

AUTHENTICITY AND VALIDITY OF THE IELTS WRITING TEST AS PREDICTOR OF ACADEMIC PERFORMANCE

Arham Mauriyat

IKIP Siliwangi
arham@ikipsiliwangi.ac.id

Abstract

The International English Language Testing System (IELTS) has become one of the most widely used measurements of English proficiency in the world for academic, professional and migration purposes. For universities in particular, it is expected that applicants' IELTS scores closely reflect their actual ability in communicating and doing their assignments in English. This study examines the authenticity and predictive validity of the writing section in the IELTS Academic Module by reviewing relevant research on IELTS within the last two decades. In general, those studies have provided evidence that the IELTS writing test suffers from low authenticity and predictive validity, and is thus an inaccurate predictor of a candidate's performance in writing real-life academic tasks.

Keywords: IELTS, Academic Writing, Authenticity, Predictive Validity

INTRODUCTION

The International English Language Testing System (IELTS) assesses one's English proficiency based on the four language macroskills: listening, reading, writing and speaking (Ingram & Bayliss, 2007). Each of those skills is tested and graded separately with a score between 1 at the lowest and 9 at the highest, and the overall score is taken from the average of the four individual scores rounded up to the nearest half. According to the official IELTS website (<https://www.ielts.org>), over 10,000 educational institutions worldwide today accept IELTS as proof of a student candidate's proficiency, many of which even set certain IELTS band scores as minimum requirement for entry. It is thereby naturally envisaged that an applicant's IELTS scores, specifically the individual band scores, are valid representations of the applicant's actual level of proficiency in each language skill.

One section to receive particular attention in this respect is writing, in that a non-native English-speaking student with a writing score above a certain cut-point is expected to be able to write assignments well with a good command of written academic English (Dooey & Oliver, 2002). However, research has revealed that overseas students continued to face difficulties in writing their assignments in standardized English although they had successfully met or surpass the required IELTS score (Feast, 2002; Lloyd-Jones et al., 2012; Paul, 2007; Yen & Kuzma, 2009). These facts raise questions over the accuracy of those students' IELTS Academic Writing band scores in gauging their actual academic writing skills, which is what this research aims to investigate further.

For the purpose of this study, such accuracy is limited to the authenticity and predictive validity of the writing tasks in the IELTS Academic Module. In terms of language testing, authenticity is the extent to which a language test closely reflects the content and skills tested (Davies et al., 1999), while predictive validity refers to "the degree of correlation between the scores on a test and some other measure that the test is designed to predict" (Brown, 2000). The authenticity of

the IELTS academic writing test can thus be understood as the extent to which the writing tasks share the attributes of real-world academic writing assignments, and its predictive validity denotes how much the test results reflect a candidate's performance in future writing tasks.

The rationale for focusing on these two principles is the close relationship between them, in that “the more the tasks and contexts in which the language is tested resemble those of real-life, the more accurately is the language test likely to predict how the candidate will cope, at least linguistically, with real-life activities” (Ingram, 2003, pp. 18-19). This also implicates the major importance of predictive validity in designing tests for selection purposes (Messick, 1986). In particular, as stated by Brown and Abeywickrama (2010), predictive validity plays a central role in the use of large-scale standardized tests, such as IELTS, as gatekeepers. The purpose of this research is thereby operationalized into the following research questions:

1. To what extent do IELTS writing tasks correspond to real-world academic writing assignments?
2. To what extent do IELTS writing test results predict performance in real-world academic writing tasks?

Results of this study may have important implications on decisions taken by stakeholders of the test regarding the use of IELTS as entry requirement, preparations of future students before enrolment in university, and, to a greater extent, calls for revision of the IELTS Academic Writing Module.

