

RESEARCH

Open Access



'Remark or retake'? A study of candidate performance in IELTS and perceptions towards test failure

William S. Pearson

Correspondence: wpearson83@gmail.com
University of Exeter, St Luke's
Campus, Heavitree Road, Exeter, UK

Abstract

It is becoming increasingly important for individuals for whom English is a second language to demonstrate their linguistic credentials for academic, work and employment purposes. One option is to undertake International English Language Testing System (IELTS), which involves attempting to meet the linguistic entrance criteria set by a gatekeeping institution in the skills of listening, reading, writing, and speaking. Yet limited information is available in the public domain concerning the success of test-takers in meeting cut-off criteria set by the 10,000 or so organisations that utilise IELTS. The present study analyses the relationship between the test results and stated band score objectives of a cohort of 600 IELTS candidates, who shared their results on a social networking platform. It was uncovered that more test-takers failed to meet their band score goals ($n = 281$) than achieved them ($n = 245$), with many requiring high level linguistic goals to maximise their prospects in immigration systems. Thematic analysis was employed to explore the seldom-heard perspectives of the test-takers who missed their targets, and thereby 'failed' the IELTS test. Far more candidates held perspectives that constituted a rejection of their overall or sub-test score in comparison with those who were accepting of their results. Candidates' incredulity was notably acute concerning the accuracy of Speaking and Writing assessment, likely fuelled by a mistrust in single examiner marking and a lack of detailed test performance feedback to explain what went wrong.

Keywords: IELTS, Language testing, Thematic analysis

Introduction

Of growing importance to global cross-border population movements for work, academic and family purposes is the need for non-native English-speaking (NNES) individuals to obtain recognised and reputable evidence of English language proficiency (ELP) in order to meet the requirements of gatekeepers (test-users), such as universities, professional and training organisations, employers and immigration authorities. Evidence of sufficient ELP can be obtained by undertaking International English Language Testing System (IELTS), an English proficiency test developed and managed by the British Council, Cambridge Assessment English and IDP Australia. As with other high-stakes language tests, the outcomes of IELTS are condensed into single scores matched to brief descriptions of proficiency, requiring interpretation from both test-users and test-takers (O'Loughlin, 2011). While there exist considerable amounts of

research data offering insights into how test-takers perform in IELTS, virtually all studies examine achievement from an institutional perspective, investigating the sufficiency of IELTS results in relation to tertiary education readiness or academic outcomes. As such, the performance data and test interpretations of 'successful' IELTS test-takers prevail in the current academic literature.

The present study explores the test performance and perceptions of 600 IELTS test-takers who publicly shared their scores on a Facebook group orientated towards providing peer support to self-directed candidates-in-preparation. The study adopts a mixed methods design with sequential quantitative/qualitative data analysis (Plano Clark & Ivankova, 2016). First, candidates' overall and sub-test IELTS scores were examined in relation to the individuals' stated band score targets. Where apparent, underperformance was calculated with respect to the number and nature of sub-tests and the magnitude of the score deficit. The study adopted a critical language testing (CLT) dimension by exploring the perceptions of the 'losers' in the testing system (Shohamy, 2001), in other words candidates who failed to achieve their desired IELTS scores. Their interpretations of achieved test scores, elaborated in their public wall posts, were analysed thematically (Braun & Clarke, 2006; Fereday & Muir-Cochrane, 2006; Terry, Hayfield, Clarke, & Braun, 2017). To date, such candidates' voices have been among the least heard in academic research, despite being the participants in the testing system upon which IELTS exerts the most notable negative impact.

The study employed the inductive framework of perceived score acceptance or rejection to frame seven themes that emerged from candidates' shared wall posts: (1) perceptions of unreliable assessment, (2) disbelief towards test results based on perceived performance, (3) claims of test unfairness, (4) acknowledgement of poor/lack of preparation, (5) speculation of mistakes made in the test, (6) discussion of affective responses and (7) outlining of difficulties encountered.

Literature review

IELTS scores as linguistic entrance requirements

Founded in 1989 and undergoing a number of minor modifications in its lifespan since (see Davies, 2007), IELTS is a high-stakes, high impact test of English that assesses a candidate's proficiency across the four skills of listening, reading, writing and speaking. The test features Academic and General Training variants, the former undertaken for admission to tertiary education, the latter usually for immigration purposes. The test is characterised by a general proficiency theoretical model (Davies, 2007; Quaid, 2018), meaning underlying the test is a belief in 'some varying technically analysable, but fundamentally indivisible body of language knowledge within each test-taker, and therefore, individuals can be ranked on the basis of this knowledge' (Quaid, 2018, p. 3). Such knowledge is constitutive of a stable proficiency construct that exists externally to test-takers and other participants in the assessment process (Fulcher, 2014). The test features no structural or functional syllabus to sample (Quaid, 2018), while performance in the test's tasks are assumed to generalise to the real world.

Test-taker performance in IELTS is condensed and simplified into a single overall and four componential scores for ease and efficiency of interpretation by various stakeholders (O'Loughlin, 2011). Candidates' raw totals (out of 40) in Listening and Reading are

equated to a *band score* between one and nine (and half bands). In Speaking and Writing, band scores from one to nine are awarded by an examiner in four respective sub-criteria (IELTS, 2007), with an overall average calculated which is, if necessary, rounded to the nearest 0.5 bands. The two Writing tasks are marked by different examiners resulting in an awarded band score that constitutes an average of the two, weighted towards the longer Task 2. Finally, an overall score calculated as an average of the four sub-tests is also given, matched to a short description of proficiency provided by IELTS (IELTS, 2014). These five numbers (the *score profile*) constitute the only performance information candidates receive from the test.

Test-users interpret candidates' test results principally in relation to established cut-off scores that are deemed by the institution to predict sufficient linguistic proficiency for the context in which the individuals will operate as second language users. Concerning academic settings, studies have uncovered that bands 6.0 and 6.5 (denoting a 'competent user') overall are most typical as minimum entry requirements onto undergraduate or postgraduate courses (Hyatt & Brooks, 2009; Thorpe, Snell, Davey-Evans, & Talman, 2017). For some institutions, band 6.0 is interpreted as a red line, below which academic failure rates are thought to accumulate at an unacceptable rate (Breeze & Miller, 2008; Ferguson & White, 1998). To guard against the admission of candidates with *jagged score profiles*, a phenomenon defined as performance in one sub-test that is inconsistent with other modules (Green, 2004), institutions usually specify a minimum level of performance in all four sub-tests (O'Loughlin, 2011), often 5.5 to 6.0 for academic programmes.

The use of General Training IELTS scores to regulate legal immigration in countries where it is accepted, primarily the UK, Canada, Australia and New Zealand (Merrifield, 2012), is far less researched in comparison. As in higher education settings, cut-off scores are implemented, varying in accordance with factors such as the extent of the right of the individual to remain, the requirement for financial resources, whether a migrant is deemed highly skilled and the possible investment they may make in the country. In addition, since immigration is perceived by some citizens as a threat to their lifestyle and safety, entrance scores are also likely contingent on the contemporary political climate of a state. Where IELTS score use for immigration purposes differs compared with tertiary education admission is that in countries which operate a points-based visa system, notably Australia and Canada, obtaining a visa or permanent residency is more achievable the higher the IELTS score (Department of Home Affairs, 2019; Government of Canada, 2019). Thus, many General Training candidates may be incentivised to retake the test in order to offset deficits in other aspects of their application.

