EDUCATIONAL ASSESSMENT & EVALUATION | RESEARCH ARTICLE

An examination of IELTS candidates’ performances at different band scores of the speaking test: A quantitative and qualitative analysis

Laleh Dashti1 and Seyyed Ayatollah Razmjoo1*

Abstract: The purpose of this mixed-methods study was to explore Iranian IELTS candidates’ strengths and weaknesses in IELTS Speaking Test in terms of IELTS’s four speaking assessment criteria, namely Fluency and Coherence (FlC), Lexical Resource (LR), Grammar Range and Accuracy (GRA), and Pronunciation (Pro). It also aimed to examine the discourse features of the candidates’ performances in part 2 of IELTS Speaking across bands 5, 6, and 7. To this end, the oral performances of 59 IELTS candidates from a series of Mock IELTS Tests were collected, re-scored, and subjected to statistical investigation. Additionally, to better understand the

ABOUT THE AUTHORS

Laleh Dashti is a Ph.D. candidate of TEFL at Shiraz University, Shiraz, Iran. She has dedicated more than a decade of her life to teaching and is currently an EFL teacher and IELTS trainer. Her research areas of interest include teacher education, curriculum, and instruction.

Seyyed Ayatollah Razmjoo is a professor in the Department of Foreign Languages and Linguistics at Shiraz University, Shiraz, Iran. His areas of interest are Advanced Testing and Assessment, Advanced Research Methods, Materials Development, and Teaching Methodology. He published and presented papers at national and international journals and conferences. Professor Razmjoo supervised more than 100 MA and Ph.D. candidates’ theses and dissertations.

ABOUT THE AUTHORS

Public Interest Statement

What makes an IELTS candidate at band 5 to achieve band 5 in IELTS Speaking? We can ask the same question about bands 6 and 7. A band score is the function of a candidate’s linguistic strengths and weaknesses. These areas are what we have explored in this study.

We found out that a candidate’s fluency and coherence more significantly contribute to a person’s total score than vocabulary or grammar does whereas pronunciation was the weakest point for the participants, originating from an EFL context. Moreover, the most salient obstacle found in the area of Fluency and Coherence was the alarming number of incorrect cohesive devices. In the Lexical Resource domain, candidates made an excessive number of inappropriate word choices. As for Grammar Range and Accuracy, the most conspicuous problem was simply forming simple sentences. Finally, mispronunciation was the error that severely influenced many candidate’s scores.
performances, we conducted a content analysis on part 2 of the Test. The results of our regression analysis showed that FIC was the greatest predictor of success in IELTS Speaking, followed by GRA and LR, respectively while Pro was found to make the least unique contribution. Furthermore, our content analysis coupled with the application of the Monte Carlo test revealed that as the total scores moved up from 5 to 7, the rate of occurrence of the uncovered faults often declined substantially. It also showed that the association between the frequency of unearthed grammatical complications and the scores 5, 6, and 7 was strong whereas those of the remaining criteria were moderate. Moreover, the most salient obstacle found in the area of FIC, LR, GRA, and Pro, was incorrect connectors and conjunctions, inappropriate word choices, inaccurate simple sentences, and mispronunciations, respectively. The study holds clear implications for IELTS trainers, language teachers, and material developers.

Subjects: Adult Education and Lifelong Learning; Education Studies; Language & Linguistics

Keywords: IELTS; IELTS Speaking; speaking assessment criteria; IELTS Test

1. Introduction
IELTS Speaking Test comprises an interaction between a candidate and an examiner, which should take about 11–14 minutes. The test has three main parts, all of which concern the quantitative aspect of the current study and the second of which is the focus of the qualitative phase. Part 1 focuses on general questions about the candidate on various familiar topics such as family, friends, or hometown. Part 2 of the test has its function: an interaction pattern, a task required, and delivered performance. The candidate is given a task card containing a prompt and is asked to talk about a given topic for one to two minutes. Before the talk, the candidate will be given one minute to prepare (IELTS, Test format, 2019b). Part 3 regards questions related to the topic in part 2, requiring discussions on more abstract ideas.

In Part 2, while the candidate is speaking, the examiner observes the performance without causing any interruptions. A detailed performance descriptor (Appendix A) has been developed by IELTS which delineates a nine-band spoken performance assessment system based on four criteria defined and detailed out by Seedhouse et al. (2014):

*Fluency and Coherence* refers to the ability to talk with normal levels of continuity, rate and effort and to link ideas and language together to form coherent, connected speech. The key indicators of fluency are speech rate and speech continuity. For coherence, the key indicators are logical sequencing of sentences, clear marking of stages in a discussion, narration or argument, and the use of cohesive devices (e.g., connectors, pronouns and conjunctions) within and between ‘sentences’.

*Lexical Resource* refers to the range of vocabulary the candidate can use and the precision with which meanings and attitudes can be expressed. The key indicators are the variety of words used, the adequacy and appropriacy of the words used and the ability to circumlocute (get round a vocabulary gap by using other words) with or without noticeable hesitation.

*Grammatical Range and Accuracy* refers to the range and the accurate and appropriate use of the candidate’s grammatical resource. The key indicators of grammatical range are the length and complexity of the spoken sentences, the appropriate use of subordinate clauses, and variety of sentence structures, and the ability to move elements around for information focus. The key
indicators of grammatical accuracy are the number of grammatical errors in a given amount of speech and the communicative effect of error.

Pronunciation refers to the capacity to produce comprehensible speech in fulfilling the speaking test requirements. The key indicators will be the amount of strain caused to the listener, the amount of unintelligible speech and the noticeability of L1 influence. (p. 5)

The research focus is, on the one hand, on performance scores, and on the other hand, the performance features of candidates only in part 2 of the IELTS Speaking Test. The reason for this selection is the restrictive challenges of the qualitative data analysis coupled with limited time and budget. The overall aim of this study is to unearth the strengths and weaknesses of spoken performances quantitatively and qualitatively in relation to the four speaking performance criteria, explicated in the public version of IELTS speaking band descriptor (IELTS, 2019a) at three bands of 5, 6, and 7. To this end, there are three research questions:

(1) What are the strengths and weaknesses of Iranian candidates’ performances in the IELTS Speaking Test based on the IELTS’s four speaking assessment criteria of Fluency and Coherence (FIC), Lexical Resource (LR), Grammatical Range and Accuracy (GRA), and Pronunciation (Pro)?

In order to answer this question, quantitative measures will be employed to discover the relative weight of each of the criteria in determining the overall scores.

(2) What key factors in FIC, LR, GRA, and Pro Criteria differentiate bands 5, 6, and 7?

Questions 2 is answered by analyzing the spoken data inductively, employing qualitative content analysis to transcripts of part 2 of the speaking tests.

(3) Is there a meaningful association between the discovered key factors and scores assigned?

Answering question 3 requires deploying a Monte Carlo test, an extension of the Chi-square test.

Informed of the discoveries of this study, IELTS trainers will be able to enhance their teaching practices by benefiting from candidates’ most salient strengths and focusing on their most probable weaknesses. Likewise, the results of the current study redound to many IELTS trainers and self-study prone candidates in that it may raise their awareness of what areas of spoken performance require greater attention, hence causing a higher likelihood of maximizing IELTS scores and overall success in the test. The findings of this study may also help material developers have a more realistic view of Iranian candidates’ speaking proficiency.