Theoretical Framework

Authenticity in language testing has proved difficult to define, as can be seen from the numerous definitions proposed by experts to date. Shakibaei (2017, p. 224) has classified those definitions by their focus, ranging from those pertaining to language produced by native or “real” speakers to those relating to “the interaction between students and teachers.” Zheng and Iseni (2017) have identified two major approaches to understanding authenticity:

- a. The real-life approach introduced by Bachman (1990) examines to what degree real-life language tasks or performance is replicated in a test, specifically in its appearance, test-takers' perception toward it and how these influence their performance in it, and to what extent the test results predict the test-takers' performance in non-test situations. This view has been criticized for its lack of clarity over what or whose “real life” it refers to, alterations of real-life task features when adapted for testing, and the lack of distinction between one's language proficiency and language performance.
- b. The correspondence approach stems from Bachman and Palmer (1996) whose definition states that authenticity is “the degree of correspondence of the characteristics of a given language test task to the features of a target language task” (p. 23). This approach is based on a framework of test task characteristics that correspond to critical features of the target language use (TLU) being tested. Bachman and Palmer introduce the domain of TLU as “a set of specific language use tasks that the test taker is likely to encounter outside of the test itself” (p. 44). However, it is practically difficult to select which TLU to assess, particularly for general proficiency tests, and impossible to precisely replicate TLU tasks in testing situations which innately differ from real-life circumstances.

Zheng and Iseni's comments echoed Buck's (2001) assertion that it is practically impossible to fully replicate real-life tasks in testing. Nevertheless, Shomoossi and Tavakoli (2010) stress that the key role of authenticity in language tests remains a general consensus among linguists, which also establishes its immediate link to predictive validity by declaring that authenticity

pertains to “the validity of testees’ future performance in real-life situations” (p. 3). Predictive validity itself derives from the concept of validity as the core principle of assessment, to which other test properties serve as evidence (Weir, 2005). In general, validity in testing refers to “the degree to which a test measures what it claims, or purports, to be measuring” (Brown, 1996, p. 231) or “the degree with which the inferences based on test scores are meaningful, useful, and appropriate” (Brualdi, 1999, p. 1). Evidence of test validity is classified into three categories, namely content validity, construct validity, and criterion-related validity. According to Brown (1996), criterion-related validity directly relates to construct validity, denoting the degree of consistency of a test taker’s performance in two different assessments of the same construct, and is at times referred to as concurrent validity or predictive validity. Whereas concurrent validity looks at achievement in two such measurements held nearly simultaneously, predictive validity examines performance in separate times. In other words, predictive validity “measures how well a test predicts performance on an external criterion” (Davies et al., 1999, p. 149), or is “the degree to which scores on a test or assessment are related to performance on a criterion or gold standard assessment that is administered at some point in the future” (Frey, 2018).

The IELTS Academic Writing Test

The International English Language Testing System (IELTS) gauges the English language proficiency of people whose first language is not English and who mostly aim to reside, work or study at places where English is mainly used to communicate, both orally and in written (<https://www.ielts.org>; Stoyneff & Chapelle, 2005). The test is provided in two versions, which are the Academic Module for higher educational or professional purposes, and the General Training Module for immigration and other purposes with lower proficiency requirements. The test providers claim that both modules measure the four major skills of language (listening, reading, writing and speaking) validly, accurately and fairly, regardless of nationality, culture, gender or special needs. This is ensured by accounting for all standard varieties of native-speaker English and extensive trials of the test with people from diverse cultural backgrounds for appropriateness and fairness (IELTS Partners, n.d.-a).

The Academic Module includes samples of academic English, and its Writing section assesses the academic writing skills of IELTS candidates through two tasks which feature elements of typical university assignments. Task 1 examines “the ability to identify the most important and relevant information and trends in a graph, chart, table or diagram, and to give a well-organized overview of it using language accurately in an academic style,” while Task 2 gauges “the ability to present a clear, relevant, well-organized argument, giving evidence or examples to support ideas and use language accurately” (IELTS Partners, n.d.-d).

Authenticity and Predictive Validity of IELTS

A number of studies have been devoted to investigating the degree to which IELTS fulfils the principles of language assessment, including authenticity, either generally or specifically by section. Among the former, based on data collected from 180 participants comprising Iranian IELTS test takers and university teachers, Shakibaei (2017) concluded that IELTS lacks authenticity in several regards and demands specific techniques that are singularly relevant to the test. Others assert that IELTS is overly “Eurocentric” and requires a considerable extent of “world knowledge” from non-European candidates (Kabir, 2018, p. 84; Moore et al., 2012, p. 62). However, in a study on the washback and impact of IELTS (Saville & Hawkey, 2004), it was commented that the test’s authentic texts, both oral and written, are one of its strengths.

