

A Comparative-Correlative Study of Test Rubrics Used as Benchmarks in Assessing IELTS and TOEFL Speaking Skills

Esmaeil Bagheridoust¹ & Yasameen Khalid Khairullah²

¹ Department of TEFL, College of Foreign Languages and Literature, Islamic Azad University, South Tehran Branch, Iran

² Department of TEFL, Islamic Azad University, South Tehran Branch, Iran

Correspondence: Yasameen Khalid Khairullah, Department of TEFL, Islamic Azad University, Iran. E-mail: khalidyasameen@gmail.com

Received: April 10, 2023

Accepted: May 8, 2023

Online Published: May 10, 2023

doi: 10.5539/elt.v16n6p1

URL: <https://doi.org/10.5539/elt.v16n6p1>

Abstract

The study focused on the comparative correlative study Test Rubrics Used as Benchmarks in Assessing International English Language Testing System (IELTS) and Test of English as a Foreign Language (TOEFL) Speaking Skills. To carry out this project, the researcher searched for IELTS and TOEFL candidates and recruited and finally selected 37 male and female candidates who took both standard tests for various reasons. The statistical results obtained (mainly Pearson's correlation coefficient of correlation moments) showed significant joint variability between IELTS Speaking scores derived from IELTS test bars and frequency band description factors, TOEFL Speaking scores derived from TOEFL test bars, and assessment scores. criteria. The results show that there is a high correlation index (0.862**) between the two tests scoring systems for speaking. The results also reflect a high correlation index (.903**) between the two tests scoring systems, assessing general English proficiency and speaking ability using both tests rubrics.

Keywords: test rubrics, benchmark and scale, speaking ability, IELTS, TOEFL, intermediate EFL learners

1. Introduction

1.1 The Problem

The term "rubrics" has several meanings and elicits different responses from teachers (Popham, 1997). A secret grading sheet in the teacher's hand that is not revealed until the students' work is marked can be a "rubric." "Rubric" is a wording of the standards expected of essays by a particular faculty or department. Rubrics can contain detailed classification logic with numbers and even formulas; Alternatively, they may not have numbers and may indicate a wide range of quality (Sadler, 2009a). One heading may use generic quality words (such as "good" or "substandard"), while another heading may explain exactly what quality looks like. Some headings eschew words in favor of graphics, from emoticons to samples of what a piece should look like with certain criteria and certain standards. To be accepted into the institution, foreign students must have a minimum level of English language skills. Several English language tests are used in various parts of the country; but the Test of English as a Foreign Language (TOEFL) and the International Language Testing System (IELTS). These two tests accepted by the institutions participating in this study. The TOEFL test was developed in 1963 as an English language proficiency assessment tool. The TOEFL was created by the Educational Testing Service (ETS), a non-profit organization focused on developing a language assessment tool that is used and recognized worldwide. More than 8,500 schools and colleges accept TOEFL scores in more than 130 countries (<http://www.ets.org/toefl>). Australian Universities and Colleges Program. IELTS is administered in more than 135 countries and the results of both tests are accepted by more than 7,000 educational institutions (www.ielts.org). Knowing the structure of the test, the results and the availability of test centers in each country is also useful information when hiring internationally. Different online TOEFL scores (TOEFL iBT) and IELTS # do not use the same numbering system. On the next page, Table 1 summarizes and compares some TOEFL iBT and IELTS composite scores.

Table 1. TOEFL iBT and IELTS Total Scores Comparison (ETS TOEFL)

TOEFL iBT	IELTS
35	5.0
46	5.5
60	6
79	6.5
94	7
102	7.5
110	8
115	8.5

The nature, task and type of internationally valid tests intended primarily to measure language ability have always been unknown and unfamiliar to language teachers, language assessors and language students. In most cases, these people have little understanding and knowledge on which such test rubrics and descriptions are based. Among the pillars on which these standardized tests are designed and built are the test rubrics, which are the primary reference point and measure to be studied, researched and analyzed. In this study, the researcher mainly focuses on describing, explaining and defining the criteria according to which evaluators and researchers evaluate, evaluate and score the grammar sections of both standard tests. The results of such assessments and scoring are displayed and published in examination reports for both tests published by ETS and the Cambridge ESOL Department and IDP Australia. The importance of research assessment and evaluation of foreign language skills have been seen in the last century as a very practical way to globally screen the most talented foreign language students and identify them as suitable candidates to study in prestigious universities in North America and other English-speaking countries, and most important immigration criteria for immigrants to English speaking countries. This research is expected to have theoretical and practical implications for both language assessment and language teaching. It is important to gain an understanding of the nature of the test rubrics and the corresponding benchmark when assessing and evaluating students' English language proficiency in comparison to native English speakers.