Organisations' IELTS entrance requirements may not be indicative of careful or considered standard setting. It has been reported that the linguistic entry requirements of some higher education institutions are not always established following a rigorous and thorough process (Chalhoub-Deville & Turner, 2000) nor monitored and reviewed regularly (O'Loughlin, 2011). The IELTS partners make non-binding recommendations for the interpretation of scores relative to the linguistic demands of academic or non-academic education programmes, summarised in Table 1, but offer no (public) advice for score setting concerning immigration. The simplicity and efficiency in which IELTS test scores can be interpreted means they are highly valued by admissions personnel (Hyatt & Brooks, 2009), even to the point of being perceived largely in 'pass/fail' terms (Coleman, Starfield, & Hagan, 2003; O'Loughlin, 2011), contradicting advice from IELTS which

Table 1 Guidance on acceptability of IELTS test scores for academic and training courses

Band score	Linguistically demanding academic courses	Linguistically less demanding academic courses	Linguistically demanding training courses	Linguistically less demanding training courses
7.5–9.0	Acceptable	Acceptable	Acceptable	Acceptable
7.0	Probably acceptable	Acceptable	Acceptable	Acceptable
6.5	English study needed	Probably acceptable	Acceptable	Acceptable
6.0	English study needed	English study needed	Probably acceptable	Acceptable
5.5	English study needed	English study needed	English study needed	Probably acceptable

Adapted from (IELTS, 2014)

encourages organisations to contextualise results in terms of the candidate’s ‘age and motivation, educational and cultural background, first language and language learning history’ (IELTS, 2007, p. 5). Consequently, underperformance in the IELTS test, be it for academic or immigration purposes, has a finite quality to it, although there exist (costly) appeal procedures and few limitations on retaking the test.

Candidate performance in the IELTS test

Increasing amounts of candidate test performance data is available through sponsored and independent research. The IELTS website offers the most detailed and up-to-date information and is the only source of test result data for the entire global candidature (although only 2017 data are available). Average performance overall and in the four sub-skills can be obtained from the breakdown of band score results according to test-takers’ gender and the 40 most common first languages, as shown in Table 2. When calculated from the gender data—a more complete dataset—an overall band score of 6.03 was achieved in 2017 in the academic test by the global candidature. Listening is the sub-test in which academic test-takers performed best in (6.22), with Writing the worst (5.61). Candidate performance calculated from an average of the 40 most commonly spoken first languages (L1 s) of test-takers indicated higher results overall and in each of the sub-tests, noticeably in Speaking (+ 0.42 of a band). This appears to show speakers of minority languages not in the presented data pull down these averages.

Candidates tend to perform better in the General Training IELTS test, particularly when comparing band scores calculated from the gender data. General Training test-takers’ performance in Writing exceeded their Academic counterparts by 0.49 of a band in 2017, which may be attributed to the contrasting task demands of writing a letter ‘to get

Table 2 Overview of global candidature performance in IELTS in 2017, measured in band scores

	Overall	Listening	Reading	Writing	Speaking
Average of gender					
Academic	6.03	6.22	6.10	5.61	5.93
General	6.47	6.64	6.26	6.10	6.60
Average of L1					
Academic	6.35	6.57	6.37	5.86	6.35
General	6.41	6.54	6.21	6.06	6.59

something done' versus reporting on graphical information, which varies according to the scope of information presented and the visualisation of the data. The test format in Reading also differs, with the General Training variant containing five or six passages, shorter in length compared to the Academic test's three passages, which are also more complex in topic and structure (IELTS, 2014). However, when calculated from the 40 most common L1 s, General Training examinees performed worse than Academic test-takers in Reading, a phenomenon uncovered in Hawkey's (2006) impact study. While providing an important cross-sectional snapshot of the global candidature's performance, absent from the data are statistics relating to test repeaters, which Hamid (2016) reports could constitute a substantial proportion of all test-takers, as well as measures of test performance in the context of individuals' personal band score goals.

A range of studies have reported on the test performance of a cohort of Academic IELTS candidates for the purpose of investigating IELTS' predictive validity (e.g. Arrigoni & Clark, 2015; Breeze & Miller, 2008; Humphreys et al., 2012; Ingram & Bayliss, 2007; Kerstjens & Nery, 2000; Lloyd-Jones, Neame, & Medaney, 2007). IELTS test results featured in predictive validity studies are usually investigated from an institutional perspective, i.e. in the context of a university's ELP cut-off scores and cross-sectional or accumulated average academic results. Such research can illuminate the extent to which individual students are linguistically prepared to undertake higher education, particularly when the score profiles of candidates are provided (Breeze & Miller, 2008; Dooley & Oliver, 2002; Ingram & Bayliss, 2007; Lloyd-Jones et al., 2007). Nevertheless, utilising predictive validity research to gain insights into IELTS test performance is limited by the often small and truncated samples, the latter due to applicants not meeting institutional ELP requirements being screened out (Arrigoni & Clark, 2015; Ferguson & White, 1998). As such, the performance of test-takers who do not meet institutional cut-off scores are rarely reported in the literature.

Test-takers' perspectives on their achieved IELTS scores

Research into NNES students' perspectives of their IELTS scores has focused on perceptions of the appropriacy of their chosen institution's entry score(s) (Coleman et al., 2003; Kerstjens & Nery, 2000), students' coping mechanisms utilised on tertiary programmes (Breeze & Miller, 2008; Kerstjens & Nery, 2000) and the perceptions of ELP development in an English language support programme relative to achieved results in IELTS (Humphreys et al., 2012). Coleman et al. (2003) surveyed the beliefs of post-test candidates enrolled on tertiary programmes regarding the appropriateness of their required entry scores across three institutions in Australia, China and the UK. They found that 54.5% of the cohort of 429 students agreed or strongly agreed with the proposition that the cut-off scores required by the university accurately reflected the required levels of English language proficiency necessary to succeed at university (of which 74.7% fell within the 5.5–6.5 range). For Kerstjens and Nery's (2000) small sample of 16 students, no clear relationship was uncovered between students' perceptions of their ELP sufficiency and institutional cut-off scores. One explanation, provided by Humphreys et al. (2012) is that students are unable to accurately self-report their level of ELP. As a corollary, the ability to interpret ELP appears even more challenging for students preparing for

IELTS, since they are expected to effectively perceive their ELP in relation to their institution's cut-off requirements.

While research that samples students' perspectives on the validity of IELTS entrance scores for academic study offers occasional insights into test-takers' sense-making of scores, its worth is limited in a number of regards. By virtue of being enrolled at tertiary institutions, candidates have achieved their goals in IELTS and are, thus, more likely to offer positive interpretations of the test. Similarly, as the IELTS test is often temporally distant at the time of data collection, score interpretations are more likely to be framed solely in terms of sufficiency for academic study. While an important and valid line of inquiry itself, candidates' perspectives of the test are certain to evolve over the course of their degree programme experiences. Finally, it must be stressed that such a line of inquiry addresses interpretations of test scores within the notion of score sufficiency, which is one of a number of ways in which the interpretation of scores in such a high-stakes test can be framed.