1.1. Literature review

1.1.1. Assessing speaking

Howarth (2001) considered speaking as the process of communicating opinions, ideas, information, or emotions, hence the importance of speaking assessment. Assessing speaking refers to evaluating one’s capacity of producing oral language (Fulcher, 2003), and it is considered an indispensable component of large scale, small scale, and classroom-based assessment (Bachman, 1990).

According to Derakhshan and Nadi Khalili (2016), speaking skill comprises two main categories: accuracy and fluency. The former is considered as the correct use of language components namely, grammar, vocabulary, and pronunciation in speaking, while the latter is “the ability to keep going when speaking spontaneously” (Gower et al., 1995). As Hedge (2000) showed, fluency is a learner’ ability to speak coherently by linking words and sentences, utilizing stress and intonation, and pronouncing the sounds in a proper way. Thornbury (2005) referred to the accuracy of vocabulary as the employment of suitable words in fitting contexts.
Fulcher (2003) believed that although both speaking and writing are thought of as the productive skills, speaking is more than mere production. It involves the verbal skill as well. Furthermore, according to Fulcher (2003), the linguistic features observed in speaking are different from those observed in writing.

There existed different approaches to assessing speaking; however, the recent approaches to assess speaking might address the abilities to get messages communicated (Bachman, 1990). Speaking performance is complex and assessing this skill becomes complicated as many variables come to play. For example, test takers’ characteristics, features of the speaking test, raters, and rubric descriptors might affect a test taker’s speaking score (Seong, 2014; Qian, 2007). That said, in addition to the linguistic knowledge (e.g., pronunciation, vocabulary, stress patterns, and rhythm), the strategies and ways of using this knowledge might introduce some other variables in an effective and successful speaking test (Fulcher, 2003).

1.2. Assessment in IELTS Speaking
The approaches of speaking assessment and the notion of L2 speaking ability have evolved and broadened dramatically over the past few decades (Purpura, 2016). IELTS speaking assessment system is a classical example of direct tests, a classification proposed by Clark (1979), evaluating speaking skills and abilities in actual performance. Direct methods are defined as “procedures in which the examinee is asked to engage in face-to-face communicative exchanges with one or more human interlocutors” (Clark, 1979, p. 36). Direct tests have the advantage of their elicitation of speaking skills in a manner that duplicates “the setting and operation of the real-life situations in which proficiency is normally demonstrated” (Shohamy, 1994, p. 100); that is, direct assessments of speaking abilities presents substantial face validity.

1.3. Studies on IELTS Speaking Test
There are a plethora of studies on the IELTS Speaking Test, investigating it from various perspectives or aspects using different research designs.

Iwashita and Vasquez (2019) studied how the distinctive features of discourse competence performance correlate to the IELTS speaking band descriptor. They undertook a detailed quantitative and qualitative examination of test-takers’ oral discourse at three proficiency levels. The features of discourse competence analyzed in the study included both cohesive devices and coherence devices. The analysis revealed that some features of discourse were more distinctly observed in the higher-level test-takers’ performance than the lower level test-takers, but other features (e.g., ellipsis and substitution) were not clearly distinguished across the levels.

Roothooft and Breeze (2019) designed a study to enhance our understanding of the differences between band scores 4, 5, 6, 7, and 8 in terms of grammatical structures and morphemes as well as the error types and rates across the designated scores. The findings contributed to our understanding of the order in which certain grammatical structures are acquired in second language acquisition. The results also showed that higher score candidates attempted more complex structures despite a considerable rate of unsuccessful instances.

Elder and Wiglesworth (2006) explored planning, proficiency, and task—three aspects of the IELTS Speaking Test, seeking to find out whether the three variables interact. The concentration of the study was on planning time, aiming to differentiate the oral performances of three groups of candidates, given no time, one minute, and two minutes to plan before they attempted the task. Neither the quantitative nor the qualitative analysis reported any significant differences between the performances of the groups, suggesting the one-minute planning time incorporated in IELTS Speaking is not likely to positively assist take-takers, yet it should remain part of the test for the sake of fairness and face validity.
Read and Nation (2006), aimed to explore vocabulary use by candidates in the current version of the IELTS Speaking Test, in which Lexical Resource is one of the four criteria applied by examiners to rate candidate performance. The results of the study showed a pattern of decrease from band 8 to band 4, but there was a considerable variance within bands, suggesting that the lexical statistics did not suggest a reliable basis for differentiating oral proficiency levels. Additionally, the findings showed that the sophistication in vocabulary use of high-proficiency candidates was characterized by the fluent use of various formulaic expressions, often composed of high-frequency words, perhaps more so than any noticeable amount of low-frequency words in their speech.

Seedhouse et al. (2014) studied the features of candidate discourse in relation to the scores given to them. The quantitative measures showed that accuracy does increase in direct relation to the given scores. Grammatical range and complexity of language were the lowest for band 5; however, surprisingly band 7 holders scored higher in this regard in comparison with band 8 candidates. The measure of fluency (pause length per 100 words) showed important differences between band scores 5 and 8. In addition, the qualitative analysis did not determine any single speaking feature that made a distinction between the band scores but suggested that in any given IELTS Speaking Test, a group of assessable speaking features can be seen to lead toward a given score.

2. Method

2.1. Design
The current study employed a mixed-methods approach to address the objectives. The quantitative data was gathered by scoring the oral performances of the subjects. The scoring was based on IELTS Speaking Descriptor (Appendix A). The data was analyzed utilizing descriptive and inferential statistics. Also, a qualitative content analysis was conducted for the further and deeper elaboration of the discourse features of performances in part 2 of the Test across bands 5, 6, and 7.

2.2. Instrument
This study will deploy two different instruments. The instrument for the quantitative phase of the study was a collection of IELTS Speaking Tests, a sample of which can be found in Appendix B. These tests were randomly selected from a test bank consisting of IELTS Speaking Tests amassed from a wide range of textbooks written on preparations for the test of IELTS. Each test entailed three parts as in a real test and was administered in its full form to the potential IELTS candidates in a series of mock IELTS tests.

Furthermore, upon scoring the speaking performances, the recorded files were transcribed for further qualitative analysis. Therefore, the second instrument used in the study was qualitative content analysis, with “an emergent framework” (Ary, Jacobs & Irvine, 2019) which assumes no variables a priori and is inductive.

2.3. Participants
This research study makes use of criterion sampling. 59 participants were chosen out of a total of 72 IELTS candidates sitting three Mock IELTS tests in Shiraz University Language Center. The reason only 59 were selected was that their marks in IELTS speaking met the criteria set for this research study, namely band-scores 5, 6, and 7.

The subjects were native speakers of Farsi from both genders, in varied age groups, with diverse educational levels, and (based on the information found in their registration forms) of upper-intermediate or higher language proficiency as it is typical of IELTS candidates.

It is noteworthy that these applicants were familiar with the procedures of the IELTS Speaking Test since they had attended IELTS preparation classes, hence being properly motivated. In addition, since the participants were typical candidates of IELTS, the sample was a fair representation of the actual population. One final note is that each candidate’s identity was transformed into
code, comprised of M/F for gender, first name initial, last name initial, and a number. Table 1 shows the demographic characteristics of the participants.

