Halo Ali

Test Reviews

Colorado State University

The demand for proficiency tests has increased lately in Iraqi Kurdistan due to more opportunities for students to study further, specifically in higher education. This increase means that there will be more and more students who will be required to take proficiency tests such as TOEFL or IELTS. Language schools try to offer preparation classes for these proficiency tests so that students become familiar with the types of questions and the skills necessary to succeed in these tests. I plan to help students in my language school by providing detailed information about TOEFL iBT and IELTS Academic so that they can prepare for the more convenient one for them.

In this project, I review two proficiency tests: TOEFL iBT and IELTS Academic. I explore these tests for young adults who will be offered preparation classes for either test in an English school that I teach in Iraqi Kurdistan. These students mostly need the results of either test for admission in universities in higher education either locally or abroad. This review will help me in determining which of these two tests to recommend for my students. In addition, this review will broaden my knowledge about some test characteristics I should be knowledgeable about when given a test to review so that I can give advice to students based on the necessary characteristics of the test(s).

International English Language Testing System (IELTS):

| | |
|---|---|
| Publisher | British Council, Bridgewater House 58 Whitworth Street Manchester, M1 6BB United Kingdom Tel: +44 161 957 7755; ielts@britishcouncil.org; <br> IDP: IELTS Australia Level 8, 535 Bourke St Melbourne VIC 3000, Australia Tel: +61 3 9612 4400 ielts@idp.com; <br> Cambridge English Language Assessment, 1 Hills Road Cambridge, CB1 2EU United Kingdom: www.cambridgeenglish.org/ helpdesk; <br> IELTS USA 825 Colorado Boulevard Suite 221 Los Angeles, CA 90041 USA Tel: +1 323 255 2771 ielts@ieltsusa.org <br><br> Website: WWW.ielts.org |
| Publication date | 1989 |
| Target population | The IELTS Academic test is for candidates entering higher education or professions where English is the main language of communication. <br><br> (IELTS, http://www.ielts.org/) |
| Cost | varies from place to place. The test costs 235 USD in Colorado, USA but 265 USD in Iraqi |
| Test Purpose | IELTS assesses candidates' language proficiency who intend to study or work in an English-speaking environment. The IELTS Academic test is suitable for admission at undergraduate or postgraduate levels, and also for career purposes. It assesses a candidate's ability and preparedness to study or train in a setting where English is the main language. It also assesses some academic elements of language used in academia. IELTS results are used by a large number of academic institutions to select their students. <br><br> (IELTS handbook, 2017) |
| Test structure | The IELTS test has all 4 sections of listening, reading, writing and speaking. The whole test takes 2 hours 45 minutes. <br> Listening: This section takes about 30 minutes with another 10 minutes as transfer time. It has 40 questions. Different question types are used. There are 4 sections: Section 1 is a conversation between two people in a social interaction. Section 2 is a monologue in a typical social interaction. Section 3 is a conversation between up to four people in an educational or training setting. Section 4 is a monologue on an academic topic. Test takers can hear each listening task only once. <br> Reading: This section takes 60 minutes. There are 40 questions. Different question types are used. It has 3 sections. In Academic Reading, Each section has one long text. The reading texts are authentic and are derived from books, newspapers, etc. Texts can have materials such as diagrams, graphs or illustrations. A glossary is given for technical terms. <br> Writing: This section takes 60 minutes. It has 2 tasks. Academic Writing In Task 1, test takers are given a graph, table, chart or diagram and are required to describe, summarize or explain the information using their own words.  In Task 2, candidates are required to write an essay about a viewpoint, argument or issue. <br> Speaking: This section takes 11-14 minutes.  It has 3 parts. Part 1 is about general introduction on familiar topics (4-5 minutes). Part 2 (3-4 minutes) is about a specific topic |

| | |
|---|---|
| | that the test taker is given and he/she needs to discuss it. Part 3 Two-way discussion (4-5 minutes) that asks further questions related to the topic of Part 2.<br><br>(IELTS handbook, 2017) |
| Test<br><br>Scoring | The Overall Band Score is the average of the four component scores, rounded to the nearest whole or half band. The component scores are weighted equally.<br>The Listening section contains 40 questions. Each correct answer is awarded one mark. Scores out of 40 are converted to the IELTS nine-band scale. Scores are reported in whole and half bands. The Reading section contains 40 questions. Each correct answer is awarded one mark. Scores out of 40 are converted to the IELTS nine-band scale. Scores are reported in whole and half bands. The table below indicates the average number of marks required to achieve a particular band score in Listening and Academic Reading. |

| Listening | | Academic Reading | |
|---|---|---|---|
| **Band score** | **Raw score out of 40** | **Band score** | **Raw score out of 40** |
| 5 | 16 | 5 | 15 |
| 6 | 23 | 6 | 23 |
| 7 | 30 | 7 | 30 |
| 8 | 35 | 8 | 35 |

Writing: Examiners use assessment criteria to award a band score for each of the four criteria: Task Achievement (for Task 1), Task Response (for Task 2), Coherence and Cohesion, Lexical Resource and Grammatical Range and Accuracy. The criteria are weighted equally and the score on the task is the average.
 Speaking: Examiners use assessment criteria to award a band score for each of the four criteria: Fluency and Coherence, Lexical Resource, Grammatical Range and Accuracy, and Pronunciation.