1.2 Review of Related Literature

Indubitably, the creation of comprehensible speech has always been considered the most difficult task of language production. Speaking is a skill that most native speakers acquire without deliberate effort, but it has always been a challenge for foreign language learners. For students of foreign languages, EFL learners who plan to carry on their studies in academic settings, speaking is usually very difficult and challenging. Different approaches to teaching and assessing speaking have been adopted over the years (Raimes, 1991). Often speaking, it was assumed that acquiring the spoken language is enough to learn the written language and that it should be a priority. Consequently, teachers mostly refrained from introducing speaking in the early stages of learning a language since EFL instructors were of the opinion that due to the differences between spelling and pronunciation could hamper the accurate speaking production (Silva and Matsuda, 2002). The main emphasis of this slant was on proper precision. Teachers had to use a controlled program to develop a systematic habit so that students would avoid mistakes. Students' speaking skills were assessed mainly through interviews and direct performance tasks. Nonetheless, such methodologies were not well-matched with the new-fangled notions of DA or Discourse Analysis that developed after 1985, which highlighted the non-linear cohort of thought and its manifestation in social communication. This reaction was mainly due to the prescription and linearity inherited from the product-based approach. Regrettably, the untainted procedure of the process-approach has not grown pervasive in academic contexts, even though many EFL/ESL instructors have adapted some of its components to language instruction. In an academic context, the concern of most fields of study is that the student should be able to perform an academic speaking task that satisfies the academic community, such as essay exams. They have little to do with process orientation (Weir, 1993). In other words, the important thing here is that speaking assessment has always been considered a type of performance-based assessment, and performance-based assessment focuses on evaluating students in the process of completing given speaking tasks. Conversely, oral measures in an educational situation are far from untainted performance assessment. A crucial defy in language testing and evaluation is the overview of the perception of performance-based assessment to generate a close bond between the test condition and genuine language use (Lynch and McNamara, 1998). Gurus have documented that performance based assessment is an imperative method to acquire a self-motivated image of learners' educational and language

development (Bachman, 1990, 1991; Gipps, 1994; Genesee and Upshur, 1996; Brown and Hudson, 1998; Chapell and Brindly, 2002).

Performance-based assessment is for certain valuable to EFL learners, since it requires the strategies learners use to validate what they previously distinguish while developing their language. In EFL situations, principally learners in EFL speaking contexts are typically controlled for errors and structures they have not yet grasped. Performance assessment chiefly appraises learners on what they can accumulate and produce, not on what they recollect and reiterate. In other words, when evaluating performance, testers must perform actual tasks with meaningful tasks, rather than more abstractly demonstrating knowledge acquisition (McNamara, 1996). According to Bachman (2000), other researchers have considered this type of assessment as an alternative (Herman et al. 1992) or genuine (Newman et al., 1998; Terwilliger, 1997, 1998; Wiggins, 1989, 1993; as mentioned in the book written by Bachman in 2010) evaluation which targets the collection of proof about how learners process data in connection with a subject area (Huerta-Macias, 1995, p. 9). True outcome-based assessment differs from traditional measurement in two ways: the process of taking the test being observed and the agreed-upon rating process (e.g., rating scale), with which the performance process is evaluated (McNamara, 1996). In other words, in performance-based evaluation, the candidate's performance is rated or graded on a scale. Therefore, an important element in evaluating speech is evaluation. scale used in standardized tests such as IELTS or TOEFL. The speaking assessment score is not only the result of the test taker and the test taker, but also the interaction of the test taker, the task, the speech, the rater(s), and the assessment. Scale (McNamara, 1996). McNamara also points out that the scale used to evaluate performance tasks, such as speaking tasks, indirectly or directly represents the theoretical basis underlying the test; that is, it incorporates the test or scale developers understanding of what abilities or skills the test measures. Weigle (2002) classifies 3 forms of rating scales to measure and score a given trait. These rating scales are known as the primary trait scales, the holistic scales, and finally the analytical scales. In the evaluation of the primary characteristics, the evaluation scale is defined according to a specific speaking task, and students' speech scripts are evaluated according to how successfully the speakers performed the tasks given to them. However, in typical comprehensive scoring, each speech script is listened to and scored against a rating scale or scoring chart that determines the scoring criteria. However, analytical scoring evaluates scripts based on multiple aspects of speaking tasks, rather than assigning a single score to scripts. Depending on the determination of the appraisal, writings may be gauged on, for instance, content, organization, cohesion and coherence, unity, vocabulary, grammar, or mechanism. Systematic scoring thus delivers more precise data about a speaker's performance in various speaking situations. Therefore, many researchers prefer analytical evaluation to holistic evaluation (Bachman and Palmer, 1996; North and Schneider, 1998; Weigle, 2002). In this study, the enquirer stresses largely on reasonable research on test rubrics and corresponding descriptors, that are mainly used as yardsticks in the assessment of IELTS and TOEFL candidates particularly in education or immigration. One of the most important skills that IELTS measures and evaluates in the exam, apart from the other three skills, is speaking; namely listening, reading and writing. In speaking tests, the evaluation of speaking performance involves the construction and application of rating scales, where IELTS examiners have a tacit agreement that students' speaking ability must be at a certain frequency band level. In this way, researchers can measure the speaking ability of students, probably in the same band range. Based on speaking assessment criteria, the exam assesses four specific skills, which are fluency and ease of movement, range of vocabulary, use and accuracy of grammar and pronunciation. By looking at these four influence rating scales, for example, the IELTS examiner reflects a candidate's ability to know their specific area of service.

IELTS speaking tests are graded 1-9, just like the other parts of the IELTS tests. Candidates' performance will be assessed in four areas: fluency and coherence, vocabulary, grammar and accuracy, and pronunciation. However, there is no minimum score required to pass the exam. The assessment of speaking ability using the TOEFL speaking test requires a more careful and faster analysis of candidates' performance in six different speaking tasks recorded online by computer systems. Both scoring rubrics, the Independent Scoring Rubric and the Integrated Scoring Rubric, define key characteristics of responses along the three important dimensions of TOEFL Speaking: Delivery, Language Use, and Topic Development. When the evaluators rate the responses, they consider all three dimensions equally.

1.3 Research Questions and Hypotheses

To illustrate the research theme, the researcher posed the following research questions and the derived corresponding hypotheses as tentative solutions;

RQ₁. To what extent the rubrics used in speaking test of IELTS are similar to those of TOEFL?

RQ₂. To what extent the rubrics used in speaking test of IELTS are different from those of TOEFL?