As for the examiners, two IELTS experts were invited to help the researcher mark and analyze oral performances. Furthermore, to guarantee that the marks given are accurate, 13 of the performances and the scores assigned to them were randomly chosen and sent to an IELTS examiner for approval, resulting in academic confidence in the attempts and efforts made by the IELTS experts.

2.4. Data collection procedure
The data for the quantitative stage of the present study was collected in a series of three full Mock IELTS tests conducted in Shiraz University Language Center. A sum of 72 candidates from a variety of age-groups and education levels attended the Mock tests. Out of these, only 59 candidates were selected as they had obtained, in their speaking tests, marks within bands 5, 6, and 7, which were required criteria for this study. Therefore, the rest of the candidates were omitted. It is noteworthy that Shiraz University Language Center, as previously agreed upon, released only the audio files of the speaking tests, keeping the rest (reading, listening, and writing tests) undisclosed.

On the exam day, all the applicants were treated the same way as in a real IELTS Test. The candidates sat the whole test, based on IELTS standard timing. The room where the speaking test was administered was fashioned after a typical room IELTS utilizes. The intention was to recreate the same situation to take into account the stress candidates typically feel in a real IELTS situation.

After the announcements of the results, the selected participants were contacted by the Language Center to obtain their permission for the use of their registration forms and recorded voices in the present research project. They were assured that their identity would be kept confidential and that participating in this study would not in any way cause any harm to them. All of the candidates consented to the request.

In the next stage, the audio files were listened to in a nonstop fashion (no pause was applied) and re-scored by the two invited IELTS experts. The inter-rater and intra-rated reliability of the scores were calculated, showing satisfactory results for the next step to commence.

In the next step, being the qualitative phase, the audio files related to part 2 of the test were fully transcribed by one of the IELTS experts. The produced documents were meticulously controlled for possible discrepancies between the audio files and transcriptions by the second IELTS expert. What followed was holding three training sessions for the two IELTS experts, acting as coders. The goal of the training sessions was to enable the coders to code the data in a calibrated and consistent fashion. To ensure the dependability of the codes, “a code-recode strategy” (Ary, Jacobs & Irvine, 2019, p. 447) was implemented involving asking the coders to review their documents and those of the fellow coder after one week of an interval, resulting in minor revisions and satisfactory intra-coder and inter-coder reliability.

| Table 1. Mock IELTS Test candidates’ demographics |
|----------------------|----------------------|
| Demographics          | Categories | Numbers |
| Gender                | Males       | 19       |
|                       | Females     | 40       |
| Age-groups            | 21-29       | 45       |
|                       | 30-35       | 9        |
|                       | 35-39       | 5        |
| Language Levels       | B2          | 53       |
|                       | C1          | 6        |
2.5. Data analysis procedure

The statistical analyses were conducted by SPSS Statistics 24. Concerning the first research question, a descriptive and multiple regression analysis was performed to identify how much each speaking criterion contributed to the total score variance, thereby determining the weakness and strengths of the candidates.

Moreover, the transcribed speaking performances were subjected to qualitative content analysis for identifying main patterns or key factors in relation to the IELTS speaking assessment criteria. For this purpose, the common mistakes, errors, and problematic areas were coded and grouped. In the next stage, the emerging themes were subjected to frequency analysis coupled with a Monte Carlo test, an extension of the Chi-square test to examine the relationships between the unearthed features and the assessment criteria.

It must be mentioned that in the qualitative analysis of the performances according to the features delineated in the Fluency and Coherence criterion, the fluency of the candidates was not examined as this variable was already explored by Iwashita and Vasquez (2019). Furthermore, to better understand the Grammar Range and Accuracy criterion and the comprising features, the textbook, Grammar for IELTS by Hopkins and Cullen (2006) was consulted.

2.6. Inter and intra-rater reliabilities

The results of intra-reliabilities are depicted in Table 2.

As can be gleaned from Table 2, the correlations between the first and second ratings of the four IELTS speaking performance aspects conducted by the first rater were high and significant: FIC \((r = .99)\), LR \((r = .99)\), GRA \((r = .99)\), and Pro \((r = .99)\).

In addition, the results of inter-rater reliabilities are summarized in Table 3.

Evident it is from the results shown in Table 3 that the correlations between the scores reported by the first and the second raters were also high and significant: FIC \((r = .98)\), LR \((r = .98)\), GRA

| Table 2. Intra-rater correlations—First rater |
|---------------------------------------------|
| | FIC (2nd time) | LR (2nd time) | GRA (2nd time) | Pro (2nd time) |
| FIC (1st time) | Pearson Correlation | .997** | .441** | .116 | .212 |
| | Sig. (2-tailed) | .000 | .000 | .380 | .107 |
| | N | 59 | 59 | 59 | 59 |
| LR (1st time) | Pearson Correlation | .439** | .997** | .276* | .214 |
| | Sig. (2-tailed) | .001 | .000 | .034 | .104 |
| | N | 59 | 59 | 59 | 59 |
| GRA (1st time) | Pearson Correlation | .113 | .291* | .995** | .393** |
| | Sig. (2-tailed) | .394 | .025 | .000 | .002 |
| | N | 59 | 59 | 59 | 59 |
| Pro (1st time) | Pearson Correlation | .230 | .217 | .364** | .991** |
| | Sig. (2-tailed) | .080 | .098 | .005 | .000 |
| | N | 59 | 59 | 59 | 59 |

**Correlation is significant at the 0.01 level (2-tailed).
*Correlation is significant at the 0.05 level (2-tailed).
(r = .98), and Pro (r = .97). Thus, it can be concluded that the ratings were highly reliable. Moreover, based on the findings above, the rating reported by the first rater was taken as the data for forthcoming computations.

### 3. Results

#### 3.1. The first research question

The first research question intended to determine the strengths and weaknesses of Iranian candidates’ performances in the IELTS Speaking Test based on the IELTS’s four speaking assessment criteria described in the public version of IELTS speaking band descriptor (IELTS, 2019a). To this end, first, Table 4 presents the descriptive statistics of the participants’ overall IELTS speaking scores and the marks for four speaking assessment criteria (Fluency and Coherence (FIC), Lexical Resource (LR), Grammatical Range and Accuracy (GRA), and Pronunciation (Pro)), examined in this study.

Table 4 shows that the participants’ total speaking score was 5.83. It also reveals that out of the four speaking assessment criteria, Pronunciation (mean = 6.27) had the highest mean score while Lexical Resource had the lowest (mean = 5.48).
In response to the first research question, a multiple linear regression analysis was conducted, and the amount of variance accounting for the total speaking mark by each speaking assessment criterion was calculated. This operation took place after the preliminary analyses were done to ensure that no violation of the assumption of normality, linearity, and multicollinearity existed.

The normality of the data was examined by calculating the ratios of skewness and kurtosis indices over their standard errors. To meet the normality assumption, skewness and kurtosis ratios over their relevant standard errors must be within the ranges of ± 1.96. Table 5 demonstrates the pertaining results.

As shown in Table 5, the absolute values of the ratios of the skewness and kurtosis were lower than ± 1.96. Thus, it could be claimed that the present data, i.e. the overall speaking scores and the scores attributed to the four speaking assessment aspects enjoyed normality.