Meanwhile, most past studies on the predictive validity of IELTS revealed that students' IELTS scores have a weak correlation with their future academic performance. In fact, the IELTS test makers themselves actually declared that IELTS only assesses language proficiency and does not predict candidates' academic success (University of Cambridge Local Examinations Syndicate, 2009). The administrators stated that the IELTS academic test only purports to measure the readiness of test takers in entering universities where English is the language of instruction and implied that candidates with borderline scores may have to take extended language training, depending on their targeted disciplines (British Council et al., 2007).

This point was proved by Ferguson and White (1993) who discovered that higher IELTS scores actually have lower predictive validity, even though they reported a weakly positive relationship between the scores and actual academic grades. Subsequent research also found weak correlations, such as studies by Cotton and Conrow (1998), Dooey and Oliver (2002), Feast (2002), and Kerstjens and Nery (2000). Kerstjens and Nery highlighted that IELTS Writing scores had a lower correlation with academic performance than Reading scores in the case of higher education students. Meanwhile, the other studies found that some students who had good IELTS scores failed in their respective courses, whereas those who did not meet the required IELTS scores, but were admitted nonetheless, ultimately achieved passing grades (Cotton & Conrow, 1998; Dooey & Oliver, 2002). These findings led the researchers to conclude that high IELTS results did not guarantee academic success and that the language proficiency of students was only one of many factors in their academic progress.

More specifically, Cotton and Conrow (1998) discovered that IELTS scores do not correlate positively with language-related problems students face at university. This finding was corroborated by Feast (2002) who observed that a few students in a South Australian university did not display adequate linguistic behavior as expected from the entry requirement of their study programs. Similarly, by comparing the language proficiency of students from the University of Melbourne in both IELTS and the real academic world using the IELTS scoring scale, Ingram and Bayliss (2007) found that the proficiency levels of students in certain faculties did not satisfy the demands of their respective courses. This phenomenon was more evident in faculties that pay greater attention to language accuracy, namely Applied Language Studies and Medical Science. Several students in those faculties displayed insufficient language levels for their respective disciplines, while their lecturers expressed concern over the inadequate vocabulary and grammatical control of some of their students, casting doubt on the reliability of those students' IELTS scores. Moreover, Al Hajr (2014) highlighted the inconsistent results of studies on the predictive validity of IELTS, ascribing them to students' different fields of study.

METHOD

This study was conducted by review of studies on the authenticity and predictive validity of IELTS writing tasks in order to draw conclusions from comparisons of the characteristics of IELTS writing tasks and real-world university assignments and the proficiency displayed in responses to both types of tasks. Official IELTS publications and books discussing IELTS and large-scale standardized testing in general were also examined to provide more accurate and in-depth information on the test.

RESULTS AND DISCUSSION

Results

Publications on the authenticity of IELTS and large-scale standardized tests in general have pointed out significant differences between IELTS writing tasks and real-world university assignments. To maintain a high degree of authenticity, the first iteration of the IELTS academic test was offered with three specialized modules in Reading and Writing to cater for candidates' selected courses of study (Clapham & Alderson, 1997; Davies, 2008). However, limited success and administrative issues prompted the IELTS test makers to revise and generalize the test (Clapham, 1996), which consequently made it less authentic (Coombe et al., 2012) and thus disadvantageous for students from less relevant disciplines (Clark & Yu, 2020). Nevertheless, this trade-off between practicality and authenticity was deemed acceptable, since attempts to introduce subject-specific modules have failed and been left out of the current IELTS version (Charge & Taylor, 1997).

