggests that "both measurements are measuring the same thing, but this is no guarantee that they are both measuring *the right thing* [italics added]" (Green, 2014). Therefore, although the TOEFL test might be a slightly better predictor of academic success than IELTS (Johnson & Tweedie, 2017), it is possible that both tests produce similar results because they are marked by similar methodological flaws. Indeed, Huang and Finn (2009) argue that "popular English proficiency tests such as TOEFL and IELTS are not adequate measures of ESOL[3] students' academic listening skills" (p. 49). Therefore, it might be the case that the construct of academic listening should be reconceptualised based on current research findings. Particularly, the lack of separate scores for academic and general listening skills might decrease the usefulness of the listening subtest and this matter should be addressed in future versions of the test. As it was demonstrated by the present critical review, the IELTS listening test's construct validity may benefit from a few other substantial modifications. The following improvements are suggested:

- (Greater) Focus on academic listening skills.
- Elimination of construct irrelevant skills (e.g., writing).
- Inclusion of at least some authentic materials.
- Inclusion of some input in video format.

As IELTS continues to evolve, it is imperative that future research investigate the usefulness of the listening module through empirical studies focussed not only on the concurrent but also on the predictive validity of the test. For example, scores from the IELTS listening test can be correlated with subsequent measures of students' abilities, such as their undergraduate course grades and even their self-perceptions of success. Exploring the validity of the listening module should be an ongoing process contributing to the overall improvement of the IELTS academic test. It is hoped that the present article has (re)ignited researchers' and educators' interest to continue this validation process.

References

Aryadoust, V. (2012). Differential item functioning in while-listening performance tests: The case of the International English Language Testing System (IELTS) listening module. *International Journal of Listening, 26*(1), 40-60.Available at: https://doi.org/10.1080/10904018.2012.639649.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice.* Oxford, UK: Oxford University Press.

Breeze, R., & Miller, P. (2011). Predictive validity of the IELTS listening test as an indicator of student coping ability in Spain. IELTS Research Reports (Vol. 12, pp. 1-34). Melbourne IDP: IELTS Australia and British Council.

Brunfaut, T. (2016). Assessing listening. In T. Dina & B. Jayanti (Eds.), Handbook of second language assessment (pp. 97-112). Berlin, Germany: De Gruyter Mouton.

Buck, G. (2001). *Assessing listening (Cambridge language assessment series).* Cambridge: Cambridge University Press.

Cheng, H. f. (2004). A comparison of multiple-choice and open-ended response formats for the assessment of listening proficiency in English. *Foreign Language Annals, 37*(4), 544-553.Available at: https://doi.org/10.1111/j.1944-9720.2004.tb02421.x.

Cooper, T. (2013). Can IELTS writing scores predict university performance? Comparing the use of lexical bundles in IELTS writing tests and first-year academic writing. *Stellenbosch Papers in Linguistics Plus, 42*(1), 63-79.Available at: https://doi.org/10.5842/42-0-155.

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. *IELTS Research Reports, 1*(4), 72-115.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213-238.

Cummins, J. (1980). The cross-lingual dimensions of language proficiency: Implications for bilingual education and the optimal age issue. *TESOL Quarterly, 14*(2), 175-187.Available at: https://doi.org/10.2307/3586312.

Douglas, D. (2000). *Assessing languages for specific purposes.* Cambridge: Cambridge University Press.

ETS. (2010). *Linking TOEFL iBT scores to IELTS scores – A research report.* Princeton, NJ: Educational Testing Service.

Flowerdew, J., & Miller, L. (2012). Assessing listening. In C. Coombe (Ed.), The Cambridge guide to second language assessment (pp. 225-233). New York: Cambridge University Press.

Geranpayeh, A. (1994). Are score comparisons across LP batteries justified? An IELTS-TOEFL comparability study. *Edinburgh Working Papers in Applied Linguistics, 5*, 50-65.