### Table 5. Descriptive statistics of tests of normality

|        | Skewness | Kurtosis |
|--------|----------|----------|
|        | Statistic | Std. Error | Statistic | Std. Error |
| FIC    | 0.303     | 0.311     | −0.643    | 0.613 |
| LR     | 0.417     | 0.311     | −0.585    | 0.613 |
| GRA    | 0.030     | 0.311     | −1.293    | 0.613 |
| Pro    | −0.294    | 0.311     | −0.966    | 0.613 |
| Total  | 0.403     | 0.311     | −0.575    | 0.613 |
To check the linearity of data, the Normal Probability Plot (P-P) of the Regression Standardized Residual was inspected. Figure 1 depicts the Normal P-P Plot, indicating linear data.

In addition, to check the multicollinearity, as one of the assumptions of the regression analysis, the correlation between the variables was inspected (Table 6).

As revealed in Table 6, the four independent variables correlated substantially with the total speaking score: FIC ($r = .67$), LR ($r = .70$), GRA ($r = .65$), and Pro ($r = .62$). Additionally, the correlations between the dependent variables were less than .90. If the correlation coefficients are above .90, it indicates that there may be multiple linear correlations between these variables (Pallant, 2005). Therefore, the correlation coefficients found in Table 6 between predictor variables may be regarded as evidence that there are no multiple linear correlations between variables.

Furthermore, to check the outliers in this model, the Mahalanobis distance was examined, and the results, summarized in Table 7. According to Pearson and Hartley (1956), the critical value for Mahalanobis distance in a study having four continuous variables is 18.47. Mahalanobis distances exceeding this criterion are regarded as extreme value (Pallant, 2005).

As Table 7 displays, the maximum value for Mahalanobis distance was less than the critical value (18.47) indicating that there were no substantial outliers.

After evaluating the assumptions for multiple regression analysis, the regression analysis was done to explore the power of four speaking assessment criteria to predict the participants' overall IELTS speaking scores. The results of the regression analysis are shown in Table 8.
As displayed in Table 8, the overall model significantly predicted the IELTS speaking scores (F (4, 54) = 320.46, p < .01). Table 9 depicts the results of the Model summary.

Table 9 demonstrated that the four speaking assessment criteria together explained 95% of the speaking mark (R = 96, R² = .95).

In Table 10, the results of multiple regression which indicate the power of four speaking assessment criteria in predicting the overall IELTS speaking scores are presented.
Table 10. Coefficients in multiple regression analysis

| Model | Unstandardized Coefficients | Standardized Coefficients | t | Sig. | 95.0% Confidence Interval for B | Correlations | Collinearity Statistics |
|-------|----------------------------|---------------------------|---|------|--------------------------------|--------------|------------------------|
|       | B  | Std. Error | Beta |     |     | Lower Bound | Upper Bound | Zero-order | Partial | Part |
|-------|----|------------|------|-----|-----|-------------|-------------|------------|---------|-------|
| 1 (Constant) | - .468 | .186 | -2.518 | .015 | -.840 | - .095 | |
| FIC | .308 | .023 | .408 | 13.186 | .000 | .261 | .354 | .671 | .874 | .361 | .782 | 1.279 |
| LR  | .248 | .023 | .343 | 10.834 | .000 | .202 | .294 | .701 | .828 | .296 | .748 | 1.336 |
| GRA | .264 | .020 | .398 | 13.077 | .000 | .224 | .305 | .653 | .872 | .358 | .808 | 1.237 |
| Pro | .268 | .027 | .298 | 9.817 | .000 | .213 | .322 | .622 | .801 | .269 | .812 | 1.231 |

*aDependent Variable: Total speaking score.*
According to Table 10, statistically significant t values were found for all independent variables in this model (FIC, LR, GRA, and Pro). Having tolerance values, which do not approach zero and having Variance Inflation Factor (VIF) values smaller than 10, revealed that there was not any multi-collinearity. As shown in Table 10, the tolerance and VIF values calculated for all independent variables had acceptable values.

The next step in the interpretation was to determine the B weights which show the strength of each independent variable in predicting the dependent variable. All four predictors had positive B values suggesting that a positive increase in the predictors indicated a positive increase in the overall speaking score.

Most importantly, the results of multiple regression in Table 10 showed that of the four speaking assessment criteria, FIC (beta = .40) was found to make the largest unique contribution. After FIC, by order of influence of B weights, GRA (beta = .39), LR (beta = .34), and Pro (beta = .29) were found to positively increase the IELTS speaking scores.

3.2. The second research question
The second research question comprises four components, parallel to the four criteria against which IELTS measures oral performances, as delineated in its public descriptor (IELTS, 2019a).

3.3. Fluency and coherence
The first component of the second research question concerns Fluency and Coherence (FIC) although in this study fluency was not included for reasons mentioned above. The qualitative analysis revealed two types of errors in FIC. Type 1 was designated the code, ‘FIC1’, which covers errors in using connectors and conjunctions. Type 2 was assigned the code, ‘FIC2’, which includes errors related to the use of pronouns. The following excerpts exemplify FIC1 and FIC2:

Example 1: I’m going to talk about cell phones that nowadays, we can say all the people have cell phones ... (FAE02)
Example 2: ... he was really generous, who’s not stingy ... (FAK05)
Example 3: I’m not sure that I really understood about sports events, if I want to talk about it ... (FFA044)

In example 1, the relative pronoun “that” was considered an error because the candidate did not omit its replacement, “cell phones” from the dependent clause. Example 2 shows that the dependent clause starting with “who” has no reference in the dependent clause. The third example demonstrates that the pronoun “it” does not agree with its antecedent, “sports events”.

Table 11 reports the frequencies and percentages of the two categories of errors (FIC1 and FIC2) across bands 5, 6, and 7.

As Table 11 illustrates, FIC1 (errors in connectors and conjunctions) was more frequent than FIC2 (errors in the use of pronouns) across the three selected band scores.

3.4. Lexical resource
The qualitative analysis probing the lexical aspect of the spoken performances generated four classes of errors, namely, LR1 (inappropriate word choice), LR2 (wrong collocation), LR3 (wrong word formation), LR4 (wrong word order). The following five quotes typify the errors, discovered in this area.
**Example 1:** ... why can you ignite your hands ... (MAR024)

**Example 2:** I think this journey was so sunshine. (FDN034)

**Example 3:** ... clothes he use[s] in the game (MKR014)

**Example 4:** ... they are not suppose to make noise too much. (FAE01)

**Example 5:** ... some of them like to hard work (MMK019)

Example 1 contains a word choice error (“ignite”), which for the examiner caused a comprehension problem. The second extract shows the same type of error; it is hard to understand what the candidate meant by choosing the word, “sunshine” as an adjective. Example 3 is an example of a wrong collocation. The word “clothes”, in the context of the test, should have been used with the verb, “to wear”. The fourth example is a word-formation error. The candidate should have said “supposed” (past participle form). The last extract shows that the quote contains a word order issue.

Table 12 summarizes the frequencies and percentages of these errors across the three bands: 5, 6, and 7.