It is also notable that large-scale standardized tests such as IELTS employ much shorter time constraints, whereas real-world university tasks allow weeks or even months for completion (Coombe et al., 2012). In fact, the IELTS Academic Writing module only requires candidates to write a minimum total of 400 words in one hour (Stoyhoff & Chapelle, 2005), which may partly account for students' difficulties in writing more complex academic tasks (Paul, 2007). Furthermore, the two timed IELTS Writing tasks cannot adequately represent the large variety of real-world academic task genres (Green, 2007). Meanwhile, Mickan and Slater (2003) pointed to the fact that IELTS raters may actually grant high scores to candidates' responses even though they lack the qualities of academic essays, namely "transparent organization", "academic objectivity", and "impersonal voice" (p. 85). Wray and Pegg (2009) voiced another concern relating to the standardized nature of IELTS Writing in that elements of candidates' responses may be previously memorized, although this act may be regarded as good learning practice, such as in Chinese education culture. However, excessive memorization may substantially disguise test takers' true writing ability and raise the issue of plagiarism in real-world university contexts (Liu, 2005; Pennycook, 1996).

Looking at each IELTS Writing task in more detail, Uysal (2010) established that Task 1 more closely reflects authentic use of English in academic contexts than Task 2. Detailed discrepancies between the latter and actual university assignments were examined by Moore and Morton (2005), who stated that both corpora differ in "genre", "information source", "rhetorical function", and "object of enquiry" (p. 47). They observed that IELTS Writing Task 2 closely resembles the essay genre of academic tasks, but essentially differs in source of information. While university tasks mainly use external primary and/or secondary sources, IELTS Task 2 does not require test takers to do so and encourages them to utilise their pre-existing knowledge, which is largely socio-cultural in nature (Mickan et al., 2000), as suggested by the standard instruction of the task: "Give reasons for your answer and include any relevant examples from your own knowledge or experience" (IELTS Partners, n.d.-c). The problem with personal background knowledge as information is that it may be anecdotal and thus not represent full comprehension of the subject of the task, which is the authentic purpose of university assignments (Horowitz, 1991). IELTS Task 2 also simply caters for an immediate display of one's prior knowledge instead of the development of one's knowledge as expected in real-life university contexts (Green, 2007).

Another point of difference presented by Moore and Morton (2005) was that IELTS Task 2 and real-world academic assignments have distinct rhetorical functions, i.e. the purpose of discourse

as instructed by the task, e.g. evaluating, comparing, describing, recommending, etc. They found that although all university tasks and IELTS Task 2 question samples they studied contained a certain degree of evaluation, none of the latter showed summarizing and describing functions, which were prevalent in university tasks across diverse disciplines. Conversely, IELTS Task 2 frequently displayed the function of “hortation” (p. 58), or statement of a writer’s position concerning the necessity of a presented course of action, a rhetorical function that rarely occurred in the university task samples. Hortation is seen as a more practical approach than the analytical nature of real-world university tasks. The last investigated dimension of difference was “object of enquiry” (p. 60) i.e. the distinction between concrete facts (e.g. behavior, events, etc.) and abstract entities (e.g. ideas, beliefs, etc.) as discussed by the subject of the tasks. Whereas both categories of enquiry objects were identified in university tasks, the IELTS Task 2 samples dealt solely with physically observable phenomena. The researchers attributed difference in enquiry objects, as well as in information source and rhetorical functions, to the limitations of IELTS Writing which relies exclusively on the test takers’ background knowledge.

Summing up their analysis, Moore and Morton (2005) put forward other characteristics of IELTS Task 2, which treats writing as a spontaneous activity separate from reading with the purpose of giving a personal opinion on real-world phenomena supported by anecdotal evidence. These features are directly at odds with university assignments, which are rarely spontaneous, are more analytical, and allow giving opinions only when substantiated by valid and academically acknowledged non-anecdotal evidence, thus requiring extensive reading. The researchers further observed that the form of discourse generated by IELTS Task 2 is more appropriate for public non-academic genres than for university contexts. This point was confirmed by Coffin and Hewings (2004), who suggested that IELTS Task 2 has a distinctive style of argumentation that cannot be modelled on real-world academic writing genres, and by Cooper (2013), who found that the lexical bundles used in the opinion-based IELTS Task 2 responses pertain more closely to spoken discourse instead of typical written discourse in academic essays. Both findings were also attributed to the absence of available external sources in IELTS Task 2. These aspects and limitations of Task 2 further expose the lack of accuracy of IELTS Writing in measuring candidates’ comprehensive academic writing ability.

