

A Critical Review of IELTS Speaking Test

Rita Seroja Ginting¹, Ahmad Amin Dalimunte^{2✉}, Muhammad Dalimunte³, Eka Yuni Kurniati⁴, Devika Adelita⁵

¹ English Education Department, Universitas Islam Negeri Sumatera Utara, North Sumatra, Indonesia,

² English Education Department, Universitas Islam Negeri Sumatera Utara, North Sumatra, Indonesia,

³ English Education Department, Universitas Islam Negeri Sumatera Utara, North Sumatra, Indonesia,

⁴ English Education Department, Universitas Islam Negeri Sumatera Utara, North Sumatra, Indonesia,

⁵ English Education Department, Universitas Islam Negeri Sumatera Utara, North Sumatra, Indonesia

✉ email: amindalimunte@uinsu.ac.id

Received:

15 October
2023

Revised:

07 November
2023

Accepted:

15 November
2023

ABSTRACT

The present study on content analysis attempts to give detailed examinations of the IELTS Speaking Test. It provides an overview of the aim of tests, as well as their reliability and validity, along with a discussion of language testing during the previous three decades. The merits and limitations of the IELTS Speaking Test were investigated, with an emphasis placed on exam scenarios. Validity, reliability, and application served as the evaluation criteria for this study. The IELTS speaking test is accurate and reliable in general; however, there are a few contentious aspects that may have an effect on the test's reliability and validity. These aspects were investigated and described in this paper. In general, the IELTS speaking test is accurate and reliable. There, the following potential proposals were made: A double-marking technique, a video conference delivery, and an intervention are all examples. The IELTS Speaking Test is one specialized component that is administered in a distinct place from the IELTS Listening, Reading, and Writing tests on the same or a different day. By using the face-to-face method, the examiner only needs to do an evaluation of the candidate's performance once, and the entire process is recorded so that there is a record in case something goes wrong. A few weeks after the test, candidates will receive the results online and by mail. Subjective factors like candidate preferences and marker performance will have a significant impact on the entire testing process and evaluation findings.

Keywords: *Assessment; Evaluation; IELTS; Speaking.*

INTRODUCTION

Assessment is one of the key concerns in language instruction because it strongly affects students' learning Tukhtabaeva, (2020). It can either motivate students to continue learning the target language or demotivate them to the point where they decide they no longer want to study it. Ameri-Golestan & Nezakat-Alhossaini (2013) said that every year, thousands of students from various nations submit applications to continue their education in English-speaking nations like Canada, Australia, and the United Kingdom. One of their most important prerequisites is that applicants present a certificate attesting to their English language skills. The International English Language Testing System (IELTS) is a comprehensive examination that was developed to assess

the level of English language proficiency of individuals whose first language is not English but who are interested in pursuing education or careers in countries or regions where English is the primary language of communication Peltekov, (2021).

The International English Language Testing System, also known as IELTS, is a standardized test that demonstrates to prospective students that they will be able to succeed in a university or country where English is the primary language of instruction. This includes the master's and PhD levels in the majority of European countries. IELTS was established in 1989, and it is currently becoming increasingly well-known all over the world Peltekov, (2021). IELTS Australia, the British Council, and the IDP: Cambridge Assessment English are all co-owners of the organization. IELTS testing has been the subject of the vast majority of empirical studies (e.g. Liao, 2021; Peltekov, 2021; and many others). It should be noted that not many studies on IELTS skills have been done. For instance, Liao, (2021) examined the homogeneity of IELTS result across testing sites as part of their investigation into the fairness of broad international language examinations.