(IELTS, http://www.ielts.org/)

| | |
|---|---|
| Statistical<br><br>distribution<br><br>of scores | The table below shows Mean and standard Deviation of Listening and Reading (2015): |

| Module | Mean | SD |
|---|---|---|
| **Listening** | 6.10 | 1.3 |
| **ACR** | 6.02 | 1.2 |

The table below shows the mean overall and individual band scores achieved by 2015 Academic test takers according to their gender.

**Mean band scores for female test takers**

|  | Listening | Reading | Writing | Speaking | OVERALL |
|---|---|---|---|---|---|
| Academic | 6.1 | 6.1 | 5.6 | 5.9 | 6.0 |

**Mean band scores for male test takers**

|  | Listening | Reading | Writing | Speaking | OVERALL |
|---|---|---|---|---|---|
| Academic | 5.9 | 6.0 | 5.4 | 5.8 | 5.8 |

(IELTS, http://www.ielts.org/)

| Standard error of Measurement (SEM) | The SEM needs to be analyzed in terms of the final band scores presented for Listening and Reading components (which are reported in half-bands). The website does not provide information on SEM for Listening and Speaking sections. |
|---|---|

| Module | SEM |
|---|---|
| Listening | 0.37 |
| Academic Reading | 0.38 |

(IELTS, http://www.ielts.org/)

| Evidence of reliability | The reliability of Listening and Reading tests is presented by using Cronbach's alpha, a reliability estimate which measures the internal consistency of the 40-item test. The following Listening and Reading material released in 2015 had sufficient candidate responses to estimate and report meaningful reliability values as follows: |
|---|---|

| Module | Alpha |
|---|---|
| Listening | 0.91 |
| Academic reading version | 0.90 |

We cannot present the reliability of the Writing and Speaking sections in the same way as the Reading/Listening sections since they do not have specific items. Instead, they are rated by experts based on in-depth descriptive criteria and rating scales. Experimental generalisability studies were carried out as part of the IELTS Speaking and Writing Revision Projects to investigate the reliability of ratings (Shaw, 2004). More recent G-studies based on examiner certification data showed coefficients of 0.83–0.86 for Speaking and 0.81–0.89 for Writing.

(IELTS, http://www.ielts.org/)

| Evidence of validity | Validity has several areas to be tested to show its presence in a test. One of them is when the test score matches with the test taker's outside ability. Kerstjens and Nery (2000) show a study between two groups' GPAs and IELTS scores. The correlation between the scores of the two are significant, which shows that the IELTS test is valid.<br>(Kerstjens, M. & Nety, C. (2000). Predictive Validity in the IELTS Test) |
|---|---|

Test of English as a Foreign language (TOEFL iBT)

| Publisher | Educational Testing Service (ETS)<br>Email: toefl@ets.org<br>Phone: 1-609-771-7100 or 1-877-863-3546 (Monday–Friday, 8am– 7:45pm U.S. Eastern Time, except U.S. holidays) Fax: 1-610-290-8972<br>Regular mail: TOEFL Services ETS PO Box 6151 Princeton, NJ 08541-6151 USA<br>Website: http://www.ets.org/toefl |
|---|---|
| Publication Date | 2005 |
| Target Population | The TOEFL test measures the ability of non-native English speakers who intend to communicate where English the language of communication. |
| Cost | varies by country; it is $195 in the US, $200 in the UK. |
| Test Purpose | The TOEFL iBT test measures test takers' abilities to use and understand English at the university level. And it evaluates how well they can combine their listening, reading, speaking and writing skills to perform academic tasks.<br>(http://www.ets.org/toefl) |
| Test Structure | The TOEFL iBT® test is administered through the Internet (internet-based). It has four sections (listening, reading, speaking, and writing). The test takes about four and a half hours to complete with check-in.<br>Reading:<br>Candidates read 3–5 passages of about 700 words each and answer 12–14 questions about each passage. There can be extra questions in this section that will not be scored. The passages have all the information to answer the questions. Background knowledge is not necessary.<br>Listening:<br>Test takers listen to 4–6 lectures, each 3–5 minutes long, and listen to 2–3 conversations, each about 3 minutes long. Listening questions are mostly multiple choice questions with a single correct answer, and some questions have more than one answer, allowing for partial-credit scores. There can also be questions in this section that will not be scored. They will be experimented for future use<br>Speaking: The Speaking section has six tasks: Two are independent. Candidates respond to a general question on a familiar topic. The other four tasks assess |