Geranpayeh, A., & Taylor, L. (2013). *Examining listening: Research and practice in assessing second-language listening (Studies in language testing; 35).* Cambridge; New York: Cambridge University Press.

Green, A. (2005). EAP study recommendations and score gains on the IELTS Academic Writing test. *Assessing writing, 10*(1), 44-60.Available at: https://doi.org/10.1016/j.asw.2005.02.002.

Green, A. (2014). *Exploring language assessment and testing: Language in action.* Abingdon, Oxon: Routledge.

Hemmati, F., & Ghaderi, E. (2014). The effect of four formats of multiple-choice questions on the listening comprehension of EFL learners. *Procedia-Social and Behavioral Sciences, 98*, 637-644.Available at: https://doi.org/10.1016/j.sbspro.2014.03.462.

Hill, K., Storch, N., & Lynch, B. (1999). A comparison of IELTS and TOEFL as predictors of academic success. *IELTS research reports, 2*(52-63).

Huang, J., & Finn, A. (2009). Academic listening tests for ESOL students: Availability, concerns, and solutions. *International Journal of Applied Educational Studies, 6*(1), 46-55.

IELTS. (2020). IELTS home of the English language test. Retrieved from: https://www.ielts.org/. [Accessed 29 December 2020].

[3] The abbreviation ESOL stands for "English to Speakers of Other Languages", which applies to both ESL (i.e., English as a Second Language) and EFL (English as a Foreign Language) contexts.

Ingram, D., & Bayliss, A. (2007). IELTS as a predictor of academic language performance, Part 1. *IELTS Research Reports, 7*(1), 137-204.
Johnson, R. C., & Tweedie, M. G. (2017). A comparison of IELTS, TOEFL, and EAP course results as predictors of English language learning success in an undergraduate nursing program. In C. Coombe et al. (Eds.), Language assessment in the Middle East and North Africa (pp. 36–53). Dubai: TESOL Arabia.
Kerstjens, M., & Nery, C. (2000). Predictive validity in the IELTS test: A study of the relationship between IELTS scores and students' subsequent academic performance. *IELTS Research Reports, 3*(4), 86-108.
Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), Handbook of second language acquisition (pp. 413-468). Boston: Academic Press.
Macaro, E., Vanderplank, R., & Graham, S. (2005). *A systematic review of the role of prior knowledge in unidirectional listening comprehension. In: Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.
McNamara, T. (2000). *Language testing (Oxford introductions to language study)*. Oxford: Oxford University Press.
Ockey, G. J. (2007). Construct implications of including still image or video in computer-based listening tests. *Language Testing, 24*(4), 517-537.Available at: https://doi.org/10.1177/0265532207080771.
Richards, J. C. (1983). Listening comprehension: Approach, design, procedure. *TESOL Quarterly, 17*(2), 219-240.Available at: https://doi.org/10.2307/3586651.
Samordna Opptak. (2020). The Norwegian universities and colleges admission service. Samordna opptak. Retrieved from: https://www.samordnaopptak.no/info.
Shohamy, E. (2001). *The power of tests: A critical perspective of the uses of language tests*. Harlow: Longman.
Taylor, L., & Geranpayeh, A. (2011). Assessing listening for academic purposes: Defining and operationalising the test construct. *Journal of English for Academic Purposes, 10*(2), 89-101.Available at: https://doi.org/10.1016/j.jeap.2011.03.002.
TOEFL. (2020). ETS TOEFL. Retrieved from: https://www.ets.org/toefl. [Accessed 30 December 2020].
Vandergrift, L. (2006). Second language listening: Listening ability or language proficiency? *The Modern Language Journal, 90*(1), 6-18.Available at: https://doi.org/10.1111/j.1540-4781.2006.00381.x.
Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language Teaching, 40*(3), 191-210.Available at: https://doi.org/10.1017/s0261444807004338.
Wagner, E. (2007). Are they watching? Test-taker viewing behavior during an L2 video listening test. *Language Learning & Technology, 11*(1), 67-86.