This lack of accuracy becomes more apparent when students display different levels of proficiency in writing real-life university tasks compared to their IELTS writing scores, exacerbated by the tendency of university courses to assign minimal IELTS score limits that are lower than recommended by IELTS administrators (Müller, 2015). For instance, a study at the business school of the University of Worcester observed that Chinese students continued to struggle in writing assignments despite achieving the minimum required IELTS Writing band score of 6.0, even though the score suggests adequate competence in academic English writing and the capability to study in English-speaking university contexts (Yen & Kuzma, 2009). Zhang and Mi (2010) also found that Chinese students in eight Australian universities expressed that they faced significant difficulties in writing academic tasks, citing cultural difference as a plausible factor that may have not been properly addressed by IELTS Writing tasks (Uysal, 2010) despite the test makers’ claim of ascertained fairness across culture (IELTS Partners, n.d.-a). In this respect, Kabir (2018), in line with the view that IELTS is Eurocentric, claims that IELTS Writing impairs non-European candidates and does not accurately measure their actual writing skills.

A similar occurrence was investigated in a follow-up to the research by Ingram and Bayliss (2007) which revealed that some students faced difficulties in producing adequate language for

academic tasks with greater complexity (Paul, 2007). Real-life university assignments involve specific orientations, interpretations, and approaches to the discourse of the tasks, all of which are aspects not covered in IELTS Writing due to the generalization of the test. Paul then concluded that IELTS scores were not adequate representation of students' ability to handle more sophisticated tasks specific to their respective disciplines, and stated that the correlation between IELTS Writing band scores and success in completing academic tasks remained unclear. Similarly, Clark and Yu (2020) uncovered that Japanese and Chinese Master's students in the UK found written university assignments much more sophisticated than IELTS writing tasks. In particular, they professed that academic writing tasks demand high levels of critical thinking, the use of readings as reference and the ability to present evidence and deliver a clear message, as opposed to IELTS writing which puts greater emphasis on lexical and grammatical range. Moreover, the standardized nature of the test is seen as a disadvantage for students from non-standard courses such as film and finance, the assignments of which greatly differ from IELTS tasks.

Another case of students with borderline IELTS scores was found in Cranfield University's postgraduate programs, where heads of study programs, or Course Directors, were concerned that their students' writing proficiency did not reach the standard level for writing Masters theses worthy of publication (Lloyd-Jones et al., 2012). Consequently, this lack of students' proficiency increased the burden of thesis supervisors in that they had to spend considerable time proofreading the theses in addition to assessing their contents. Because of this, Course Directors from the university's School of Management (SOM) even shared a common preference that applicants with borderline language proficiency levels be screened out rather than be given additional language support, an expense deemed irrelevant and wasteful. As a matter of fact, this view was also shared by Müller (2015) and justified by findings that some students' writing ability did not improve or even declined during their university study (O'Loughlin & Arkoudis, 2009). Meanwhile, a Course Director from the School of Engineering (SOE), which applied a different selection method, remained dissatisfied with the poor writing ability of non-native English-speaking students despite having raised the minimum required IELTS score from 6.5 to 7, whereas another SOE Course Director opined that the students' lack of ability in writing a thesis or an extensive report was the sole source of problems. Most Course Directors also did not take borderline IELTS scores for granted and sought additional evidence of students' writing proficiency through emails and instant messages, as they found it nearly impossible to devise a highly authentic assessment of writing skills. Overall, the position and course of action of the Course Directors reflected skepticism on the accuracy of IELTS in assessing students' ability in writing at thesis level.