RQ₃. Is there any reliable correlation between the scores of speaking obtained from TOEFL and IELTS tests by the same candidates?

And here are the formulated null-hypotheses:

H₀₁. There is NO reliable correlation between the SPEAKING scores obtained from TOEFL and IELTS tests by the same candidates.

H₀₂: There is NO reliable correlations among the Overall Scores of TOEFL and IELTS with TOEFL and IELTS Speaking Scores.

H₀₃. There is NO reliable correlation between the OVERALL scores obtained from TOEFL and IELTS tests by the same candidates.

H₀₄. There is NO reliable correlation between the IELTS Speaking Scores and IELTS Reading, Listening, and Writing Scores.

H₀₅. There is NO reliable correlation between the TOEFL Speaking Scores and TOEFL Reading, Listening, and Writing Scores.

H₀₆. There is NO reliable difference between the IELTS and TOEFL Speaking Scores of Male and Female Candidates.

2. Method

2.1 Participants

Thirty-Seven (37) IELTS test takers (candidates) were recruited and interviewed from amongst a large number of candidates who took the IELTS test in Iran. The candidates were selected based on the fact that if they have taken the TOEFL test as well. The main reason is that the researcher intended from the square one to compare their experience in speaking section of both TOEFL and IELTS tests while measuring the common variance between two scores obtained by the same candidates. The researcher inquired the candidates' language background as well as their plans for their future.

The participants were both male and female students from a variety of educational background. The age range was from 19 to 37, where the youngest candidate was a high school student and the eldest candidate was a PhD holder. All candidates took part in both IELTS and TOEFL test for different reasons, some of which are reflected in chapter 4 based on the interview the researcher ran with them.

2.2 Procedure and Instruments

The study was carried out through the correlational analysis of the scores obtained by the same participants (IELTS and TOEFL candidates) who experienced both tests and obtained scores in both IELTS and TOEFL tests. These participants were rare and hard to find, so the inquirer had hard time to spot such participants and it was not possible without the direct support of her thesis advisor to find and invite these candidates with their reports on IELTS and TOEFL tests.

According to David (1938) a sample size equal or superior to 25 suffices. So in this study, the researcher recruited a sample size equal or superior to 25 IELTS candidates, 37 in this study, who took the TOEFL test as well as the IELTS.

The researcher primarily ran a one-on-one interview, including 5 questions, with the participants of the study to find out how they felt about both IELTS and TOEFL tests.

Table 2. Questions of a personal interview with the candidates

R	Questions	Answer
1	Which test, IELTS or TOEFL, reflected your true ability?	
2	Which test, IELTS or TOEFL, was more user-friendly to you?	
3	In which test, IELTS or TOEFL, you believe you obtained a better overall score?	
4	With which test, IELTS or TOEFL, you had a more convenient experience?	
5	With which speaking test, IELTS or TOEFL, you had a less stressful experience?	
6	Write your overall impression and experience with IELTS and TOEFL in one or two sentences. Answer:	

The researcher then compared the total TOEFL and IELTS scores of the group and all the scores on the report card, including the four language skills: listening, speaking, reading and writing. The reason for such a comprehensive comparison was that the researcher was interested in finding out whether other language skills acquired during speaking are related or NOT.

First, the researcher compared the candidates' overall language scores on their report cards to see how well their scores on both the IELTS and TOEFL tests correlate to find the degree of variance shared between the two tests. Second, the questioner compared their performance in listening, reading and writing as secondary skills in this research project, with the aim of gaining availability and recommendations for future researchers.

Finally, the researcher compared the candidates' performance in both IELTS and TOEFL speaking sections to find a preliminary solution and answer the main question of this study: "Is there a significant relationship between the speaking performance of the tests by the same? candidates?" The only reason behind such a comparison is that the rubrics used in the IELTS and TOEFL speaking tests produced consistent results and converged in the same direction.

2.3 Speaking Rating Rubrics

Speaking in the TOEFL is recorded with a microphone and sent to the test-takers, while the IELTS speaking is a face-to-face test for all candidates. TOEFL speaking section would last for 20 minutes. IELTS would be 10-15 minutes' maximum. An important differentiator between TOEFL and IELTS is the ability to take the speaking test before other IELTS tests, while TOEFL does not have this option. English proficiency test scores are an important part of university applications in English-speaking countries. All non-native English speakers must complete these tests to apply. The most popular are IELTS and TOEFL. Both assess candidates' reading, writing, listening and speaking skills, but in different formats. I wonder why should I choose IELTS over TOEFL or vice versa? If so, here is a comprehensive blog on IELTS vs. TOEFL. The IELTS speaking test is taken in front of an examiner, the TOEFL speaking test is taken on a computer screen and recorded by the TOEFL speaking raters for later scoring.

Table 3. Comparing TOEFL iBT and IELTS Speaking scores

TOEFL iBT Speaking (0-30)	IELTS Speaking (0-9)
0-11	0-4
12	4.5
14	5*
16	5.5*
18	6*
20	6.5*
23	7*
24	7.5*
26	8*
28	8.5*
30	9

The rubric to assess the speaking ability of the participants is usually chosen from the descriptors of the Common European Framework of Reference (CEF). The CEF descriptions are versatile documents that prepare a description for independent language use and communication skills at six levels, such as A1, A2, B1, B2, C1 and C2; Levels A, B and C refer to Basic User, Independent User and Advanced User. This framework aims to improve language teaching and promotes linguistic and cultural diversity and multilingualism. The CEF is a tool for planning and evaluating language learning so that qualifications are mutually recognized and policies coordinated. (MacDonough, Shaw, and Masuhara, 2013, p. 48)

In addition, the CEF indicates a certain skill level for a given task that students should perform and how well they should perform the tasks. Also, according to the Council of Europe (2012), CEF is considered one of the standardized criteria to assess a student's language skills. Therefore, in this experiment, the researcher used the Creative Writing Criterion of CEF to evaluate the writing tasks of the participants in the classroom and to evaluate the General Writing Criterion of CEF in the pre-test and post-test.