Evident from Table 12 is that LR1 (inappropriate word choice), LR2 (wrong collocation), and LR3 (wrong word formation) were present in band scores 5, 6 and 7. However, in two cases, LR4 (wrong word order) occurred only within band score 6, while it was absent in 5 and 7.
The qualitative exploration of the Grammar Range and Accuracy of the IELTS speaking performances resulted in four types of errors: GRA1 (simple sentence), GRA2 (complex sentence), GRA3 (global error), and GRA4 (mistakes or tongue slips). The following excerpts from the oral performances clarify the discovered errors:

Example 1: *It’s make the connection easier.* (FAE02)

Example 2: *And I always try to figure out what is difference between different nations and different cooking ...* (FAK04)

Example 3: *I’m not the most important on the most important part is no love should be society.* (FSHM4)

Example 1, an utterance articulated about cellphones, typifies the sort of errors, designated as simple sentence error. Example 2 exemplifies the errors in complex structures. The third example is one of the global errors found in the performances. As can be seen, what the speaker meant is incomprehensible. The GRA4 errors, which are mistakes, taking place as the result of tongue slip, are not mentioned here because their occurrences were rare and specific to faster speakers at band 7.
Table 13. Frequency of key factors in Grammar Range and Accuracy

| Band score | Categories of Criteria | Total |
|------------|------------------------|-------|
|            | GRA1 | GRA2 | GRA3 | GRA4 |       |
| 5          | 144  | 1    | 26   | 0    | 171  |
| % within Band score | 84.2% | 0.6% | 15.2% | 0.0% | 100.0% |
| % within Categories of Criteria | 57.1% | 16.7% | 47.3% | 0.0% | 54.1% |
| % of Total | 45.6% | 0.3% | 8.2% | 0.0% | 54.1% |
| 6          | 82   | 2    | 19   | 0    | 103  |
| % within Band score | 79.6% | 1.9% | 18.4% | 0.0% | 100.0% |
| % within Categories of Criteria | 32.5% | 33.3% | 34.5% | 0.0% | 32.6% |
| % of Total | 25.9% | 0.6% | 6.0% | 0.0% | 32.6% |
| 7          | 26   | 3    | 10   | 3    | 42   |
| % within Band score | 61.9% | 7.1% | 23.8% | 7.1% | 100.0% |
| % within Categories of Criteria | 10.3% | 50.0% | 18.2% | 100.0% | 13.3% |
| % of Total | 8.2% | 0.9% | 3.2% | 0.9% | 13.3% |
| Total      | 252  | 6    | 55   | 3    | 316  |
| % within Band score | 79.7% | 1.9% | 17.4% | 0.9% | 100.0% |
| % within Categories of Criteria | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% |
| % of Total | 79.7% | 1.9% | 17.4% | 0.9% | 100.0% |

*Criteria = GRA.

Table 13 summarizes the results of the analysis across bands, 5, 6, and 7.

As displayed in Table 13, within band scores 5, 6, and 7, a considerable number of grammatical errors were of GRA1 kind (simple sentence). The results also revealed that band scores 5 and 6 were devoid of GRA4 (mistake), while band score 7 included three cases of it.

### 3.6. Pronunciation

The qualitative examination of the candidates’ performances in pronunciation is summarized in Table 14, depicting the types of error, discovered across the three bands, namely Pro1 (wrong word stress), Pro2 (wrong intonation), and Pro3 (mispronunciation). To illustrate, the following utterances were chosen as examples:

Example 1: ... watch it in the TV in our high school (MHN032)

Example 2: ... so close ... (MRS023)

In example 1, the word, “TV”, was pronounced Tv; it should have been pronounced tV. The second example is a mispronunciation of the word “close” which in this context should have been pronounced as/kloʊ/ rather than /kloz/. Examples of Pro2 are difficult to be presented here because they typically concern the flat intonation coming from Farsi, which can affect understanding the meaning or intention of the speaker.

Following Table 14, all key factors of Pronunciation, Pro1 (wrong word stress), Pro2 (wrong intonation), and Pro3 (inaccurate pronunciation) were present in band scores 5, 6, and 7.
3.7. The third research question

In order to find out if there was any significant difference between the frequencies of key factors discovered in the qualitative examination of the spoken performances, the Monte Carlo test was deployed. The reason for this was that the assumption of the expected value was violated, and the problem was too large for an exact option (more than 20% of the expected counts were fewer than 5), thereby, using the Monte Carlo test, an extension of Chi-square test, “provides very accurate estimates of exact p values” (Mehta & Patel, 2012, p. 30). Also, the acceptable significance level was deemed to be \( p < .05 \).

3.8. Fluency and coherence criterion

The frequency of the two types of errors unearthed from the transcriptions of the spoken data were compared to bands 5, 6, and 7 employing a Monte Carlo test. Table 15 summarizes the findings.

As presented in Table 15, the Monte Carlo test yielded significant results \( (X^2 (2, 90) = 5.84, p = .40) \), indicating that there was a significant difference between the three band scores in IELTS Speaking Test in terms of the frequency of FIC errors found in this study. It shows that FIC1 (wrong connectors and conjunctions), which was the most common, was higher in frequency in band 5 \( (F = 47) \) compared with band scores 6 \( (F = 20) \) and 7 \( (F = 10) \). The results also demonstrated that within band scores 5 \( (F = 4) \), 6 \( (F = 4) \), and 7 \( (F = 5) \), the frequency of FIC2 had almost the same frequency. In addition, the difference between the frequency of FIC1 and FIC2 was more significant in band score 5. Finally, the results of the test of the strength of association (Cramer’s \( V = .26 \)) confirmed that there was a moderate association between the three band scores and the error types in FIC.

---

Table 14. Frequency of key factors in Pronunciation

| Categories of Criteria | Total |
|------------------------|-------|
|                        | Pro1  | Pro2  | Pro3  |
| Band score 5           | 95    | 35    | 55    |
| Count                  | 43    | 13    | 19    |
| % within Band score    | 51.4% | 18.9% | 29.7% |
| % within Categories of Criteria | 61.3% | 53.8% | 67.1% |
| % of Total             | 31.5% | 11.6% | 18.2% |
| Band score 6           | 43    | 13    | 19    |
| Count                  | 17    | 17    | 8     |
| % within Band score    | 57.3% | 17.3% | 25.3% |
| % within Categories of Criteria | 27.7% | 20.0% | 23.2% |
| % of Total             | 14.2% | 4.3%  | 6.3%  |
| Band score 7           | 17    | 17    | 8     |
| Count                  | 155   | 65    | 82    |
| % within Band score    | 51.3% | 21.5% | 27.2% |
| % within Categories of Criteria | 100.0% | 100.0% | 100.0% |
| % of Total             | 51.3% | 21.5% | 27.2% |