IELTS test-takers' perspectives of the scores they achieve have rarely been the focus of research, especially those individuals who miss their band score targets. Candidates' interpretations are likely to be underpinned by a range of factors that are not a concern for test-users, and hence are beyond the scope of institutionally orientated predictive validity research. These include the personal experiences of undertaking IELTS, the extent to which candidates believe their test results are a true and fair representation of their ELP, their affective responses to receiving high-stakes assessment information and possible explanations of underperformance, all of which are possible to quickly share online to a mass international audience. This appears to be a worrying omission, since as Hamp-Lyons (2000) stresses, it is the individual test candidates across the world whose voices are heard the least yet have the most to gain or lose in the testing system.

Notably omitted are the views of 'failed' test-takers, who constitute a notable proportion of the global candidature (Hamid, 2016). Under a CLT perspective (Shohamy, 2001), IELTS represents an instrument of power, potentially blocking these individuals' academic, migration or work ambitions. Proponents of CLT argue that the social dimensions and consequences of high-stakes tests must be monitored, and if necessary challenged, in order to mitigate exclusionary or discriminatory test outcomes (Shohamy, 2001), such as the propensity of high-stakes testing to drive social inequality through favouring individuals with greater economic means (Templer, 2004). Crucial in this endeavour is the need to consider a diverse group of voices, notably the losers as well as the winners in the testing system. From a more practical perspective, the views of test underperformers must be garnered to provide a more complete research base into students' attitudes towards IELTS, since candidates successfully admitted into university or through an immigration system are likely to hold more positive dispositions towards IELTS.

In light of the current limited evidence of candidate perspectives of IELTS test scores, the present study seeks to generate knowledge concerning how individuals who have taken the IELTS test perceive their scores. The present study aims to answer the following questions:

1. How do IELTS test-takers sharing their scores on a public Facebook group perform in the test relative to their chosen target band scores?
2. What perceptions do the candidates who underperform relative to their band score goals hold towards their test results?

Method

Settings

The study analyses data drawn from a public IELTS-orientated Facebook group constituting approximately 293,000 members (as of March 2019) who are preparing for IELTS. The Facebook group encompasses an online sphere where individuals can gather, interact and seek and share ideas (Ahern, Feller, & Nagle, 2016; Pi, Chou, & Liao, 2013) on the theme of IELTS. Prospective and veteran test-takers from disparate locations around the world (though with a high proportion of Asian candidates) utilise the functionality of this and other IELTS-themed Facebook groups—wall posts, replies and sub replies and private chat—to obtain and disseminate test tips and preparation advice, locate partners for synchronous online speaking practice, share preparatory content and seek and provide feedback on practice writing compositions (Pearson, 2018). Perusal of a typical day's wall posts, especially around the dates IELTS test results are transmitted reveals that score sharing is a common activity undertaken by the group's participants, the nature and purposes of which are explored in this study.

Sharing the results of a high-stakes assessment, not only IELTS, on social media platforms such as Facebook and Twitter represents the evolution of the 'exam post-mortem' from the sphere of the private conversation in the corridors of an exam hall to a mass international discussion (Lebus, 2016). The development of social networking interfaces, allowing users to convey their perspectives through emojis and memes as well as written text, has fundamentally altered the nature of high-stakes assessments. The phenomenon of trending hashtags allows for conversations about test questions to rapidly become national or even global in scope (Sutch & Klir, 2017). Along with national exam boards, the IELTS co-owners must establish controls to ensure the security of their assessments as well as implement measures to guarantee insider information transmitted online across time zones does not prejudice test fairness (Sutch & Klir, 2017). These platforms would appear to empower IELTS test candidates, providing a convenient and efficient locale in which disparately located individuals can dissect aspects of the test, mitigating the potential isolation of the individual, on-demand nature of IELTS preparation and test-taking. Nevertheless, there is very little research on the role of social media in language test preparation, and only the present author's thematic analysis located in an IELTS context (Pearson, 2018).

Data collection

A sample of 600 wall posts that featured users publicly disseminating their test results was extracted from the IELTS Facebook group utilising NVivo's NCapture functionality. All posts that were retrieved originated from the period March 2016 to March 2019 using the search term 'results' in order to generate a selection of relevant hits. Searching was preferable to browsing since it was deemed too time consuming to sift through the myriad of posts that did not feature score sharing. This meant that if candidates

had not employed the term ‘results’ in their wall post, they were omitted from the study. Two forms of information were collected and imported into NVivo; the candidate’s test score and any accompanying written commentary. Test results that were communicated as either a screenshot image from an official IELTS communication or typed by the test-taker in the post were included. While screenshotted scores of official communiqués offered greater authenticity, and hence validity, there appeared to be no obvious motivation for test-takers to share inauthentic scores through typed posts.

Data analysis

The researcher analysed each test score and accompanying information quantitatively using a deductive coding scheme, extracting the information into Microsoft Excel according to the variables presented in Additional file 1. All complete score profiles (overall band score plus the four sub-test scores) were included, and none discounted based on the score achieved. Where test results for the four skills were presented without an overall score (9% of all posts), a spreadsheet formula was used to calculate an average, which was then rounded to the nearest 0.5 (of a band). The outlined coding scheme facilitated the emergence of a dataset that could be analysed quantitatively in order to generate global insights into the nature of score sharing on the IELTS Facebook group. A range of simple statistical measures such as averages and frequency counts were generated for each band score (and half score) overall and for the four sub-tests, allowing for the creation of a score profile for the cohort. For the other variables, frequency counts (K) and percentage calculations in relation to the whole cohort were adopted as the means of quantitative data analysis.

In addition to the quantitative analysis of individuals’ IELTS test score data, qualitative data in the form of the utterances contained in users’ wall posts were analysed. An inductive thematic approach was adopted based on identifying and coding patterns within the data (Braun & Clarke, 2006; Fereday & Muir-Cochrane, 2006; Terry et al., 2017). A list of anticipated themes was drawn up, which were applied to a sample set of 100 posts. The posts were read and re-read by the researcher, leading to most themes being refined, some being abolished and new themes being created. This eventually resulted in a set of seven themes subsumed under three superordinate themes (Additional file 1). The enhanced coding scheme was then applied to the full dataset, with a number of smaller adaptations made to some of the codes as the coding progressed. Extracts from the participants are used as sources of evidence both illustratively and analytically (Terry et al., 2017). The former denotes the use of quotations to exemplify key elements of the analytic narrative, while the latter embodies the double hermeneutic tradition of the dual layering of participant/researcher perceptions of a phenomenon as the basis for knowledge claims (Terry et al., 2017). The quantitative and qualitative strands of data were analysed sequentially, with the qualitative data used to generate knowledge (perceptions of test performance) that complemented and elaborated the quantitative test result information (Plano Clark & Ivankova, 2016).

Ethical considerations

Social networking services offer an exciting sphere in which researchers can generate new and interesting insights into various phenomena in the social sciences. They are

also associated with ethical risks and pose challenges for the traditional pillars of ethically sound research; informed consent, confidentiality and anonymity (Wilkinson & Thelwall, 2011; Willis, 2017; Zimmer, 2010). The present study is framed within the documentary analysis tradition of social media research (Wilkinson & Thelwall, 2011). This involves the researcher adopting an archival approach to recording online data that is in a public locale, rather than the real-time observation of human subjects in the private sphere of their personal pages. In such research, informed consent for the extraction of the data is not usually required (Willis, 2017). In terms of confidentiality, only data that had been shared on the public page of the group was collected, while no identifying background information on the users was gathered. Participant quotations, utilised for illustrative analysis, were partially reformulated to help ensure they could not be matched to the individual who uttered them.