CONCLUSION

In general, a considerable number of studies have examined the authenticity and predictive validity of IELTS, some particularly focusing on the writing test. Despite some exceptions, these studies have shown considerably consistent results regarding the correlation between university students' IELTS Writing scores and their performance in real-life academic tasks, and revealed that IELTS Writing lacks predictive validity and is not sufficiently authentic. Research has also found that students with borderline scores displayed lower writing proficiency in academic tasks compared to their IELTS writing scores. It can thus be concluded that IELTS writing scores are not accurate predictors of the ability of future university students in writing actual academic works.

Therefore, it is recommended that the minimum IELTS individual band score for writing score required to enroll in university programs be raised (Feast, 2002; Müller, 2015; Paul, 2007), although Feast acknowledged that such decision may eventually result in significant financial losses from the universities' point of view. Alternatively, student candidates should undertake an academic writing skills preparation program along with their preparation for IELTS (Moore & Morton, 2005), as such support may not be adequately available in universities, particularly for overseas postgraduate students (Clark & Yu, 2020). It may also be suggested that the IELTS Writing test, particularly the second task, be reverted to an older version which integrated writing and reading skills (Wallace, 1997) to increase authenticity by covering the skills of accessing external sources (Moore & Morton, 2005). In addition, Kabir (2018) suggests that the writing topics access not only the Western but also the Eastern side of 'world knowledge' to improve the test's fairness in regard to the candidates' diverse cultural backgrounds.

Research concerning this topic has generally used different measurements in comparing between performance in IELTS and in real-life academic tasks, with the exception of the study by Ingram and Bayliss (2007). Hence, in order to reduce bias and errors in analyzing the data, further studies should employ trained IELTS raters or their equivalents to examine academic essays using IELTS assessment criteria (IELTS Partners, n.d.-b). The essay genre is chosen due to its aforementioned similarity to IELTS writing tasks (Moore & Morton, 2005). Moreover, to augment inter-rater reliability, assessors of the university task samples should refer to clear English writing standards and conventions as presented by academic writing handbooks, such as those by Bailey (2011), Faigley (2015) and Oshima and Hogue (2006). Such rigorous effort is necessary to prompt interventions to make the IELTS test a more accurate assessment of future students' English language proficiency and to appropriately address students' language-related difficulties in writing real-world academic assignments. Nevertheless, further research needs to be undertaken to confirm the conclusion of this study, either by refining data measurement to increase the validity of correlations or by investigating similar populations in other contexts.

ACKNOWLEDGMENTS

All praise and gratitude belong to Allah, for without His guidance and consent all this humble work of mine would not have been made possible. I would also like to thank the TESOL Department faculty members and staff of Flinders University, South Australia, for their insights, tuition and facilities which have helped inform my study. Last but not least, I would like to thank my colleagues at IKIP Siliwangi for encouraging and facilitating the publication of this paper.

REFERENCES

- Al Hajr, F. (2014). The predictive validity of language assessment in a pre-university programme in two colleges in Oman: A mixed-method analysis. *International Multilingual Journal of Contemporary Research*, 2(2), 121-147.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests* (Vol. 1). Oxford University Press.
- Bailey, S. (2011). *Academic writing: A handbook for international students* (3rd ed.). Routledge.