The IELTS exam is broken up into four different parts: speaking, listening, reading, and writing. These sections test a student's ability to varying degrees in the English language. Speaking is a participatory activity, which is one of the reasons why it is such an important skill, as stated by Hughes (2003). The speaking portion of the IELTS test is delivered in the form of an oral interview that takes place in a realistic setting. In contrast to the components that are provided by computer or on paper, the speaking portion of the exam lasts anywhere from 11 to 14 minutes and is conducted in person with a qualified examiner. After that, the results are scored using a scale ranging from 1 to 9, with both full and half bands being used. On the IELTS exam, students are required to exhibit integrated speaking skills based on the following four criteria: fluency and coherence, lexical resource, grammatical breadth, and correctness in pronunciation. A multitude of validation studies have been undertaken to investigate the various factors that impact the performance of individuals taking the IELTS speaking test (e.g., Green A. & Hawkey R., 2012; Wigglesworth & Elder, 2010). Numerous elements have been recognized in scholarly literature as exerting influence on language testing. The parameters under consideration include the test task or format (Brown A., 2006; Iwashita & Vasque, 2015), timing Wigglesworth & Elder, (2010), and topic (Khabbazzbashi, 2017). Prior studies have been undertaken to investigate the procedures employed by examiners and the evaluation methods they employ Carey & Mannell, (2009).

However, the IELTS speaking test is still controversial due to its marking systems on test-takers' language proficiency, despite the fact that it is utilized extensively and acknowledged by organizations, corporations, and governments all over the world. In other words, it is essential for both test-takers and units that utilize test results to establish a person's competence to know whether or not test results accurately reflect test-takers' language behaviors during the test and whether or not these behaviors accurately reflect test-takers' actual language competence. In addition, it is important to know whether or not these behaviors accurately reflect test-takers' actual language competence. Thus,

establishing the legitimacy of the IELTS speaking test is of the utmost importance and this research offers the novelty in the kind of a deeper analysis on IELTS speaking session.

The Goals of the Test

According to Fulcher (2006), standardized testing of languages serves a variety of purposes both in the classroom and in society as a whole. As we make thoughtful judgments about language examinations, we should take into account not only the historical evolution of testing and assessment but also the appropriate roles that testing should play in equitable societies. So, while analyzing actual applications, it is necessary to have a solid grasp of the aim of testing Chalhoub-Deville & Turner, (2000). According to Hughes, (2003), the four most important examinations are the proficiency test, the accomplishment test, the diagnostic test, and the placement test.

According to Quaid (2018), the majority of proficiency exams are designed to determine how applicants have attained a particular standard in relation to a particular set of competencies. Furthermore, according to Fulcher & Davidson, (2007), standardized proficiency assessments are typically norm-referenced rather than criterion-referenced. Fulcher, (2006) claims instead of evaluating a candidate's achievement to that of other candidates taking the same test, the rating criteria standard is used to evaluate a candidate's performance during the proficiency test. Moreover, as demonstrated by Bachman, L. F., & Palmer, (2010) proficiency examinations are accessible to individuals of all skill levels, provided that applicants are adequately prepared to take them. These examinations assess the four most fundamental linguistic abilities (listening, speaking, reading, and writing). Thus, the IELTS might be regarded as a competency test.

Achievement tests, on the other hand, are directly tied to how individuals and groups of students perform in language classes. According to Fulcher (2006), achievement assessments primarily assess students' understanding at the time of the test, irrespective of any prior language skills applicants may have acquired.

According to Li (2019), diagnostic tests are significant since they primarily identify the learning that still has to occur. Additionally, Moland, C. W., & Oetting, (2021) stated that effective diagnostic tests are crucial for self-education. Diagnostic testing may reveal gaps in language proficiency, but creating a diagnostic exam requires a lot of effort and the desire of test developers, as Shums'kyi, (2020) noted. As a result, a diagnostic test is difficult to execute.

According to Fulcher, (2006), a placement test can be considered as a comprehensive diagnostic test because it classifies pupils according to their respective strengths and weaknesses in order to provide suitable education. In other words, the objective of a placement test is to ensure that students receive the necessary teaching. According to Bachman, L. F., & Palmer, (2010), placement examinations will assist teachers in assigning students to the appropriate phase of a lesson plan based on their aptitude. This is because the test's name, "placement," conveys its function.