Results and discussion

Research question 1

Data were collected concerning which variant of the IELTS test (Academic versus General Training) candidates had undertaken and how many times they had attempted IELTS. Unfortunately, only approximately 18% of candidates reported either data. Of the 110 test-takers who stated the test variant, 70% specified they had undertaken the Academic test and 30% General Training. This ratio roughly corresponds to IELTS' global candidature data, where 78.1% of all test-takers attempting the test in 2017 undertook the Academic test and 21.9% General Training (IELTS, 2018). One hundred and eleven individuals stated whether the scores they shared were achieved on their first attempt and/or how many previous attempts they had employed to realise their results. From this figure, 61.3% of candidates reported that their score was attained on their first attempt, 25.2% on the second try and 6.3% needing three attempts. Interestingly, eight candidates reported that they had taken IELTS four times or more. This data appear to support Hamid's (2016) finding that a notable proportion of IELTS test-takers are veterans of the test.

Performance of the cohort in IELTS

Analysis of test results revealed that, on the whole, test-takers tended to perform well overall and in most of the sub-tests. It was notable that the mean overall band score was 7.07, denoting a 'good user' according to IELTS. This score is generally adequate for admission onto most undergraduate and postgraduate tertiary courses as well as being favourable for immigration purposes. It is also higher than the average candidate's performance according to IELTS' official data (2018) (6.03 in the Academic test and 6.47 in General Training) and when compared with studies investigating candidate performance in the test (Arrigoni & Clark, 2015; Hawkey, 2006; Thorpe et al., 2017). Table 3 presents the score profiles of the cohort. Across the four components of the test, performance in the receptive skills tests, and Listening in particular, was the highest. Surprisingly, band 8.5 was the most commonly achieved score in Listening, 0.5 of a band short of the maximum possible. In contrast, test-takers fared less well in the productive skills, notably Writing. Scores in Writing were, on average, 0.6 of a band lower than the aggregated overall and 1.08 bands lower than Listening. Nevertheless, 84.8% achieved at least a 6.0 in Writing, generally sufficient for admission onto undergraduate and postgraduate programmes.

Table 3 Profile of candidates' IELTS scores overall and by sub-test

Band ^a	Overall	Listening	Reading	Writing	Speaking
9.0	1	44	63	0	6
8.5	18	148	89	1	23
8.0	88	86	79	10	72
7.5	178	105	62	56	121
7.0	131	70	102	126	149
6.5	90	53	73	173	124
6.0	58	51	67	143	69
5.5	28	32	33	63	27
5.0	5	5	20	20	9
4.5	3	4	5	5	0

^aScores below 4.5 were omitted

Interestingly, the difference in performance between highest and lowest sub-test scores is more sizeable in the present study compared with official data and studies of IELTS test results. Variations between the highest and lowest sub-test components, as calculated from the gender data provided by IELTS, were 0.60 for females and 0.62 for males. In General Training, these were closer; 0.51 for females and 0.57 for males. It appears that many of the candidates sharing test scores on the Facebook group exhibited jagged score profiles, a phenomenon rarely explored in the academic literature on IELTS. The dataset was analysed to uncover how many candidates exhibited inconsistencies in performance across the four sub-tests, defined arbitrarily as a gap of two bands (since IELTS provides no formal definition of a jagged profile). Under this characterisation, a surprising 224 individuals exhibited a jagged profile across the four components. Of this figure, 55.8% demonstrated a score variance of 2.0 bands, 33.5% 2.5 bands and 10.7% a significant gap of three bands. Consequently, individuals indicating jagged profiles, potentially a substantial cohort of test-takers, may find themselves falling short of componential cut-off scores, despite excelling in certain skills, such as listening.

Candidate performance relative to band score goals

Analysis of the 600 wall posts revealed that the most prevalent lens through which IELTS test scores were framed was whether the test results met the aims designated by candidates' chosen (but rarely articulated) organisations. Surprisingly in light of the high levels of test performance, the cohort of candidates who shared results accompanied by an explicit statement or indicator that they had not met their target was higher (46.8%) than those who stated they had (40.8%) (with 12.3% making no explicit indication). This finding suggests that IELTS test-takers primarily interpret their achieved scores in finite pass or fail terms, mirroring the behaviours of administrators whose role is to analyse the sufficiency of IELTS scores for tertiary admission (Coleman et al., 2003). Few candidates exhibited expectations that a deficit in a particular sub-test could be offset by other evidence of ELP or negotiated with the designated institution. As such, many of these underperforming candidates appeared resigned to retaking the test in the future.

Of the 281 reported candidate score profiles that fell short of test-user requirements, only 22 involved instances where the overall band score requirement had not been met. Far more common was a deficit between actual and desired performance in one or more sub-tests. It was found that of the cohort who underperformed, 57.7% ‘failed’ IELTS by missing their desired target in one sub-test only. Further, 12.8% fell short of their target in two sub-tests, while the figures for underperformance in three or four sub-tests, an indicator that a candidate was underprepared for the test, both accounted for approximately 5–6%. Collectively, this totalled 398 instances of IELTS test components that the cohort underperformed in, although in 50 of these the specific component was not explicitly reported.

The 348 cases of sub-test failure were analysed to identify which components they occurred in and how significant the performance deficit was, with the results shown in Table 4. Writing represented by far the most common component in which candidates failed to achieve their targets (46.2%), followed by Speaking (17.3%), and Reading (14.1%). Writing was also the module which evinced the most significant underperformance, with 41 candidates missing their targets by 1.0 band. While this offers tentative evidence that test-takers struggle with the IELTS Writing test, it was also discovered that 117 underperformers required scores of 6.5 or 7.0 in Writing, considered demanding targets regardless of settings (Coleman et al., 2003; Hyatt & Brooks, 2009; Thorpe et al., 2017). Many candidates revealed that they were attempting to meet the score profiles designated for immigration, such as Listening 8.0, Reading 7.0, Writing 7.0 and Speaking 7.0, a configuration that drastically improves an individual’s prospects of permanent residency in Canada (Government of Canada, 2019). Unfortunately, satisfactory scores achieved in the three other sub-tests are considered invalid since test-users usually require cut-off scores to be achieved in a single sitting (Hamid, 2016). Thus, the implementation of minimum scores across the four sub-tests appears to represent a more serious barrier to test-takers than mean overall performance bands.