- British Council, IDP: IELTS Australia, & University of Cambridge ESOL Examinations. (2007). *IELTS Handbook 2007*. University of Cambridge Local Examinations Syndicate.
- Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices* (2nd ed.). Pearson Education.
- Brown, J. D. (1996). *Testing in language programs*. Prentice Hall Regents.
- Brown, J. D. (2000). What is construct validity? *Shiken: JALT Testing & Evaluation SIG Newsletter*, 4(2), 8-12. http://hosted.jalt.org/test/bro_8.htm
- Brualdi, A. (1999). *Traditional and modern concepts of validity* (ED435714). ERIC/AE Digest. <https://files.eric.ed.gov/fulltext/ED435714.pdf>
- Buck, G. (2001). *Assessing listening*. Cambridge University Press.
- Charge, N., & Taylor, L. B. (1997). Recent developments in IELTS. *ELT Journal*, 51(4), 374-380.
- Clapham, C. (1996). *The development of IELTS: A study of the effect of background on reading comprehension* (Vol. 4). Cambridge University Press.
- Clapham, C., & Alderson, J. C. (Eds.). (1997). *Constructing and trialling the IELTS test*. University of Cambridge Local Examinations Syndicate.
- Clark, T., & Yu, G. (2020). Beyond the IELTS test: Chinese and Japanese postgraduate UK experiences. *International Journal of Bilingual Education and Bilingualism*, 1–19. <https://doi.org/10.1080/13670050.2020.1829538>
- Coffin, C., & Hewings, A. (2004). IELTS as preparation for tertiary writing: Distinctive interpersonal and textual strategies. In L. Ravelli & R. A. Ellis (Eds.), *Analysing academic writing: Contextualized frameworks* (pp. 153-171). Continuum.
- Coombe, C., Davidson, P., O'Sullivan, B., & Stoyloff, S. (2012). *The Cambridge guide to second language assessment*. Cambridge University Press.
- 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, 63-79. <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.
- Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., & McNamara, T. (1999). *Dictionary of language testing* (Vol. 7). Cambridge University Press.
- Davies, A. (2008). *Assessing academic English: Testing English proficiency, 1950-1989 - The IELTS solution*. Cambridge University Press.
- Dooley, P., & 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. <http://hdl.handle.net/20.500.11937/77607>
- Faigley, L. (2015). *The little Penguin handbook* (3rd ed.). Pearson Australia.
- Feast, V. (2002). The impact of IELTS scores on performance at university. *International Education Journal*, 3(4), 70-85.
- Ferguson, G., & White, E. (1993). *A predictive validity study of IELTS: A report commissioned by UCLES*. University of Edinburgh, Institute for Applied Language Studies.
- Frey, B. (2018). *The SAGE encyclopedia of educational research, measurement, and evaluation* (Vols. 1-4). SAGE Publications. <https://dx.doi.org/10.4135/9781506326139.n535>
- Green, A. (2007). *IELTS washback in context: Preparation for academic writing in higher education*. Cambridge University Press.
- Horowitz, D. (1991). ESL writing assessments: Contradictions and resolutions. In D. Hamp-Lyons (Ed.), *Assessing second language writing in academic contexts* (pp. 71-85). Ablex.
- IELTS Partners. (n.d.-a). *Ensuring quality and fairness*. IELTS. <https://www.ielts.org/about-the-test/ensuring-quality-and-fairness>