*Criteria = Pro.
### Table 15. Monte Carlo test to compare the band scores 5, 6, and 7 in terms of the frequency of Coherence error types

| Test Type               | Value        | df | Asymp. Sig. (2-sided) | Monte Carlo Sig. (2-sided) | 99% Confidence Interval | Cramer’s V | Approx. Sig. |
|------------------------|--------------|----|-----------------------|---------------------------|-------------------------|------------|--------------|
|                        |              |    |                       |                           | Lower Bound             | Upper Bound | Value        |              |
| Pearson Chi-Square     | 6.2254      | 2  | .044                  |                           | .039 c                  | .034       | .043         | .263 .044    |
| Likelihood Ratio       | 5.567        | 2  | .062                  |                           | .076 c                  | .070       | .083         |              |
| Fisher’s Exact Test    | 5.844        | 2  |                       |                           | .040 c                  | .035       | .045         |              |
| Linear-by-Linear       | 5.96112     | 1  | .015                  |                           | .016 c                  | .013       | .020         |              |
| Association            |              |    |                       |                           |                         |            |              |              |
| N of Valid Cases       | 90           |    |                       |                           |                         |            |              |              |
|                          | Value | df | Asymp. Sig. (2-sided) | Monte Carlo Sig. (2-sided) | Cramer's V |
|--------------------------|-------|----|-----------------------|---------------------------|------------|
|                          |       |    |                       | Sig.                      |            |
|                          |       |    |                       | 99% Confidence Interval   |            |
|                          |       |    |                       | Lower Bound               | Upper Bound|
|                          |       |    |                       | Value                     | Approx. Sig.|
| **Pearson Chi-Square**   | 13.792 | 6  | .032                  | .033 ^c                   | .029       | .038     | .145     | .032     |
| **Likelihood Ratio**     | 13.891 | 6  | .031                  | .029 ^c                   | .025       | .033     |
| **Fisher’s Exact Test**  | 12.718 |     |                       | .035 ^c                   | .030       | .040     |
| **Linear-by-Linear Association** | 1.157^d | 1  | .282                  | .289 ^c                   | .277       | .300     |
| **N of Valid Cases**     | 326   |    |                       |                           |            |          |          |          |

Table 16. Monte Carlo test to compare the band scores 5, 6, and 7 in terms of the frequency of error types in Lexical Resource

Dashti & Razmjoo, Cogent Education (2020), 7: 1770936
https://doi.org/10.1080/2331186X.2020.1770936
3.9. Lexical resource criterion
The statistical comparison of the four classes of errors related to Lexical Resource found in the candidates’ spoken performances with the three bands of 5, 6, and 7 is summarized in Table 16.

The significance level presented in Table 16 demonstrated that there was a significant difference between the three band scores in terms of the frequency of the error types in Lexical Resource ($X^2 (6, 326) = 12.71, p = .03$). This implies that the distribution of the discovered errors related to Lexical Resource is not equal in band scores 5, 6, and 7. The results also showed that although LR1 (inappropriate word choice) was the dominant key factor in the three chosen band scores, it was more frequent in band 5 (F = 107) compared with bands 6 (F = 60) and 7 (F = 17). LR2 (wrong collocation) and LR3 (wrong word formation) were also more frequent in band 5. Also, Cramer’s V (.14) revealed that there was a medium degree of association between the band scores and the key factors in Lexical Resource.

3.10. Grammar range and accuracy criterion
The investigation of the transcriptions of the oral performances concerning GRA generated three kinds of errors, whose frequencies were statistically compared with regards to the three bands: 5, 6, and 7 and summarized in Table 17.

Based on the results of the Monte Carlo test, the difference among the frequencies of the key factors of GRA across the three different band scores was significant ($X^2 (6, 316) = 19.86, p = .00$). Cramer’s V test revealed that there was a high degree of association (.22) between the band scores and the key factors (Pallant, 2005). GRA1 (simple sentence), as the dominant grammatical problem, had the highest frequency of occurrence in band score 5 (F = 144) compared to band scores 6 and 7. GRA3 (Global error) also had a higher frequency in level 5 (F = 26) compared to bands 6 (F = 19) and 7 (F = 10). Additionally, the lowest possible frequency belonged to GRA2 (complex sentence), which was present within band scores 5 (F = 1), 6 (F = 2), and 7(F = 3). Being absent in bands 5 and 6, GRA4 (mistake) appeared in band score 5 just three times.

3.11. Pronunciation criterion
Across the three bands of 5, 6, and 7, the distribution of the errors found in the pronunciation of the candidates was calculated and summarized in Table 18.

As displayed in Table 18, statistically a significant difference between the band scores in terms of the frequency of the key factors in Pronunciation was found ($X^2 (2, 90) = 6.22, p = .04$), and the Cramer’s V = .13 represented a medium association. Moreover, compared to band scores 6 and 7, band score 5 enjoyed the highest frequency of mistakes in Pronunciation (Pro1 (F = 95), Pro2 (F = 35), and Pro3 (F = 55)). Additionally, band scores 5 and 6 included the three key factors of the Pronunciation with the highest frequency for Pro1 (band 5 (F = 95), band 6 (F = 43)) and the lowest frequency for Pro2 (band 5 (F = 35), band 6 (F = 13)) whereas band score 7 included higher frequency of Pro1 (F = 17) and Pro2 (F = 17) than Pro3 (F = 8).

4. Discussion
The current study sought to explore the strengths and weaknesses of Iranian IELTS candidates in the IELTS Speaking Test based on scores, given by the use of the IELTS speaking descriptor (public version) (IELTS, 2019a). It also aimed to identify the comprising features or key factors of the speaking performances in part 2 of the Test across bands 5, 6, and 7 and examine whether there existed a relationship between the features and overall marks given.

4.1. The quantitative aspect
The results of the regression analysis revealed that of the four speaking assessment criteria, FIC made the largest distinctive impact. After that, in descending order of influence, GRA, LR, and Pro were found to positively improve the IELTS speaking marks.

The prominence of the contribution of Fluency and Coherence (FIC) criterion to the overall score suggests that Iranian candidates, unlike what is commonly expectedly, are fluent and coherent, relative to their
|                | Value     | df | Asymp. Sig. (2-sided) | Monte Carlo Sig. (2-sided) | 99% Confidence Interval | Cramer's V | Approx. Sig. |
|----------------|-----------|----|-----------------------|---------------------------|--------------------------|------------|--------------|
|                |           |    |                       | Sig.                      | 99% Confidence Interval  |            |              |
|                |           |    |                       |                           | Lower Bound              | Upper Bound |              |
| Pearson Chi-Square | 30.844\(b\) | 6  | .000                  | .000 \(c\)               | .000                     | .000       | .221         | .000         |
| Likelihood Ratio    | 21.477    | 6  | .002                  | .002 \(c\)               | .001                     | .003       |              |              |
| Fisher's Exact Test | 19.866    | 6  | .002                  | .001 \(c\)               | .000                     | .002       |              |              |
| Linear-by-Linear Association | 8.703\(a\) | 1  | .003                  | .003 \(c\)               | .002                     | .005       |              |              |
| N of Valid Cases     | 316       |    |                       |                           |                          |            |              |              |

Table 17. Monte Carlo test to compare the band scores 5, 6, and 7 in terms of the frequency of error types in Grammar

\(a\) indicates a Pearson Chi-Square value.
\(b\) indicates a Likelihood Ratio value.
\(c\) indicates a Fisher's Exact Test value.
4.2. The qualitative aspect

The qualitative content analysis revealed that the comprising errors found in each band score decreased in number as candidates’ band scores increased. This finding is corroborated by Brown (2006).