Table 4 Instances of candidate IELTS sub-test underperformance, correlating sub-test scores with the extent of band score deficit

	Listening	Reading	Writing	Speaking	Total
0.5 bands	17	29	87	45	178
1 band	14	8	41	12	75
1.5 bands	5	6	10	4	25
2 bands	1	4	2	1	8
2.5 bands	1	1	1	0	3
Not reported	0	8	43	7	58
Total	39	56	184	69	348

Research question 2

The majority of candidates who missed their targets exhibited one or more perceptions in the wall post they shared on the IELTS Facebook group. These diverse perspectives were condensed into seven themes and three superordinate themes, the frequency of which are outlined in Table 5. Concerningly, the most common superordinate theme expressed among candidates who underperformed in IELTS was a rejection of one or

Table 5 Occurrences of perceptions held towards IELTS test results of underperforming candidates

Superordinate themes/themes	K	%
Rejection of test results	126	70.0%
Unreliable assessment	93	73.8%
Disbelief of result based on perceptions of performance	30	23.8%
Accusation of test unfairness	3	2.4%
Acceptance of test results	35	19.4%
Under-preparation	14	40.0%
Mistakes made in the test	7	20.0%
Affective responses in the test	7	20.0%
Difficulties with the content of the test	7	20.0%
Unsure of sufficiency of achieved score	19	10.6%

more of their scores (70%). The most common reason behind this was a perception that the awarded score (mostly) in Speaking and/or Writing was inaccurate (73.8%), and that by applying for remarking by a different examiner the band score might increase. A further 23.8% of rejected results were characterised by a disbelief towards the given band score relative to the individual’s perception of test performance.

In contrast, there were far fewer instances of candidates sharing scores which they accepted (19.4%), constituted by four general perceptions, the most common of which was admittance of a lack of adequate preparation (40%). It was also evident that 10.6% of candidates neither explicitly accepted nor rejected their scores since they were unsure of their sufficiency, revealing that a number of individuals took IELTS without seeking to attain specific goals. The prevalence of candidates who rejected their scores versus those that were accepting may be explained by the purposes to which individuals utilised the platform. Test-takers incredulous towards their scores employed the group to seek advice from peers or merely to vent, whereas those who were accepting felt more assured in the preparation activities that could help them perform better next time, thus, perceiving little reward in sharing personal test result information. Regardless of the individuals’ dispositions towards their scores, rarely were task prompts or test questions dissected in any detail, likely reflecting the lack of repetition in and the unpredictability of future test content.

Perspectives of failed candidates who perceived their scores as illegitimate

A surprisingly high 93 candidates shared their scores for the purposes of inquiring into the potential of remarking to upgrade an aspect of their result, usually in Speaking or Writing. According to test regulations, a candidate may apply for remarking, known as Enquiry on Result (EoR) in one or more sub-tests and pay an administrative fee of approximately GBP 60, which is reimbursed should a score be upgraded (Pearson, 2018). While the high cost (relative to retaking the test) may dissuade many candidates, it is not possible for the original score(s) to be downgraded through an EoR. Since test-users generally require candidates to obtain the required minimum scores in one sitting, some test-takers perceive EoR checks as an easier and more cost-effective means in which to attempt to bridge a minor shortfall in performance. Frequently, test scores were presented along with an indication of the gap in performance, with the user requesting advice on

whether to undertake an EoR or retake the test; "My band scores are L8.5, R9, W6.5, S8, but I need 7 in writing. Should I ask for a remark or just sit the exam again?" As such, the platform allows some individuals to explore other candidates' experiences of remarking, to better inform their own decision whether to undertake an EoR (or retake the test). As a consequence, this and other IELTS-orientated Facebook groups serve as mass international locations in which the issue of test reliability can be discussed, with ramifications for how individuals who participate in online discussions as well as 'lurkers' who merely read posts perceive the test's fairness, validity and reliability.

Expressions of doubt towards band scores awarded in Speaking and Writing were also evinced in wall posts where test-takers expressed that they had received a score that was lower than what they felt they deserved, based on perceptions of their performance in the test:

I got 6.5 in writing but overall I got 8. I find this unbelievable. I do not think this shows my writing ability correctly.

Underlying the rejection of the awarded scores were a range of perspectives rooted in test-takers' beliefs regarding language proficiency and assessment. As exemplified in the above excerpt, jagged score profiles seemed to be considered as evidence of scorer unreliability by certain candidates, even though heavily contrasting Speaking and Writing scores necessitate a procedural remark by a second examiner. Additionally, some test-takers who, by admission, were test repeaters attributed stagnant or reduced test performance to unreliability:

Does anyone here know someone who has been successful at an EOR request? I am quite sure I would have scored more than 6 based on my previous attempts. So disappointed!

Others exhibited frustration with achieved scores in the context of the amount of time or effort that had been invested in preparation:

I need 7 in each test. I've had 5 private lessons and written several practice essays. I've improved my hand-writing, my grammatical range and structure. The topic was also OK.

Unfortunately, time invested in preparation is not a guarantee of successful test outcomes, a position IELTS itself has adopted through the removal of the guideline principle of a one band increase resulting from 200 h of study after 2002 (Green, 2004).

The tendency for advice-seeking on EoR checks underscores a surprising and worrying lack of trust in the reliability of the assessment of (mainly) IELTS Speaking and Writing. This finding contrasts with attitudinal research that has indicated candidates hold generally favourable perceptions towards trust and fairness in IELTS (Arrigoni & Clark, 2015; Coleman et al., 2003; Hawkey, 2006; Merrylees, 2003). Clearly, it matters whose attitudes are investigated and at what stage in the testing process, since individuals who have been successfully admitted onto tertiary programmes or achieved their migration ambitions are more likely to hold positive dispositions towards IELTS. Perceptions of unreliability likely

arose out of concerns with single-rater marking (Uysal, 2009). Yet data from IELTS examiner certification indicates a coefficient alpha of 0.83–0.86 for Speaking and 0.81–0.89 for Writing (IELTS, 2018), relatively high correlations for productive skills tests sampling candidates' general proficiency. Likely accentuating perceptions of unreliability is the lack of detailed feedback on performance in Speaking and Writing (Hamid & Hoang, 2018; Pearson, 2019), making it difficult for candidates to identify what went right and wrong and how to close gaps in performance. While both online and printed test preparation materials often provide textual models for learners to compare with their own rehearsal compositions, an absence of research means it is far from clear whether prospective test-takers are able to utilise these to enhance their band scores.

Perceptions of IELTS results considered legitimate

A smaller cohort of candidates ($n = 35$) utilised the platform to share unsatisfactory scores with the community which they had appeared to accept as legitimate. Underlying this behaviour were perceptions that are represented in four key themes, the most prevalent of which was poor or lack of preparation (40%). One candidate reflected on not undertaking preparatory tasks in conditions that simulated the test, a typical activity type in IELTS preparation classes (Hawkey, 2006; Hayes & Read, 2008): "During my preparation, I did not put myself under exam conditions, so I really struggled in reading and writing to keep to the time". It was also evinced that a lack of preparation arose out of a deliberate decision to undertake IELTS for diagnostic purposes:

I did not prepare for the IELTS test. I just took it to check my English level. Now I would like to know how much time I need to prepare and get a good score.

For this candidate, a diagnostic IELTS test was carried out in order to establish a benchmark to guide the focus and intensity of preparation. For others, the purpose appeared to be for confidence-building: "This was my IELTS results in first attempt. It's not satisfactory but without preparation it's not bad either. Feeling motivated to start preparing". This approach is likely limited to those test-takers with the financial means and willingness to invest more heavily in undertaking IELTS, known to be an expensive test to undertake globally (Pearson, 2019; Templer, 2004).