- IELTS Partners. (n.d.-b). *IELTS scoring in detail*. IELTS. <https://www.ielts.org/ielts-for-organisations/ielts-scoring-in-detail>
- IELTS Partners. (n.d.-c). *Sample test questions*. IELTS. <https://www.ielts.org/about-the-test/sample-test-questions>
- IELTS Partners. (n.d.-d). *Test format*. IELTS. <https://www.ielts.org/about-the-test/test-format>
- Ingram, D. E. (2003, July 10-12). *Towards more authenticity in language testing* [Paper presentation]. AFMLTA National Conference 2003: Languages Babble, Babel and Beyond, Brisbane, Australia.
- Ingram, D., & Bayliss, A. (2007). IELTS as a predictor of academic language performance, Part 1. *IELTS Research Reports*, 7. https://www.ielts.org/-/media/research-reports/ielts_rr_volume07_report3.ashx
- Kabir, S. M. (2018). IELTS Writing Test: Improving cardinal test criteria for the Bangladeshi context. *Journal of NELTA*, 23(1-2), 76-89. <https://doi.org/10.3126/nelta.v23i1-2.23352>
- 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, 85-108. https://www.ielts.org/-/media/research-reports/ielts_rr_volume03_report4.ashx
- Liu, D. (2005). Plagiarism in ESOL students: Is cultural conditioning truly the major culprit? *ELT Journal*, 59(3), 234-241.
- Lloyd-Jones, G., Neame, C., & Medaney, S. (2012). A multiple case study of the relationship between the indicators of students' English language competence on entry and students' academic progress at an international postgraduate university. *IELTS Research Reports*, 11, 129-184. https://www.ielts.org/-/media/research-reports/ielts_rr_volume11_report3.ashx
- Messick, S. (1986). The once and future issues of validity: Assessing the meaning and consequences of measurement. *ETS Research Report Series*, 1986(2), i-24. <https://doi.org/10.1002/j.2330-8516.1986.tb00185.x>
- Mickan, P., & Slater, S. (2003). Text analysis and the assessment of academic writing. *IELTS Research Reports*, 4, 59-88.
- Mickan, P., Slater, S., & Gibson, C. (2000). Study of response validity of the IELTS writing subtest. *IELTS Research Reports*, 3, 29-48. https://www.ielts.org/-/media/research-reports/ielts_rr_volume03_report2.ashx
- Moore, S., Mahony, P., & Stroupe, R. (2012). Perceptions of IELTS in Cambodia: A case study of test impact in a small developing country. *IELTS Research Reports*, 13. https://www.ielts.org/-/media/research-reports/ielts_rr_volume13_report6.ashx
- Moore, T., & Morton, J. (2005). Dimensions of difference: A comparison of university writing and IELTS writing. *Journal of English for Academic Purposes*, 4(1), 43-66. <https://doi.org/10.1016/j.jeap.2004.02.001>
- Müller, A. (2015). The differences in error rate and type between IELTS writing bands and their impact on academic workload. *Higher Education Research & Development*, 34(6), 1207-1219. <https://doi.org/10.1080/07294360.2015.1024627>
- O'Loughlin, K., & Arkoudis, S. (2009). Investigating IELTS exit score gains in higher education. *IELTS Research Reports*, 10, 1-86. https://www.ielts.org/-/media/research-reports/ielts_rr_volume10_report3.ashx
- Oshima, A., & Hogue, A. (2006). *Writing academic English* (4th ed.). Pearson Longman.
- Paul, A. (2007). IELTS as a predictor of academic language performance, Part 2. *IELTS Research Reports*, 7. https://www.ielts.org/-/media/research-reports/ielts_rr_volume07_report4.ashx
- Pennycook, A. (1996). Borrowing others' words: Text, ownership, memory, and plagiarism. *Tesol Quarterly*, 30(2), 201-230.

- Saville, N., & Hawkey, R. (2004). The IELTS impact study: Investigating washback on teaching materials. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 73-99). Lawrence Erlbaum Associates.
- Shakibaei, G. (2017). On the authenticity of IELTS academic tests. *Iranian Journal of Applied Language Studies*, 9, 223-254. <https://dx.doi.org/10.22111/ijals.2017.4231>
- Shomoossi, N., & Tavakoli, M. (2010). Authenticity and authentication in language testing: An operational perspective. *MJAL*, 2(1), 1-26.
- Stoynoff, S., & Chapelle, C. A. (2005). *ESOL tests and testing*. TESOL.
- University of Cambridge Local Examinations Syndicate. (2009). IELTS candidate performance in 2008. *Research Notes*, 36, 30-32.
- Uysal, H. H. (2010). A critical review of the IELTS writing test. *ELT Journal*, 64(3), 314-320. <https://doi.org/10.1093/elt/ccp026>
- Wallace, C. (1997). IELTS: Global implications of curriculum and materials design. *ELT Journal*, 51(4), 370-373. <https://doi.org/10.1093/elt/51.4.370>
- Weir, C. J. (2005). *Language testing and validation*. Palgrave MacMillan.
- Wray, A., & Pegg, C. (2009). The effect of memorized learning on the writing scores of Chinese IELTS test-takers. *IELTS Research Reports*, 9, 191-216. https://www.ielts.org/-/media/research-reports/ielts_rr_volume09_report5.ashx
- 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. <https://eprints.worc.ac.uk/id/eprint/811>
- Zhang, Y., & Mi, Y. (2010). Another look at the language difficulties of international students. *Journal of Studies in International Education*, 14(4), 371-388. <https://doi.org/10.1177/1028315309336031>
- Zheng, Y., & Iseni, A. (2017). Authenticity in language testing. *Anglisticum*, 6(8), 9-14. [http://dx.doi.org/10.0001/\(aj\).v6i](http://dx.doi.org/10.0001/(aj).v6i)