Reliability

The dependability of a test can be defined as the degree to which test results are consistent across multiple test administrations and test assignments Bachman, L. F., & Palmer, (2010). According to Himelfarb, (2019), dependability is the degree to which test takers' language results remain stable over several administrations of the same assessment technique. Consistently gauging an individual's competency at any given moment and in any given setting is a hallmark of reliable tests Hughes, (2003). When both the testing method and the possibility of variability between manufacturers are considered, the results of the test are thought to be more trustworthy Quaid, (2018). Individuals' linguistic abilities and test scores are highly susceptible to random, uncontrollable influences. They include things like illness, a lack of drive, and test-day jitters.

In addition, Bachman, L. F., & Palmer (2010) point out that the variations in testing settings, as well as weariness and worry, can have an effect on applicants' performance, which can result in scores that are inconsistent from one instance to the next. When examiners have a full understanding of the influence these factors have, they are able to reduce measurement error while simultaneously increasing reliability Bachman, (1990). In a similar vein, in order to improve the reliability of tests, the following factors should be taken into consideration: the conditions under which the test is administered; the psychometric properties that are embodied in the difficulty indices of test tasks; and the standard error of measurement, particularly near passing scores Chalhoub-Deville & Turner, (2000). Aside from the initial quantitative indicator of the test's usefulness, which is its reliability Quaid, (2018). Validity, an additional interconnected term, is equally deserving of interpretation in connection to language testing. In point of fact, reliability is an essential condition for validity, and the reason for this is that invalid interpretation and application are seldom ever guaranteed by unreliable test outcomes Bachman, L. F., & Palmer, (2010). In addition, reliability is related to the effects of measurement error that are associated with the least severe consequences, whereas validity is related to the effects of language ability that are associated with the most severe consequences that may be evaluated Bachman, (1990). After that, the definition of validity and the issues that are connected to it will be presented as follows:

Validity

According to Hughes (2003), a test is only deemed credible if it is able to measure the construct for which it was created to assess in an accurate manner. Validity is an evaluation of the appropriateness of conclusions and actions based on test scores. After that, there will be a series of questions about the different types of relevant scores, the degree to which these scores can assist in decision-making, and the question of whether or not those who opt to take tests experience any beneficial impacts as a result of doing so Fulcher, (2006).

The evaluation of language tests is rapidly adopting the use of additional subordinate forms of validity, such as content validity and criterion-related validity. A

further category of validity is known as "face validity." Im, G. H., Shin, D., & Cheng, (2019) said that validity is a determination of the amount to which empirical data and theoretical rationales support the appropriateness and suitability of conclusions and actions based on test scores. This can be done by looking at the empirical evidence and theoretical rationales. It is necessary to use a sample that is typical of the language abilities and structures that are intended to be evaluated in order to determine whether or not a test has a valid content. This is done so that the content validity of the test can be determined. This can be done by selecting questions from a bank of ones that have been used in tests that have already been administered Bachman, L. F., & Palmer, (2010).

Criterion-related validity, on the other hand, is demonstrated by the empirical relationship between test scores and criterion scores, as opposed to content validity, which is demonstrated by examining the relationship between the two. This is done in order to determine whether or not the two objects under investigation are related. In contrast, content validity is determined by examining the content itself, hence this is an entirely separate idea Pearson, (2019).

The Various Language Tests Periods

According to Spolsky (1997), the development of language testing can be broken down into three distinct phases: the pre-scientific, the psychometric-structuralist, and the psycholinguistic-sociolinguistic phases. The prescientific period comes before the scientific phase in human history. It is important to note that each period of testing languages was strongly tied to the corresponding developments in the fields of teaching, psychology, and sociology during that period. In addition, these three movements are occurring simultaneously in time and manner.