In terms of FIC (more specifically, Coherence), the more the candidates’ competence improved as indicated by their band scores, the fewer errors they had in their oral performances. In fact, band 5 was where the number of FIC errors was significantly higher. This was confirmed by findings by Seedhouse et al. (2014) and Iwashita and Vasquez (2019). The regression analysis also marked Fluency and Coherence as the most important predictor of the overall mark. This suggests that the learners were influenced by a context of learning in which Fluency and Coherence were among the most salient aspect of teaching.

As for Lexical Resource, there appears to be a direct link between a candidate’s speaking competence and the overall score as confirmed by Read and Nation (2006). Band 5 was laden with many of the discovered errors, but band 7 showed far fewer errors. The most predominant issue, which was inappropriate word choice across the three bands suggests that the candidates attempted memorized words, expressions, or phrases either without contexts or in a limited number of meaningful settings, thereby a lack of ability to apply a vocabulary item to the right context while answering questions. Note worthy is that caution should be taken with linking this interpretation of the findings to the issue of negative washback caused by the IELTS Test as these observations are limited and cannot be extrapolated to a global scale.

Similar to FIC and LR, the errors in Grammar Range and Accuracy declined significantly as the band scores improved (hence the strongest association between errors and overall band scores among all criteria), implying that the candidates came from a learning context in which Grammar Range and Accuracy were the most accentuated part of the learning process in comparison with other components of language. This finding was in accordance with the results reported by Roothooft and Breeze (2019) and Seedhouse et al. (2014).

However, the unexpected finding was the existence of global errors in bands 6 and 7, which according to the public version of IELTS Speaking band descriptor, are not anticipated to surface if the candidate’s level of proficiency is above 5, yet these comprehension problems were detected through the qualitative analysis.
at bands 6 and 7. The problem might be that the examiners did not notice the errors while marking the audio files, meaning they thought they understood the candidates perfectly well. This could be because they did not use the pause bottom (as if rating a real candidate in the testing room) and listened to the audio without interruption. Another reason might be that the examiners and the candidates shared the same mother tongue, suggesting this could make it hard for the examiners to spot comprehension problems while concentrating on the stream of words being spoken at the time because both sides of the room shared one worldview shaped by one language. Another explanation can be that detecting global errors while listening to an individual who is articulating ideas, opinions, and positions using body language is probably more complicated than it is believed.

The errors detected in pronunciation, likewise, declined in number as the overall scores improved, suggesting a direct relationship between a learner’s competence and his/her overall band. The most striking feature in this area was “wrong word stress” issues as the commonest problem across all bands, which indicates a probable lack of reliance on a dictionary for pronunciation check by the learners. This can also be rooted in a learning process where accurate pronunciation is not praised or paid attention to as much Grammar Range and Accuracy is.

5. Conclusion and implications
This study probed into the strengths and weaknesses of Iranian candidates in the IELTS Speaking Test based upon four assessment criteria, delineated in the IELTS speaking descriptor (public version) (2019). Employing qualitative content analysis, it also examined the performances of candidates in part 2 of the IELTS Speaking Test across three bands of 5, 6, and 7, in the search for their distinctive features.

The results of the study demonstrated that Iranian IELTS candidates, participating in the study, greatly benefited from their Fluency and Coherence capability, which can be contributed to their language competence or the outcome of their attempts to prepare themselves for the second part of the test. Either way, IELTS trainers should capitalize on this in order to assist candidates in achieving higher scores. Conversely, Pronunciation was found to be the weakest predictor of the overall score. Therefore, EFL teachers are suggested to compensate for the fact that their language learners live in a non-English speaking country by assigning a multitude of tasks, activities, or projects, which are capable of bringing about a tremendous amount of exposure to English.

The results also revealed that there was a negative correlation between the scores given and the number of problematic areas, errors, or mistakes unearthed via content analysis. The indication is that EFL teachers and IELTS trainers should view the errors and their frequency as a sign of the candidates’ current band score (language competence) and strive to help learners become aware of the errors, and show them how they can improve their scores by resolving the problematic areas. Helping students based upon the type and frequency of an error in IELTS preparation courses, which are typically short, can potentially be effective.

The most common coherence problem was connectors and conjunctions. This invites EFL teachers and IELTS trainers to specify a course covering this aspect of coherence. Also, the remarkable number of inappropriate word choices suggest that language learners and IELTS candidates might not be aware of the fact that memorizing vocabulary even in contexts can be disadvantageous because a lexical item can fit multiple contexts, each with a distinct meaning. This can also be indicative of the fact that IELTS candidates are inflicted with a misconception about appropriate ways to improve their Lexical Resource score, which should encourage IELTS trainers to examine if they are the sources of these misconceptions being transferred to learners. Trainers are to raise candidates’ consciousness about this profound problem, meaning IELTS candidates must become informed that big words do not necessarily mean big scores. Furthermore, the frequency of errors in the Grammar Range and Accuracy of the oral performances was strongly associated with the scores given, meaning this aspect requires a greater deal of explicit attention specifically if students come from CLT contexts, where grammatical accuracy is not typically as a focal matter as fluency is. Moreover, the wrong intonation and word stress comprised a staggering number of errors. EFL teachers, as well as IELTS trainers, should take the disadvantage of learning English in a non-
English speaking country more seriously. They are, therefore, suggested to encourage candidates to expose themselves to natural English in various ways as often as possible.

The results of this study carry important implications. It can help IELTS trainers to have a realistic view of the language competence of their learners. This will probably result in better-tailored lesson plans, targeting weak areas and investing in strengths. Another implication concerns teacher training programs to broaden the prospective IELTS trainers’ knowledge concerning teaching practices and qualities of IELTS candidates’ proficiency, which should be termed, “IELTS candidate language”. The findings of this study may conduce to IELTS trainers to make guided recommendations to candidates, specifically those, opting for self-study. It also affords an opportunity for trainers to broaden their understanding of the speaking skill so they can design effective lesson plans for the candidates.

Informed of the limitations of the study, namely (a) a lack of at least two IELTS examiners to mark the audio files, (b) the use of content analysis instead of conversation analysis, and (c) a limited geographical context, future researchers are suggested to conduct the same research in a wide variety of educational contexts, including ESL and EFL settings.

Funding
The authors received no direct funding for this research.

Author details
Laleh Dashti1
E-mail: dashti_laleh@yahoo.com
Seyyed Ayatollah Razmjoo1
E-mail: razmjoo@rose.shirazu.ac.ir
1 Department of Foreign Languages and Linguistics, Shiraz University, Shiraz, Iran.

Citation information
Cite this article as: An examination of IELTS candidates’ performances at different band scores of the speaking test: A quantitative and qualitative analysis, Laleh Dashti & Seyyed Ayatollah Razmjoo, Cogent Education (2020), 7: 1770936.