Affective responses to the test items and settings were perceived as responsible for undermining performance by seven candidates. Perhaps unsurprisingly, the most common of these reactions was anxiety, with one candidate noting: "I got so nervous during the speaking test. I know I could have done so much better". Previous studies have noted that undertaking IELTS can induce potentially debilitating anxiety in some candidates (Estaji & Tajeddin, 2012; Hamid & Hoang, 2018; Issitt, 2008). This may be particularly prevalent in the Speaking test, where there is pressure to perform in an intimidating face-to-face encounter with the examiner (Issitt, 2008). For another test-taker, the prospect of undertaking IELTS was sufficiently anxiety-inducing that it interfered with their ability to properly sleep prior to the test: "I was exhausted when I did the exam cos I did not sleep well for 2 days before the exam". It is evident that the high-stakes nature of IELTS is the principle contributing factor to inducing feelings of stress and nerves, which can overcome some candidates. Other contributory factors that elevate the pressure to perform include the cost of the test borne by the

candidate, which may prohibit repeated attempts, and the time and financial resources involved in travel to the test venue (perhaps requiring an overnight stay in a hotel as the Speaking test may be held on a different day).

Seven candidates who shared their test scores included reflections on the difficulties they experienced in the test which they perceived as contributing to underperformance. Interestingly, all but one instances of perceived difficulties concerned Reading. One candidate attributed test failure to a lack of preparation, stating: "Didn't get enough practice, especially in reading. The texts were unfamiliar, and I did not get to finish the questions". The pressure of only 60 min allocated to complete the Reading test was referenced by two individuals, with one noting, "I could not read the final passage so I had to guess half the questions". These findings are consistent with other studies that have uncovered candidates' perceptions of difficulty in IELTS (Coleman et al., 2003; Hamid & Hoang, 2018; Hawkey, 2006; Merrylees, 2003). In Hawkey's (2006) impact study, 49% of survey respondents noted that Reading was the most difficult module, with the time pressure and unfamiliar topics being cited as key causes of perceived difficulty, a finding mirrored in Cotton and Conrow (1998). In Hamid and Hoang's (2018) thematic investigation, the types of questions were also noted as being difficult, particularly the 'guessing game' of True/False/Not Given (p. 12). Nevertheless, it should be remembered that in the present study, candidates generally performed well in Reading, with 56 (out of 348) instances of underperformance in Reading, alone or in combination with other sub-test underperformance, costing candidates their results.

An interesting phenomenon as prevalent as affective responses and perceptions of test difficulties was reflections on possible test-taking errors which may have contributed to underperformance. One test-taker elaborated an error of judgement resulting from not following the demanding time pressures in the test:

I think I spent too much time doing [Writing] task one so by the time the invigilator announced there was 5 min left, I was only on my second paragraph on task 2. I had to finish the conclusion quickly and did not proof-read my text.

Such an error may result from a mismatch between test-takers' expectations of the type of graphical representation of data in Task 1 and that which is present in the test. Concern caused by a lack of proof-reading in Writing was also referenced by another candidate: "I wasn't able to read my task 2 when I finished writing because there wasn't any time". Proof-reading is beneficial in the Writing test as candidates may be able to notice and correct surface-level lexical or grammatical errors. However, the pencil-and-paper nature of IELTS means content/organization-level editing are not practical in the time allocated, thus it is unlikely a lack of proof-reading would be the sole cause of failure in IELTS. Two other candidates perceived writing too much had been a mistake in the Writing test, with one speculating: "the only problem with my essay may be that, I wrote 345 words". There is no upper limit on the word count of candidates' Task 2 compositions, and therefore text length itself is not explicitly penalised. In fact, research has shown that test-takers who produce lengthier scripts tend to be rated more highly (Mayor, Hewings, North, Swann, & Coffin, 2007; Riazi & Knox, 2013). This may reflect a perception held by examiners that longer texts evince greater linguistic fluency.

Beyond accepting and rejecting IELTS test results

In contrast to the majority of cohorts of candidates who rejected or accepted their IELTS test scores, a notable sub-section of the sample expressed uncertainty over the sufficiency of their test results. It is likely that these candidates took the test either for diagnostic purposes or without having researched the linguistic requirements of the relevant gatekeepers. It was evident that some candidates were not taking IELTS with a specific academic programme in mind: "Writing 5.5 but speaking 7. I want to know if the results could allow me to apply for Master's degree?" While it is important for prospective candidates to research the linguistic requirements prior to undertaking IELTS, it is also apparent that entry scores vary across institutions (Hyatt & Brooks, 2009; Thorpe et al., 2017), and that it is time-consuming and impractical for prospective applicants to verify the entry requirements of the over 130 universities in the UK, for example. Other test-takers exhibited a lack of knowledge of their entry scores in relation to immigration requirements: "I received this result.... can anyone tell...is this OK for Canadian immigration?" This issue is more complex since IELTS scores for migration constitute one element of an individual's application. Higher achievement in IELTS can offset deficiencies in other areas of an application, and vice-versa.

Conclusions

Limitations

The present study documented the performance of 600 IELTS candidates vis-à-vis the linguistic targets set by gatekeeping organisations for academic, emigration and other purposes. Test-takers utilised the Facebook group to share their scores with a vast international community of candidates. As such, the sample does not accurately typify the wider IELTS test candidature, being skewed towards individuals who utilise social networking platforms for preparation and to those who felt comfortable sharing their personal test results with a large anonymous online community. Thus, there are limitations to the generalisability of the findings beyond this context of test-takers. In addition, the study is restricted by the level of detail supplied in candidates' wall posts, with substantial gaps relating to the number of test attempts and which variant of the test had been undertaken. Questionnaire research could be utilised to generate a more comprehensive and complete dataset, featuring candidates' backgrounds, information concerning their test taking and how their stated band score goals arose.

Main conclusions

In this study, it was found that candidates who shared their IELTS scores publicly on the Facebook group tended on average to outperform both the global candidature (IELTS, 2018) and various cohorts of mostly academic test-takers explored in sponsored or independent research (Arrigoni & Clark, 2015; Hawkey, 2006; Thorpe et al., 2017). With an average overall score of 7.07 and 6.42–7.50 in the sub-tests, it is perhaps surprising that 46.8% of candidates stated they did not achieve the scores required by their respective institution. Writing was evinced as major stumbling block, accounting for 46.2% of instances where performance in one or more sub-tests was deemed insufficient. Further research is required to explore whether the reasons behind this are task, learner, or context related. It was evident that over a third of test-takers exhibited

a jagged score profile, an issue seldom explored in detail in the literature. Unfortunately for these candidates, excelling in skills such as listening and reading does not protect test results from being invalidated by often minor shortfalls (especially in writing), resulting in frustrating outcomes for many individuals.

The qualitative dimension of the study drew on the CLT tradition of exploring the interpretations of candidates whose IELTS test scores had fallen below their targets. A major rationale for this behaviour was the seeking of help, support or advice characterised by the candidate's rejection of (an aspect of) their achieved score profile. Most often, this perspective was evinced through candidates questioning the reliability of the scoring of Writing and/or Speaking and whether marking by a second examiner through the official process of EoR checking was warranted. This was explained by the high stakes involved, doubts over single examiner marking and the lack of feedback information. There were far fewer instances of test-takers presenting results framed in more accepting terms, perhaps because such individuals saw limited value in sharing their results. Rarely did candidates dissect in detail the content or semantics of the task prompts themselves, although it is acknowledged that such activities likely take place in wall posts that were filtered out under the study's inclusion criteria.