Table 4. Comparing Features of TOEFL iBT and IELTS

Details	IELTS	TOEFL
Conducting Body	British Council and IDP Education Ltd.	Educational Testing Services
Medium	Online and Offline	Online and Offline
Variants	IELTS Academic and IELTS General Training IELTS Indicator	TOEFL iBT
Duration	2 hours 45 minutes	4 hours
Breaks	No breaks during the listening, reading and writing sections, but the speaking portion can be taken up to a week before listening and speaking sections or after the rest of the test	A 10-minute break between listening and speaking sections
Sections	Reading, Writing, Listening and Speaking*	Reading, Writing, Listening and Speaking
Test Delivery Format	Paper-based	Computer-based
Score Range	0-9	0-120
Average Minimum Score Required by Universities	Top 6.3	78
Exam Frequency	48 times in a year	More than 50 times a year
Results Timeline	Results are issued 13 days after the test	Results are issued approximately 10 days after the test
Where Accepted	All are the Scores Accepted by more than 9,000 organizations globally	Accepted by more than 10,000 organizations globally
Cost	INR 14,700/-	INR 13,780/ (TOEFL iBT)

To facilitate the calculation of CEF scores, they have been converted to a 1-6 ordinal scale, with CEF score A1 replaced by 1 and C2 score replaced by 6. Each candidate has a question. consider which to choose, IELTS or TOEFL. First of all, find out which tests are allowed by the educational institutions you are interested in. If you don't have time constraints, consider which IELTS vs TOEFL test is right for you. The TOEFL is longer but fully computer-based, so if you are concerned about your writing skills or your English when speaking to a native speaker, this could be a good choice for you. IELTS is a shorter, potentially more rigorous exam with more questions and tasks, including face-to-face interaction with the examiner.

2.4 Study Design and Analysis Plan

This comparative-correlative study was conducted by the researcher based on the speaking rubrics and corresponding descriptors used as benchmark to measure speaking ability of the IELTS and TOEFL candidates as well as the computation of the correlation between the speaking scores obtained by the candidates in both tests. Running pre-tests and post-tests in both groups, the researcher collected the scores. To analyze the pre-test and post-test data, she employed the statistical package SPSS version 21. Given the nature of the variables, design, procedure of the study and the research questions, both descriptive and inferential statistics accommodating correlational analysis and interpretation (Pearson-product moment) will be run to determine whether the difference in means between the two groups – if exist – will be significant at the .05 level. Since the design of the study is correlative-comparative, the statistical test the researcher devised was Bivariate correlation employing Pearson correlational analysis, inter-rater reliability analysis, and finally the univariate analysis.

2.5 Contribution to Knowledge

The study is meant to provide more intellectual insight into innovative pedagogies of IELTS and TOEFL scoring rubrics and descriptors through the employment of the correlational analysis and interpretations of the scores reported for one candidates in both tests. It is expected that this study will improve students' oral fluency performance if applied to all schools, colleges, and universities across the country. The approach, with the support from the Government towards organizing workshop for teachers in schools, will equip teachers of English as

Foreign Language in the country with the required methodological skills in making the test participatory programs more meaningful and convenient for test takers.

3. Results

3.1 Descriptive Analysis

In the square one of the analysis, the researcher employed descriptive statistics for the four skills being measure in both tests, along with the overall band score they obtained in the tests they took. Table 5 a shows the mean score of all the scores obtained, along with the minimum and maximum scores obtained in different skills. As the indices show, in both tests, the results are very similar in 4 skills since the scores in IELTS do not exceed band score 6 in all skills, and in TOEFL the scores do not exceed 22. The same story is true for the standard deviation indices in Table 5. Figures 1 and 2 graphically compare the scores obtained in both IELTS and TOEFL skills. As shown in the Figures, candidates had almost the same performance in dealing with the four skills of Listening, Speaking, Reading, and Writing.

Table 5. Descriptive Statistics for IELTS and TOEFL Speaking, Overall Scores, and other language skills

	N	Minimum	Maximum	Mean	
	Statistic	Statistic	Statistic	Statistic	Std. Error
IELTS Speaking Scores	37	5.00	8.50	6.5405	.15424
TOEFL Speaking Scores	37	17.00	29.00	22.6216	.61555
IELTS Overall Scores	37	4.50	8.00	6.4324	.12430
TOEFL Overall Scores	37	37.00	110.00	82.7568	2.77034
IELTS Reading Scores	37	4.00	8.00	6.3243	.14397
TOEFL Reading Scores	37	18.00	28.00	22.7568	.38634
IELTS Listening Scores	37	5.00	8.00	6.5270	.13555
TOEFL listening Scores	37	17.00	27.00	22.0000	.39895
IELTS Writing Scores	37	5.00	8.00	6.4189	.12483
TOEFL Writing Scores	37	18.00	29.00	22.6216	.48141
Valid N (list wise)	37				