| | |
|---|---|
| | integrated skills. On two of these tasks, test takers respond to both an oral and a written stimulus; in the other two integrated tasks, they respond to an oral stimulus. The tasks follow this format: independent and integrated. Test takers will speak into a microphone.<br><u>Writing</u><br>This section has two parts:<br>Independent Writing and Integrated Writing Tasks: For the independent, candidates write about a rather general topic using their own background knowledge and/or opinions. For the integrated, candidates read a passage; then they listen to a lecture that usually casts doubt (opposes) to the reading passage.<br>(TOEFL iBT Test Framework, 2011) |
| Scoring of<br><br>the Test | |

| Skill | Score Range | level |
|---|---|---|
| Reading | 0-30 | High (22-30)<br>Intermediate (15-21)<br>Low ( 0-14) |
| Listening | 0-30 | High (22-30)<br>Intermediate (15-21)<br>Low (0-14) |
| Speaking | 0-30 score scale | Good (26-30)<br>Fair (18-25)<br>Limited (10-17)<br>Weak (0-9) |
| Writing | 0-30 score scale | Good (24-30)<br>Fair (17-23)<br>Limited (1-16) |
| Total Score | 0-120 | |

The score range for each of the four test sections is from 0 to 30, but each section is a separate measure and each measure has its own scale. Therefore, scores obtained on a section can be compared to other scores from the same section, but it is not appropriate to compare scores across different sections.

The Reading and Listening sections are scored by computer with a score range from 0 to 30. The Reading section has 36–56 tasks based on reading passages from academic texts and answering questions. The Listening section has 34–51 tasks based on listening to lectures, classroom discussions and conversations, then answering questions.

Speaking Section:

Each of six tasks is rated from 0 to 4. The sum is converted to a scaled score of 0

| | |
|---|---|
| | to 30.<br><br>Writing Section:<br><br>Two tasks are rated from 0 to 5. The sum is converted to a scaled score of 0 to 30.<br><div align="right">http://www.ets.org/toefl</div> |
| Statistical<br><br>Distribution<br><br>of Scores | Research was carried out for TOEFL test takers between January/2016 and<br><br>December/2016 showing Means and Standard deviation for the male test takers<br><br>around the world as follows:<br><br><table><tr><td></td><td>Reading</td><td>Listening</td><td>Speaking</td><td>Writing</td></tr><tr><td>Mean</td><td>20.6</td><td>20.1</td><td>19.9</td><td>20.7</td></tr><tr><td>STD</td><td>6.8</td><td>6.8</td><td>4.4</td><td>4.9</td></tr></table><br><div align="right">(TOEFL, 2017; Test and Score Data)</div> |
| Standard<br><br>Error of<br><br>Measurement | The data below presents the average section and total score reliability estimates and standard errors of measurement based on operational data from 2007.<br><br>Score               Scale        SEM<br>Reading           0-30         3.35<br>Listening         0-30         3.20<br>Speaking        0-30         1.62<br>Writing           0-30         2.76<br>Total             0-120      5.64<br><br><div align="right">(Reliability and Comparability,2011)</div> |
| Evidence of<br><br>Reliability | In the TOEFL iBT test, the reliability estimation for the Reading and Listening sections that include selected response questions is conducted using a method based on item response theory (IRT) (Lord, 1980). For the Speaking and Writing sections that has constructed response tasks, generalizability theory (G-theory) is employed (Brennan, 1983).<br>As we see in the data below, the reliability measurements of Listening, Speaking and Reading are noticeably high in comparison to the reliability of the writing section. (Breland, Bridgeman, & Fowles, 1999) argue that this measurement is typical for writing sections that have only two tasks and are time-consuming.<br>The following data shows the score reliability estimates based on operational data from 2007:<br>Score                       Reliability Estimate<br>Reading                  0.85<br>Listening              0.85<br>Speaking             0.88<br>Writing               0.74 |

| | (Reliability and Comparability,2011) |
|---|---|
| Evidence of<br><br>Validity | Lawrence (2008) considers some propositions (e.g., content relevance, scoring appropriateness and appropriate use of test results) along with proof for the propositions that show that the TOEFL iBT is valid. For example, as far as content of the test is concerned, it aligns with the types of tasks students see in academic environments. The proof given for this proposition is the reviews of hands-on studies and research of language use at colleges and universities. |

## Discussion

The target group I envision myself working with will be students at an English-learning school back in Iraqi Kurdistan. At this school, both TOEFL iBT and IELTS Academic are offered to students with the focus of familiarizing them with the parts of the tests and the strategies they need to know in these high-stake tests. Most of these students need the results of one of these proficiency tests for admission in a university or college in higher education. In terms of students' level of language proficiency, one of the requirements for students to be able to register in the preparation class is to be at least at the intermediate level, which can be determined through the school's placement test. After a review of the two tests, I will recommend the more suitable test for these students.