Implications

One implication of this study is that the reputation of IELTS (and other similar high stakes international ELP tests) for a valid, reliable and fair assessment could end up suffering damage through continuous, large-scale online discussions where individuals repeatedly question the accuracy of the marking. To mitigate this, IELTS cannot rest solely on claims of reliable marking contained in documentation targeted at researchers and test-users (e.g. IELTS, 2014, 2018), probably rarely seen by candidates. Rather, the test's co-owners need to explore additional means of providing insightful test preparation support online, perhaps through the use of their own group hosted on Facebook, moderated by staff employed by the co-owners.

It is understandably frustrating for test-takers to miss out on achieving the required cut-off scores in a single component of an expensive high-stakes test, particularly if performance is more than adequate in other modules. Highly condensed test result information, such as provided by IELTS, linked to a very short general description of performance makes bridging deficits in ELP a challenging task for test-takers. As a consequence, the partners could investigate ways in which to enhance the scope of test feedback, so test-takers do not become trapped in a cycle of costly repeated test failure. While textual models abound both online and in printed preparation materials, learners do not always make connections with current/future rehearsal compositions or in the test itself. The efficacy of their use in an IELTS preparation classroom context has not yet featured as a focus of research, a suggestion for future investigation.

There are also a number of implications of the study to the theory and practice of language testing research more generally. From a practical perspective, the study emphasised the important link between the results of attitudinal research into high-stakes ELP testing and who is being investigated and at what stage in the assessment process. These facets must be considered integral to the sampling decisions of future researchers of candidate perspectives on high-stakes ELP tests, requiring explicit discussion in the research output.

Theoretically, the present study largely framed candidate perspectives towards their IELTS test scores in terms of the perceived legitimacy of the results, or lack of. Existing research into language test fairness approaches the issue from the perspective of the validation undertaken by the test's designers (see Xi, 2010). As such, this study may be used by future researchers to develop a candidate-centred framework for analysing perceptions of the fairness of high-stakes ELP tests, for example by accounting for learner, test and contextual factors that constitute score acceptance and rejection.

Additional file

Additional file 1: Data Coding Scheme (DOCX 17 kb)

Abbreviations

CLT: Critical language testing; ELP: English language proficiency; EoR: Enquiry on result; IELTS: International English Language Testing System; L1: First language; NNES: Non-native English-speaking

Acknowledgements

Not applicable.

Author's contributions

The author read and approved the final manuscript.

Funding

The author received no specific funding for this work.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

The author declares that he has no competing interests.

Received: 14 June 2019 Accepted: 19 August 2019

Published online: 10 October 2019

References

- Ahern, L., Feller, J., & Nagle, T. (2016). Social media as a support for learning in universities: An empirical study of Facebook groups. *Journal of Decision Systems*, 25(1), 35–49. <https://doi.org/10.1080/12460125.2016.1187421>.
- Arrigoni, E., & Clark, V. (2015). *Investigating the appropriateness of IELTS cut-off scores for admissions and placement decisions at an English-medium university in Egypt* (IELTS research report series, no. 3). Canberra: IELTS Australia Pty limited Retrieved from <https://www.ielts.org/teaching-and-research/research-reports/online-series-2015-3>.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>.
- Breeze, R., & Miller, P. (2008). *Predictive validity of the IELTS Listening test as an indicator of student coping ability in Spain* (IELTS research reports volume 12). Canberra: IELTS Australia Pty Limited Retrieved from <https://www.ielts.org/teaching-and-research/research-reports/volume-12-report-5>.
- Chalhoub-Deville, M., & Turner, C. E. (2000). What to look for in ESL admission tests: Cambridge certificate exams, IELTS, and TOEFL. *System*, 28(4), 523–539. [https://doi.org/10.1016/S0346-251X\(00\)00036-1](https://doi.org/10.1016/S0346-251X(00)00036-1).
- Coleman, D., Starfield, S., & Hagan, A. (2003). *The attitudes of IELTS stakeholders: Student and staff perceptions of IELTS in Australian, UK and Chinese tertiary institutions* (IELTS research reports volume 5). Canberra: IELTS Australia Pty Limited Retrieved from <https://www.ielts.org/teaching-and-research/research-reports/volume-05-report-4>.
- 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 volume 1). Canberra: IELTS Australia Pty Limited Retrieved from <https://www.ielts.org/teaching-and-research/research-reports/volume-01-report-4>.
- Davies, A. (2007). Assessing academic English language proficiency: 40+ years of U.K. language tests. In J. Fox, M. Wesche, D. Bayliss, L. Cheng, C. E. Turner, & C. Doe (Eds.), *Language testing reconsidered* (pp. 73–86). Ottawa: University of Ottawa Press.
- Department of Home Affairs. (2019). Meeting our requirements: English language. Retrieved June 12, 2019, from <https://immi.homeaffairs.gov.au/help-support/meeting-our-requirements/english-language>
- 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.
- Estaji, M., & Tajeddin, Z. (2012). The learner factor in washback context: An empirical study investigating the washback of the IELTS academic writing test. *Language Testing in Asia*, 2(1), 5–25. <https://doi.org/10.1186/2229-0443-2-1-5>.
- Fereday, J., & Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International Journal of Qualitative Methods*, 5(1), 80–92. <https://doi.org/10.1177/160940690600500107>.
- Ferguson, G., & White, E. (1998). A small-scale study of predictive validity. *Melbourne Papers in Language Testing*, 7(2), 15–63.