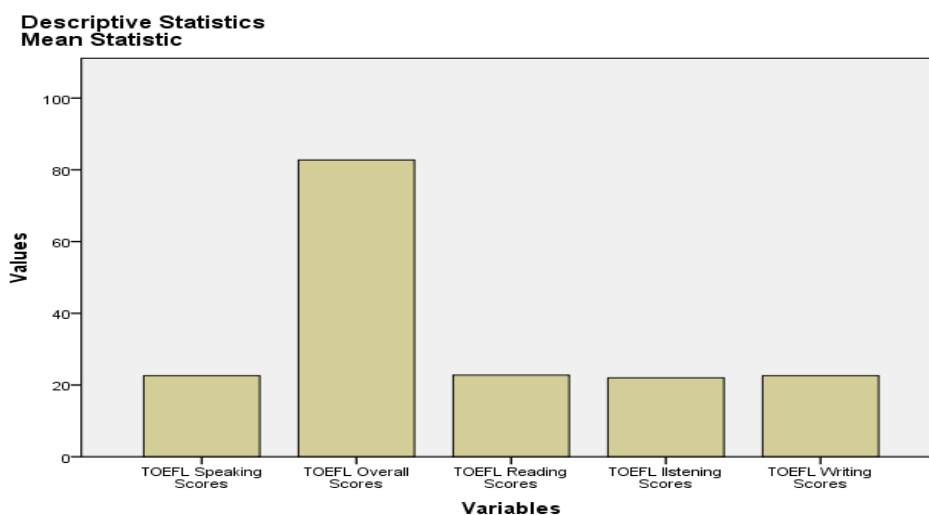


Figure 1. Comparative bar chart of TOEFL scores in each skill and the overall score

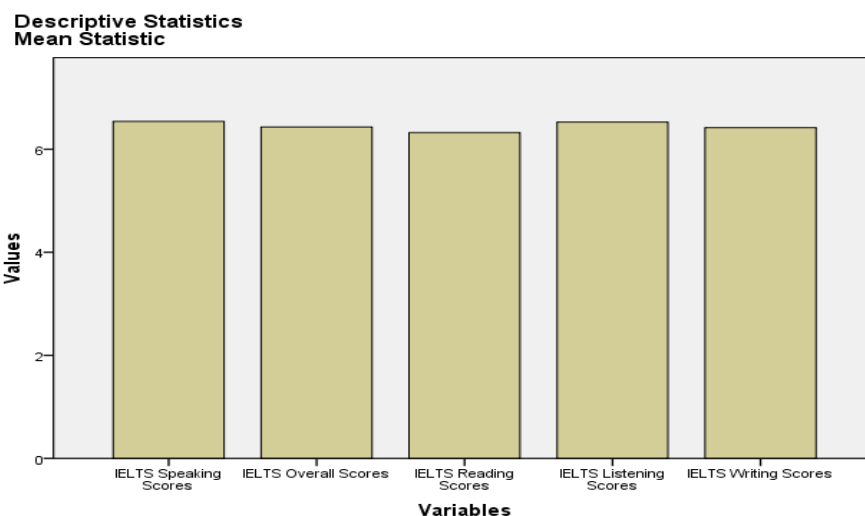


Figure 2. Comparative bar chart of IELTS scores in each skill and the overall score

As discussed above, Figure 1 and 2 have one concept in common. Candidates in both tests, namely IELTS and TOEFL, have proven almost similar performance in all skills including speaking skill. This fact indicates that both tests function as reliable as possible in dealing with the elicitation of data out of the students.

3.2 Correlational Analysis

In order to explore the degree of common variance among the variables of this study, primarily speaking skill scoring systems based on the rubrics and descriptors defined and designed by IELTS and TOEFL organizations, Cambridge ESOL Examinations and ETS, the researcher employed Pearson Correlation coefficient analyses.

In the first step of the correlational analysis, as indicated in Table 6, the inquirer ran correlation test between 4 variables, namely IELTS Speaking Scores, TOEFL Speaking Scores, IELTS Overall Band scores, and TOEFL Overall scores.

Table 6. Descriptive Statistics of IELTS and TOEFL scores

	Mean	Std. Deviation	N
IELTS Speaking Scores	6.5405	.93822	37
TOEFL Speaking Scores	22.6216	3.74426	37
IELTS Overall Scores	6.4324	.75611	37
TOEFL Overall Scores	82.7568	16.85132	37

As indicated in Table 6, there is a high index of correlation between the scoring system of IELTS and TOEFL in assessing speaking. The index of .862** is indicative of a high common variance between the two sets of scores of IELTS and TOEFL. The obtained correlation index of .862** is large enough to reject the first null-hypothesis which assumes that “There is NO reliable correlation between the SPEAKING scores obtained from TOEFL and IELTS tests by the same candidates”.

Table 7. Correlations between Speaking of TOEFL and IELTS

		IELTS Speaking Scores	TOEFL Speaking Scores
IELTS Speaking Scores	Pearson Correlation	1	.862**
	Sig. (2-tailed)		.000
	N	37	37
TOEFL Speaking Scores	Pearson Correlation	.862**	1
	Sig. (2-tailed)	.000	
	N	37	37
IELTS Overall Scores	Pearson Correlation	.905**	.854**
	Sig. (2-tailed)	.000	.000
	N	37	37
TOEFL Overall Scores	Pearson Correlation	.834**	.808**
	Sig. (2-tailed)	.000	.000
	N	37	37

** . Correlation is significant at the 0.01 level (2-tailed).

Table 7 also reflects a high index of correlation between the scoring system of IELTS and TOEFL in assessing the overall English language proficiency and assessing speaking skill using both IELTS and TOEFL rubrics. Such a result approves of the fact that standardized assessment and testing is a reliable measurement instrument to assess foreign language learners' performance in speaking. The index of .905** indicates a high correlation between the IELTS Overall Band scores and IELTS Speaking scores. The corresponding correlation index of (.854**) also indicates a high correlation between the IELTS Overall Band scores and TOEFL Speaking scores.