References
Ary, D., Jacobs, L. C., & Irvine, C. K. (2019). Introduction to research in education. Cengage.
Bachman, L. F. (1990). Fundamental consideration in language testing. Oxford University Press.
Brown, A. (2006). An examination of the rating process in the revised IELTS Speaking Test. IELTS Research Reports, 6, 1-30.
Clark, J. L. (1979). Direct vs. semi-direct tests of speaking ability. In E. J. Briere & F. B. Hinoftos (Eds.), Concepts in language testing: Some recent studies (pp. 35-49). TESOL.
Derakhshan, A., & Nadi Khalili, A. (2016). Developing EFL learner’s speaking ability, accuracy and fluency. English Language and Literature Studies, 6(2), 177–186. https://doi.org/10.5539/elss.v6n2p177
Elder, C., & Wiggleworth, G. (2006). An investigation of the effectiveness and validity of planning time in Part 2 of the IELTS Speaking Test. In P. McGovern & S. Walsh (Eds.), IELTS Research Reports (Vol. 6, pp. 1–28). Canberra: IELTS Australia, Pty Ltd & British Council.
Fulcher, G. (2003). Testing second language speaking. Longman/Pearson Education.
Gower, R., Phillips, D., & Walters, S. (1995). Teaching practice handbook. MacMillan Education.
Hedge, T. (2000). Teaching and learning in the language classroom. Oxford University Press.
Hopkins, D., & Cullen, P. (2006). Cambridge grammar for IELTS. Cambridge University Press.
Howorth, P. (2001). Process speaking: Preparing to repeat yourself. MET, 10(1), 39–44.
IELTS. (2019a, December 22). IELTS scoring in detail. www.ielts.org: https://www.ielts.org/ielts-for-organisations/ielts-scoring-in-detail
IELTS. (2019b, December 1). Test format. ielts.org: https://www.ielts.org/about-the-test/test-format
Iwasheshi, N., & Vasquez, C. (2014, November 3). An examination of discourse competence at different proficiency levels in IELTS speaking part 2 proposals. IELTS Research Report Series, 5, 1–44. Retrieved from www.ielts.org/researchers. ielts.org: https://www.ielts.org/teaching-and-research/research-reports/online-series-2015-5
Mehta, C. R., & Patel, M. R. (2012). SPSS exact tests. IBM Corporation.
Pollant, J. (2005). SPSS survival guide: A step by step guide to data analysis using SPSS for windows. Open University Press.
Pearson, E. S., & Hartley, H. O. (1956). Biometrika tables for statisticians. Cambridge University Press.
Purpura, J. (2016). Second and foreign language assessment. The Modern Language Journal, 100(51), 190–208. https://doi.org/10.1111/mlj.12308
Qian, D. (2007). Assessing university students: Searching for an English language exit test. RELC Journal, 38(1), 18–37. https://doi.org/10.1177/0033688206076156
Read, J., & Nation, P. (2006). An investigation of the lexical dimension of the IELTS speaking test. IELTS Research Reports, 1-25.
Roothooft, H., & Breeze, R. (2019). IELTS: Investigating the development of ‘grammatical range and accuracy’ at different proficiency levels in the IELTS speaking test. IELTS Research Reports Online Series, No. 1, 2-36. https://www.ielts.org/teaching-and-research/research-reports
Seedhouse, P., Harris, A., Naeb, R., & Ustunel, E. (2014). The relationship between speaking features and band descriptors: A mixed methods study. IELTS Research Reports, 1-30. www.ielts.org/researchers
Seong, Y. (2014). Strategic competence and L2 speaking assessment. Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics, 14(1), 13–24.
Shohamy, E. (1994). The validity of direct versus semi-direct oral tests. Language Testing, 11(2), 99–123. https://doi.org/10.1177/026553229401100202
Thornbury, S. (2005). How to teach speaking. Longman.
## Appendix A

| Band | Fluency and Coherence | Lexical Resource | Grammatical range and accuracy | Pronunciation |
|------|-----------------------|------------------|--------------------------------|---------------|
| 9    | • speaks fluently with only rare repetition or self-correction; • any hesitation is content-related rather than to find words or grammar • speaks coherently with fully appropriate cohesive features • develops topics fully and appropriately | • uses vocabulary with full flexibility and precision in all topics • uses idiomatic language naturally and accurately | • uses a full range of structures naturally and appropriately • produces consistently accurate structures apart from “slips” characteristic of native speaker speech | • uses a full range of pronunciation features with precision and subtlety • sustains flexible use of features throughout • is effortless to understand |
| 8    | • speaks fluently with only occasional repetition or self-correction; hesitation is usually content-related and only rarely to search for language • develops topics coherently and appropriately | • uses a wide vocabulary resource readily and flexibly to convey precise meaning • uses less common and idiomatic vocabulary skillfully, with occasional inaccuracies • uses paraphrase effectively as required | • uses a wide range of structures flexibly • produces a majority of error-free sentences with only very occasional inappropriacies or basic/non-systematic errors | • uses a wide range of pronunciation features • sustains flexible use of features, with only occasional lapses • is easy to understand throughout; L 1 accent has minimal effect on intelligibility |
| 7    | • speaks at length without noticeable effort or loss of coherence • may demonstrate language-related hesitation at times, or some repetition and/or self-correction • uses a range of connectives and discourse markers with some flexibility | • uses vocabulary resource flexibly to discuss a variety of topics • uses some less common and idiomatic vocabulary and shows some awareness of style and collocation, with some inappropriate choices • uses paraphrase effectively | • uses a range of complex structures with some flexibility • frequently produces error-free sentences, though some grammatical mistakes persist | • shows all the positive features of Band 6 and some, but not all, of the positive features of Band 8 |
| 6    | • is willing to speak at length, though may lose coherence at times due to occasional repetition, self-correction or hesitation • uses a range of connectives and discourse markers but not always appropriately | • has a wide enough vocabulary to discuss topics at length and make meaning clear in spite of inappropriacies • generally paraphrases successfully | • uses a mix of simple and complex structures, but with limited flexibility • may make frequent mistakes with complex structures though these rarely cause comprehension problems | • uses a range of pronunciation features with mixed control • shows some effective use of features but this is not sustained • can generally be understood throughout, though mispronunciation of individual words or sounds reduces clarity at times |
| 5    | • usually maintains flow of speech but uses repetition, self-correction and/or slow speech to keep going | • manages to talk about familiar and unfamiliar topics but uses vocabulary with limited flexibility | • produces basic sentence forms with reasonable accuracy • uses a limited range of more complex structures, but these | • shows all the positive features of Band 4 and some, but not all, of the positive features of Band 6 |

(Continued)
Appendix B

**SPEAKING**

**PART 1**
The examiner asks the candidate about him/herself, his/her home, work or studies, and other familiar topics.

**EXAMPLE**

**Weekends**
- How do you usually spend your weekends? [Why?]
- Which is your favourite part of the weekend? [Why?]
- Do you think your weekends are long enough? [Why/Why not?]
- How important do you think it is to have free time at the weekends? [Why?]

**PART 2**
Describe someone you know who does something well.
You should say:
who this person is
how you know this person
what they do well
and explain why you think this person is so good at doing this.

You will have to talk about the topic for one to two minutes. You have one minute to think about what you are going to say. You can make some notes to help you if you wish.

PART 3
Discussion topics:
Skills and abilities
Example questions:
What skills and abilities do people most want to have today? Why?
Which skills should children learn at school?
Are there any skills which they should learn at home? What are they?
Which skills do you think will be important in the future? Why?

Salaries for skilled people
Example questions:
Which kinds of jobs have the highest salaries in your country? Why is this?
Are there any other jobs that you think should have high salaries? Why do you think that?
Some people say it would be better for society if everyone got the same salary. What do you think about that? Why?

Source: IELTS Cambridge 10 Test 1