- Fulcher, G. (2014). Philosophy and language testing. In A. J. Kunnan (Ed.), *The companion to language testing* (pp. 1431–1451). Hoboken: Wiley.
- Government of Canada. (2019). Language testing—Skilled immigrants (express entry). Retrieved June 12, 2019, from <https://www.canada.ca/en/immigration-refugees-citizenship/services/immigrate-canada/express-entry/documents/language-requirements/language-testing.html>
- Green, A. (2004). *Making the grade: Score gains on the IELTS Writing test* (research notes issue 16). Cambridge: Cambridge University Press Retrieved from <https://www.cambridgeenglish.org/images/23132-research-notes-16.pdf>.
- Hamid, M. O. (2016). Policies of global English tests: Test-takers' perspectives on the IELTS retake policy. *Discourse: Studies in the Cultural Politics of Education*, 37(3), 472–487. <https://doi.org/10.1080/01596306.2015.1061978>.
- Hamid, M. O., & Hoang, N. T. H. (2018). Humanising language testing. *TESL-EJ*, 22(1).
- Hamp-Lyons, L. (2000). Social, professional and individual responsibility in language testing. *System*, 28(4), 579–591. [https://doi.org/10.1016/S0346-251X\(00\)00039-7](https://doi.org/10.1016/S0346-251X(00)00039-7).
- Hawkey, R. (2006). *Impact theory and practice: Studies of the IELTS test and Progetto Lingue 2000* (studies in language testing 24). Cambridge: Cambridge University Press Retrieved from <http://www.cambridgeenglish.org/images/329230-studies-in-language-testing-volume-24.pdf>.
- Hayes, B., & Read, J. (2008). IELTS test preparation in New Zealand: Preparing students for the IELTS academic module. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 97–111). Mahwah: Lawrence Erlbaum Associates Publishers.
- Humphreys, P., Haugh, M., Fenton-Smith, B., Lobo, A., Michael, R., & Walkinshaw, I. (2012). *Tracking international students' English proficiency over the first semester of undergraduate study* (IELTS research report series, no. 1). Canberra: IELTS Australia Pty Limited Retrieved from <https://www.ielts.org/teaching-and-research/research-reports/online-series-2012-1>.
- Hyatt, D., & Brooks, G. (2009). *Investigating stakeholders' perceptions of IELTS as an entry requirement for higher education in the UK* (IELTS research reports volume 10). Canberra: IELTS Australia Pty Limited Retrieved from <https://www.ielts.org/teaching-and-research/research-reports/volume-10-report-1>.
- IELTS. (2007). *The IELTS handbook*. Cambridge: University of Cambridge Local Examinations Syndicate, The British Council, IDP Australia.
- IELTS. (2014). Guide for educational institutions, governments, professional bodies and commercial organisations. Retrieved from <https://www.ielts.org/-/media/publications/guide-for-institutions/ielts-guide-for-institutions-uk.ashx?la=en>
- IELTS. (2018). Test performance 2017. Retrieved May 22, 2019, from <https://www.ielts.org/teaching-and-research/test-performance>
- Ingram, D., & Bayliss, A. (2007). *IELTS as a predictor of academic language performance, part 1* (IELTS research reports volume 7). Canberra: IELTS Australia Pty Limited Retrieved from <https://www.ielts.org/teaching-and-research/research-reports/volume-07-report-3>.
- Issitt, S. (2008). Improving scores on the IELTS speaking test. *ELT Journal*, 62(2), 131–138. <https://doi.org/10.1093/elt/ccl055>.
- 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 volume 3). Canberra: IELTS Australia Pty Limited Retrieved from <https://www.ielts.org/teaching-and-research/research-reports/volume-03-report-4>.
- Lebus, S. (2016). Summer Series 2016—Postmortems, petitions and practicalities [Blog post]. Retrieved from <https://www.cambridgeassessment.org.uk/blog/summer-series-2016-postmortems-petitions-and-practicalities/>
- Lloyd-Jones, G., Neame, C., & Medaney, S. (2007). *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 volume 11). Canberra: IELTS Australia Pty Limited Retrieved from <https://www.ielts.org/teaching-and-research/research-reports/volume-11-report-3>.
- Mayor, B., Hewings, A., North, S., Swann, J., & Coffin, C. (2007). A linguistic analysis of Chinese and Greek L1 scripts for IELTS academic writing task 2. In L. Taylor & P. Falvey (Eds.), *IELTS collected papers: Research in speaking and writing assessment* (pp. 250–315). Cambridge: Cambridge University Press.
- Merrifield, G. (2012). *The use of IELTS for assessing immigration eligibility in Australia, New Zealand, Canada and the United Kingdom* (IELTS research report volume 13). Canberra: IELTS Australia Pty Limited Retrieved from <https://www.ielts.org/teaching-and-research/research-reports/volume-13-report-1>.
- Merrylees, B. (2003). *An impact study of two IELTS user groups: Candidates who sit the test for immigration purposes and candidates who sit the test for secondary education purposes* (IELTS research reports volume 4). Canberra: IELTS Australia Pty Limited Retrieved from <https://www.ielts.org/teaching-and-research/research-reports/volume-04-report-1>.
- O'Loughlin, K. (2011). The interpretation and use of proficiency test scores in university selection: How valid and ethical are they? *Language Assessment Quarterly*, 8(2), 146–160. <https://doi.org/10.1080/15434303.2011.564698>.
- Pearson, W. S. (2018). Utilising Facebook community groups for IELTS preparation: A thematic analysis. *The Asian EFL Journal*, 20(12.2), 5–59.
- Pearson, W. S. (2019). Critical perspectives on the IELTS test. *ELT Journal*, 73(2), 197–206. <https://doi.org/10.1093/elt/ccz006>.
- Pi, S.-M., Chou, C.-H., & Liao, H.-L. (2013). A study of Facebook groups members' knowledge sharing. *Computers in Human Behavior*, 29(5), 1971–1979. <https://doi.org/10.1016/j.chb.2013.04.019>.
- Plano Clark, V. L., & Ivankova, N. V. (2016). How to use mixed methods research?: Understanding the basic mixed methods designs. In *Mixed methods research: A guide to the field* (pp. 105–134). Thousand Oaks: SAGE Publications Ltd.
- Quaid, E. D. (2018). Reviewing the IELTS speaking test in East Asia: Theoretical and practice-based insights. *Language Testing in Asia*, 8(2). <https://doi.org/10.1186/s40468-018-0056-5>.
- Riazi, A. M., & Knox, J. S. (2013). *An investigation of the relations between test-takers' first language and the discourse of written performance on the IELTS Academic Writing Test, task 2* (IELTS research report series, no. 2). Canberra: IELTS Australia Pty Limited Retrieved from <https://www.ielts.org/teaching-and-research/research-reports/online-series-2013-2>.
- Shohamy, E. (2001). Democratic assessment as an alternative. *Language Testing*, 18(4), 373–391. <https://doi.org/10.1177/026553220101800404>.
- Sutch, T., & Klir, N. (2017). Tweeting about exams: Investigating the use of social media over the summer 2016 session. *Research Matters: A Cambridge Assessment Publication*, 23, 2–9 Retrieved from <https://www.cambridgeassessment.org.uk/images/375441-tweeting-about-exams-investigating-the-use-of-social-media-over-the-summer-2016-session.pdf>.
- Templer, B. (2004). High-stakes testing at high fees: Notes and queries on the international English proficiency assessment market. *The Journal of Critical Education Policy Studies*, 2(1), 189–226.

- Terry, G., Hayfield, N., Clarke, V., & Braun, V. (2017). Thematic analysis. In *The SAGE handbook of qualitative research in psychology* (pp. 17–36). London: SAGE Publications Ltd.
- Thorpe, A., Snell, M., Davey-Evans, S., & Talman, R. (2017). Improving the academic performance of non-native English-speaking students: The contribution of pre-sessional English language programmes. *Higher Education Quarterly*, 71(1), 5–32. <https://doi.org/10.1111/hequ.12109>.
- Uysal, H. H. (2009). A critical review of the IELTS writing test. *ELT Journal*, 64(3), 314–320. <https://doi.org/10.1093/elt/ccp026>.
- Wilkinson, D., & Thelwall, M. (2011). Researching personal information on the public web: Methods and ethics. *Social Science Computer Review*, 29(4), 387–401. <https://doi.org/10.1177/0894439310378979>.
- Willis, R. (2017). Observations online: Finding the ethical boundaries of Facebook research. *Research Ethics*, 15(1), 1–17. <https://doi.org/10.1177/1747016117740176>.
- Xi, X. (2010). How do we go about investigating test fairness? *Language Testing*, 27(2), 147–170. <https://doi.org/10.1177/0265532209349465>.
- Zimmer, M. (2010). "But the data is already public": On the ethics of research in Facebook. *Ethics and Information Technology*, 12(4), 313–325. <https://doi.org/10.1007/s10676-010-9227-5>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