In the fourth row of the same Table, the index of .834** indicates a high correlation between the TOEFL Overall Band scores and IELTS Speaking scores. The corresponding correlation index of (.808**) also indicates a high correlation between the TOEFL Overall Band scores and TOEFL Speaking scores. These results blatantly bear witness to a reliable and strong rejection of the second null hypothesis of this study which indicates that "There is NO reliable correlations among the Overall Scores of TOEFL and IELTS with TOEFL and IELTS Speaking Scores." By a glance at rows 3 and 4 of Table 4, 5, which reflect a large correlation index (.903**) between IELTS Overall scores and TOEFL Overall scores, we can conveniently reject the 3rd null-hypothesis, that is to say, there is NO reliable correlation between the OVERALL scores obtained from TOEFL and IELTS tests by the same candidates.

Table 8. Correlations between speaking and the overall scores in both tests

		IELTS Scores	Overall TOEFL Scores	Overall Scores
IELTS Speaking Scores	Pearson Correlation	.905**	.834**	
	Sig. (2-tailed)	.000	.000	
	N	37	37	
TOEFL Speaking Scores	Pearson Correlation	.854**	.808**	
	Sig. (2-tailed)	.000	.000	
	N	37	37	
IELTS Overall Scores	Pearson Correlation	1	.903**	
	Sig. (2-tailed)		.000	
	N	37	37	
TOEFL Overall Scores	Pearson Correlation	.903**	1	
	Sig. (2-tailed)	.000		
	N	37	37	

** . Correlation is significant at the 0.01 level (2-tailed).

Figure 3 below shows graphically the high Pearson correlation indices between the overall IELTS and TOEFL scores with the corresponding speaking scores in both tests.

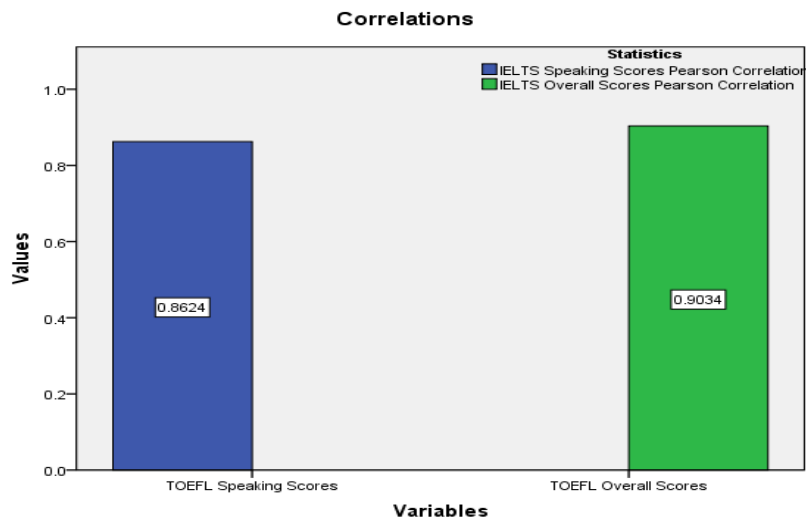


Figure 3. Comparative bar chart of the overall TOEFL and IELTS scores with Speaking

Table 9 clearly reflects the correlation indices between the IELTS Speaking Scores and IELTS Reading, Listening, and Writing Scores. The first row consecutively shows the Pearson correlation indices of .803**, .591**, and .824**, which indicate clearly a high level of correlation among the four language skills being measured by IELTS. The indices are significantly large enough to reject the fourth null hypothesis in this study which holds that there is NO reliable correlation between the IELTS Speaking Scores and IELTS Reading, Listening, and Writing Scores.

Table 9. Correlations between IELTS Speaking score and IELTS Reading, Listening, and Writing Scores

		IELTS Speaking Scores	IELTS Reading Scores	IELTS Listening Scores	IELTS Writing Scores
IELTS Speaking Scores	Pearson Correlation	1	.803**	.591**	.824**
	Sig. (2-tailed)		.000	.000	.000
	N	37	37	37	37
IELTS Reading Scores	Pearson Correlation	.803**	1	.430**	.594**
	Sig. (2-tailed)	.000		.008	.000
	N	37	37	37	37
IELTS Listening Scores	Pearson Correlation	.591**	.430**	1	.514**
	Sig. (2-tailed)	.000	.008		.001
	N	37	37	37	37
IELTS Writing Scores	Pearson Correlation	.824**	.594**	.514**	1
	Sig. (2-tailed)	.000	.000	.001	
	N	37	37	37	37

** . Correlation is significant at the 0.01 level (2-tailed).

By the same token, Table 10 clearly reflects the correlation indices between the TOEFL Speaking Scores and TOEFL Reading, Listening, and Writing Scores. The first row consecutively shows the Pearson correlation indices of .469**, .394*, and .584**, which indicate clearly a high level of correlation. The indices are reliably large enough to reject the 5th Null hypothesis in this study which holds that there is NO reliable correlation between the TOEFL Speaking Scores and TOEFL Reading, Listening, and Writing Scores.

Table 10. Correlations between TOEFL Speaking score and TOEFL Reading, Listening, and Writing Scores

		TOEFL Speaking Scores	TOEFL Reading Scores	TOEFL listening Scores	TOEFL Writing Scores
TOEFL Scores	Speaking				
	Pearson Correlation	1	.469**	.394*	.584**
	Sig. (2-tailed)		.003	.016	.000
	N	37	37	37	37
TOEFL Scores	Reading				
	Pearson Correlation	.469**	1	.516**	.612**
	Sig. (2-tailed)	.003		.001	.000
	N	37	37	37	37
TOEFL Scores	listening				
	Pearson Correlation	.394*	.516**	1	.438**
	Sig. (2-tailed)	.016	.001		.007
	N	37	37	37	37
TOEFL Scores	Writing				
	Pearson Correlation	.584**	.612**	.438**	1
	Sig. (2-tailed)	.000	.000	.007	
	N	37	37	37	37

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

The important conclusion that can be obtained out of the correlational analyses and interpretation is that the common variance between IELTS scores obtained in speaking and writing are much stronger than that of TOEFL scorings.