These proficiency tests both assess non-native English speakers' abilities who intend to study or work in English-speaking countries or where English is the language of communication. Both tests assess all four skills of listening, reading, speaking and writing. In terms of reliability, both tests are reliable as shown in the statistical data in the tables above. The range for reliability is between 0.00 to 1.00 and the closer the number is to 1.00, the more reliable it is. Both tests have pretty high scores close to 1.00.

As far as practicality is concerned, the IELTS Academic seems to be more practical for my target group because it requires less time to finish the whole test. The difference in time between the two tests is over an hour which may disinterest both test takers and test

administrators. My target group is not used to taking four-hour tests like the TOEFL iBT, so they might be at a disadvantage to take the TOEFL. It can be challenging for students to be able to stay focused for over four hours with only 10 minutes break after the first two sections. Moreover, the TOEFL iBT sometimes has extra readings in the Reading section that will not be graded which can be discouraging or tiring to students.

When it comes to the speaking sections of both tests, the Speaking section in IELTS seems to be more authentic for my target group as the test taker will speak to an actual person that allows for having the interviewer repeat a question, whereas the test taker will speak into a microphone in the Speaking section of the TOEFL iBT which seems to be less authentic and more problematic if my target group struggle to understand the question or issue they are required to talk about. This issue can be a challenge for my target students as their listening skills might not be at the level of hearing something from a machine for the first time and fully understand the question immediately. Relating authenticity to validity, one aspect of validity is when test takers are encountered with tasks that resemble real-world tasks, so as far as validity, the IELTS will be a more practical test in the speaking section for my target group.

In terms of the Speaking and writing sections of both tests, IELTS may be more appropriate and suitable for my target group because there is no integration of different skills in one task, whereas both sections in the TOEFL iBT have integrated tasks which can be really challenging for my students to be able to master that. In addition, if test takers do not understand the listening portion of the task to integrate it with another skill later on, they will not have much to produce as what they write/speak will partially be based on what they listen to.

In conclusion, based on what is discussed above in terms of authenticity, practicality and validity, the IELTS test will be more appropriate for my target students and they may obtain

more success with IELTS than the TOEFL iBT.  In terms of language difficulty of sections like the Speaking and Writing sections, the tasks in IELTS are easier for my target group as they do not have the integrated tasks. However, my recommendation may change if my target group changes in different settings.

# References

Breland, H., Bridgeman, B. & Fowles, M. E. (1999). Writing assessment in admission to higher education: Review and framework (ETS Research Rep. No. 99-03). Princeton, NJ: ETS. Retrieved on (April 12, 2017) from www.ets.org/Media/Research/pdf/RR-99-03-Breland.pdf

Brennan, R. L. (1983). Elements of generalizability theory. Iowa City, IA: American College Testing Program.

Criper, C. & Davies, A. (1988) IELTS Validation Project Report: Research Report 1(i). The British    Council/University of Cambridge Local Examinations Syndicate

IELTS handbook, (2017) Information for Candidates Introducing IELTS to test takers. Retrieved on (April 11, 2017) from https://www.ielts.org/-    /media/publications/information-for-candidates/ielts-information-for-candidates-english- us.ashx

IELTS (2008). Predictive validity of the IELTS Listening Test as an indicator of student coping ability in  Spain. Retrieved on (April 14, 2017) from https://www.ielts.org/~/media/research-reports/ielts_rr_volume12_report5.ashx

Kerstjens, M. & Nety, 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. IELTS Australia.

Lawrence, I. (2008). Validity evidence supporting the interpretation and use of TOEFL IBT Scores.  TOEFL IBT Research Insight, 4, 1-16.Retrieved from: http://www.ets.org/s/toefl/pdf/toefl_ibt_insight_s1v4.pdf

Lord, F. M. (1980). Applications of item response theory to practical testing problems. Hillsdale, NJ: Lawrence Erlbaum Associates.

Reliability and comparability of TOEFL iBT scores. (2011). Insight: TOEFL iBT Research, 1(3), 1-8. Retrieved on (April 14, 2017) from https://www.ets.org/s/toefl/pdf/toefl_ibt_research_s1v3.pdf

Shaw, SD (2004) IELTS writing: revising assessment criteria and scales (Phase 3). Research Notes 16, PP.3–7

TOEFL iBT test framework and test development. (2011). *Insight: TEOFL iBT Research, 1*(1), 1-10. Retrieved on (April 10, 2017) from https://www.ets.org/s/toefl/pdf/toefl_ibt_research_insight.pdf