In order to measure the differences between the performance of male and female candidates of IELTS and TOEFL tests, the researcher ran the Case Processing Summary (CPS) test. Indices in Table 11 indicate clearly the gender performance differences.

Table 11. Male and Female Case Processing Summary

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
IELTS Speaking Scores * Male and Female Candidates	37	97.4%	1	2.6%	38	100.0%
TOEFL Speaking Scores * Male and Female Candidates	37	97.4%	1	2.6%	38	100.0%
IELTS Overall Scores * Male and Female Candidates	37	97.4%	1	2.6%	38	100.0%
TOEFL Overall Scores * Male and Female Candidates	37	97.4%	1	2.6%	38	100.0%

a. Limited to first 100 cases.

Based on the results obtained in Table 11, we can conclude that female candidates have outperformed male candidates in both IELTS and TOEFL Speaking tests as well as their OVERALL score in both IELTS and TOEFL tests. Figure 4, below graphically indicates such a difference. Therefore, we can reject the 6th null-hypothesis which holds that there is NO reliable difference between the IELTS and TOEFL Speaking Scores of male and female Candidates.

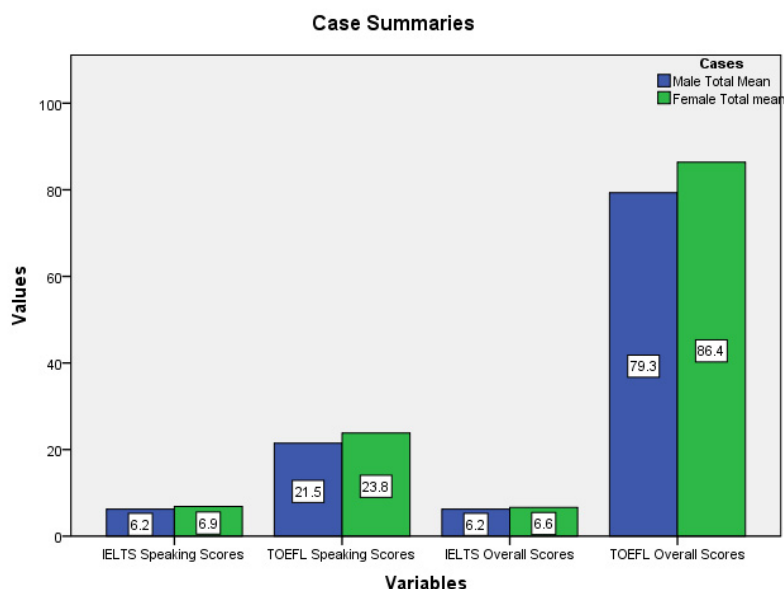


Figure 4. Comparative bar chart of the overall TOEFL and IELTS scores of Male and Female Candidates

4. Discussion

Data analysis and interpretation of both IELTS and TOEFL tests results indicate that there is a positive and high correlation between the two standardized language proficiency tests. As for the speaking rubrics of both tests being used by the examiners and raters of IELTS and TOEFL tests, we found a high correlation existing not only between the speaking tests of IELTS and TOEFL, but also between the speaking scoring and writing scoring of both tests. The results also indicate a highly positive correlation between the speaking scores of IELTS and Overall band scores obtained by the same candidates. The same story is true for the TOEFL speaking scores and the overall scores obtained by the same candidates. Last but not least is that, we can conclude that female candidates have outperformed male candidates in both IELTS and TOEFL Speaking tests as well as their OVERALL score in both IELTS and TOEFL tests.

5. Conclusion

Based on what we can read and understand from the Correlations Tables reflected in Chapter 4, and the results obtained, the researcher has come up with the following findings. There is a high index of correlation between the scoring system of IELTS and TOEFL in assessing speaking. The index of $.862^{**}$ is the indicative of a high common variance between the two sets of scores of IELTS and TOEFL. The obtained correlation index of $.862^{**}$ is large enough to reject the first null-hypothesis which assumes that “There is NO reliable correlation between the SPEAKING scores obtained from TOEFL and IELTS tests by the same candidates”.

The results also reflect a high index of correlation between the scoring system of IELTS and TOEFL in assessing the overall English language proficiency and assessing speaking skill using both IELTS and TOEFL rubrics. The results also reflect a large correlation index ($.903^{**}$) between IELTS Overall scores and TOEFL Overall scores, we can conveniently reject the 3rd null-hypothesis, that is to say, there is NO reliable correlation between the OVERALL scores obtained from TOEFL and IELTS tests by the same candidates.

The high Pearson Product correlation indices between the IELTS Speaking scores and IELTS Reading, Listening and Writing Scores are indicated in chapter 4. Based on the results obtained, the degree of correlation between IELTS Speaking scores and Writing Scores is higher than the two others ($.824$ against $.803$ and $.591$). This can be the indicative of the fact that the test rubrics and related band descriptors in measuring and assessing speaking and writing as two productive skills by the IELTS speaking and writing examiners are highly reliable, consistent, dependable, and accurate.

The results of the analysis and the correlation indices between the TOEFL Speaking scores and TOEFL Reading, Listening and Writing Scores. Based on the graph, the degree of correlation between TOEFL Speaking scores and Writing Scores is higher than the two others ($.584$ against $.469$ and $.394$).

Based on the results obtained, we can conclude that female candidates have outperformed male candidates in both IELTS and TOEFL Speaking tests as well as their OVERALL score in both IELTS and TOEFL tests.

Last but not least, candidate who took both IELTS and TOEFL tests prove to have almost similar experience in both tests and the performance of these candidates are very similar in all skills.

The results of this study has some hints for English instructors to pay more attention to the test rubrics and descriptors as reliable benchmark for assessing speaking ability of their students. This would definitely assist learners to be more successful and convenient in their speaking performance since they know the rubrics and descriptors along with the corresponding band scores used by raters and examiners. To introduce speaking test rubrics of both IELTS and TOEFL tests, language instructors can avail from the Tables and charts formulated by the ETS and Cambridge ESOL examinations Department in their speaking classes and familiarize students with such benchmarks.

In this study, TOEFL and IELTS applicants are expected to be exposed to the context in which they will be given the chance to act more dynamically and enthusiastically while dealing with speaking tasks based on the corresponding scales of both IELTS and TOEFL.

References

- Bachman, L. F. (2008). Modern language testing at the turn of the century: Assuring that what we count counts. *Language Testing*, 17(1), 1-42. <https://doi.org/10.4236/oalib.1108741>
- Brown, J. D., & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly*, 32(4), 653-675. Cambridge certificate exams, IELTS, and TOEFL. *System*, 28(4), 523-539. <https://doi.org/10.4236/jss.2018.68023>
- Chapelle, C. A., & Brindly, G. (2002). Assessment. In N. Schmitt (Ed.), *An introduction to applied linguistics*, (pp. 267-288). London, Arnold. <https://doi.org/10.4236/jss.2018.612029>
- Chapelle, C.A., Enright, M.K., & Jamieson, J.M. (Eds.) (2008). *Building a validity argument for the test of English as a foreign language*. New York: Routledge. Educational Testing Service TOEFL. Retrieved from <http://www.ets.org/toefl>
- Elder, C., & Wigglesworth, G. (2003). An investigation of the effectiveness and validity of planning time in part 2 of the IELTS speaking test. *IELTS Research Reports, 2003, Volume 6*, 1-28. <https://doi.org/10.4236/jss.2018.612029>
- Genesee, F., & Upshur, J. A. (1996). *Classroom-evaluation in second language education*. Cambridge, Cambridge University Press. <https://doi.org/10.4236/ce.2012.326134>
- Gipps, C. V. (1994). *Beyond testing: Towards a theory of educational assessment*. London: The Flamer Press. <https://doi.org/10.4236/jsea.2015.82005>
- Huerta-Macias, A. (1995). Alternative assessment: Responses to commonly asked questions. *TESOL Journal*, 5, 8-10. <https://doi.org/10.4236/jdm.2022.124015>
- Huerta-Macias, A. (2002). Alternative assessment. In J. C. Richards, & W. A. Renandya (Eds.). *Methodology in language teaching: an anthology of current practice*. Cambridge: Cambridge University Press.
- Herman, J. L., Aschbacher, P. R., & Winters, L. (1992). *A practical guide to alternative assessment*. Alexandria, VA: Association for Supervision and Curriculum Development. [https://doi.org/10.1016/S0191-491X\(01\)00032-3](https://doi.org/10.1016/S0191-491X(01)00032-3)
- IELTS Handbook. (1997). *Cambridge: University of Cambridge Local Examinations Syndicate*. 75 IELTS Partners - British Council, IELTS Australia Pty Ltd (solely owned by IDP Education Pty Ltd) and the University of Cambridge ESOL Examinations. (2009-2011).
- Lynch, B. K., & McNamara, T. F. (1998). Using G-Theory and Many Facet Rasch Measurement in The Development of Performance Assessments of ESL Speaking Skills of Immigrants. *Language Testing*, 15, (2), 158-188. <https://doi.org/10.4236/ce.2022.132044>
- McNamara, T. F. (2002). Language testing. Oxford: Oxford University Press. *Melbourne Papers in Language Testing*, 2(1), 68-85. <https://doi.org/10.4236/oalib.1108741>
- Pearson, N. (2002). *Assessment of TOEFL scores and ESL classes as criteria for admission to career & technical education and other selected Marshall University graduate programs*. Retrieved from ERIC database. (ED473756)

- Popham, W. J. (1997). What's wrong - and what's right - with rubrics. *Educational Leadership*, (2), 72. <https://doi.org/10.4236/ojl.2018.71003>
- Raimes, A. (2002). Ten steps in planning a Writing Course and Training teachers of writing. In J. C. Richards, & W. A. Renandya (Eds.). *Methodology in language teaching: an anthology of current practice*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511667190.043>
- Sadler, D.R. (2009a). Indeterminacy in the use of preset criteria for assessment and grading in higher education. *Assessment and Evaluation in Higher Education*, 34 (Online DOI: 10.1080/02602930801956059) [Taylor & Francis Online], [Web of Science ®], [Google Scholar]. <https://doi.org/10.1080/02602930801956059>
- Silva, T., & Matsuda, P. K. (2002). Writing. In N. Schmitt (Ed.). *An introduction to applied linguistics*, (pp. 251-266). London: Arnold.
- Weigle S. C. (2002). *Assessing-writing*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511732997>
- White, R., & Arndt, V. (1991). *Process writing*. London: Longman. <https://doi.org/10.4236/ce.2012.326134>
- Weir, C. (1993). *Understanding & developing language tests*. Hertfordshire, Prentice Hall. <https://doi.org/10.4236/ce.2013.412A2005>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).