Therefore, determining a learner's proficiency in the pre-scientific period mostly consisted of testing their command of grammatical and lexical resources, as this was the only method available at the time. At this point, there was no consideration given to either the concept of dependability or validity. On the other hand, linguistics and psychology had tremendous growth throughout the 20th century. At first, linguists who subscribed to the structuralism school of thought and psychologists who adhered to the behaviorism school of thought believed that the process of learning a language entailed carrying out structural exercises that were repetitive in nature and were based on phrase patterns that were common among "native speakers." After a period of time, a second derogatory term that labeled language as having inborn norms eventually became the most common one. Since that time, getting correct input has been one of the most important priorities Read, (2022). Later on, at the psycholinguistic-sociolinguistic level, Canale and Swain Fulcher & Davidson, (2007) introduced the idea of communicative competence. The evaluation of linguistic competence at this level was informed by sociolinguistics, functional linguistics, and sociocultural psychology, all of which contributed individually and collectively. Students were reminded that they are social beings and that the ability to communicate effectively through language is an essential component of successful social interaction. At the moment, an emphasis is being made

on integrative testing, with concepts of communicative competence serving as the primary topic of discussion. As a direct consequence of this, the focus of the test has shifted to evaluating the candidate's ability to use the target language effectively for a variety of situations and functions in their day-to-day lives Hughes, (2003).

The IELTS speaking test, which is widely considered to be one of the most relevant measures of knowledge in the modern period, includes components that are comparable to those that may be discovered in examinations from the post-scientific era. The IELTS speaking test is designed to evaluate candidates' level of oral language ability using rating criteria as the benchmark. Production of spoken language relies on social interaction, which requires communicative language abilities to function well. In addition, candidates are given sufficient opportunities to talk for extended periods of time and significantly demonstrate their skill in the English language Taylor, (2001). The IELTS speaking test is recognized as being valid and trustworthy in accordance with the overarching principles that govern the post-scientific era.

Context

The British Council and IDP, IELTS Australia are the organizations that support and administer the examination based on the 2019 IELTS Syllabus. As one of the four sections, the speaking test aims to evaluate a wide range of skills, including the ability to naturally communicate opinions and information on everyday topics; speak at length on a given topic in an appropriate manner; organize ideas coherently; express and justify opinions; and analyze and discuss issues fluently (<http://www.ielts.org/>). Speaking is one of the most significant components of the IELTS examination. The following are important details about the IELTS speaking exam.

Candidates

Around 10,000 different organizations all over the world recognize the results of the International English Language Testing System (IELTS) (<http://www.ielts.org/>). IELTS stands for the International English Language Testing System. Individuals who wish to live in countries such as Australia, Canada, New Zealand, or the United Kingdom are required to have IELTS scores that are satisfactory in order to be eligible for these activities. These countries include those countries. Information regarding the test, including its specifics, can be found on its official website. Users who are registering for the IELTS will find that it is quite straightforward to go around and use for their needs. Once the personally identifiable information disclosure form has been submitted, test types, and test schedule, the examination fee, which is around 170 British Pounds, needs to be paid. After some time has passed, participants will receive emails and messages explaining to them when and how they will take the test, as well as how to finish the procedure in its entirety. There are even instructional and learning films that can be accessed for free on the official website to help you get ready for the IELTS. Nearly 78 percent of candidates are expected to join academic IELTS in 2017, according to demographic figures. Participants in Greece, Germany, and Canada all do very well in the speaking portion, achieving a band score of 7. The majority of test takers, both male and female, score higher than band 6 on the speaking portion of the exam.

Examiners

More than 7,000 IELTS examiners have their grading standards monitored and kept up to date by the IELTS Professional Support Network, which was established and is managed by the British Council in collaboration with IDP: IELTS Australia. Despite how skilled these examiners are, they all need to undergo frequent retraining and reassessment testing. Examiners are sent to the speaking session in a randomized and private manner. At the start of the speaking examination, one examiner will be waiting in the testing location for the candidate who was sent there. Crucially, the examiner only scores the candidate's competence once, and the entire procedure is preserved for safety. Examiners need to adhere to the guidelines and execute a script in order to maximize test reliability and validity.

Test Conditions

The International English Language Testing System (IELTS) might administer the listening, reading, and writing examinations on the same day as the spoken test, or they might be administered on different days. Each person who is taking the exam and the person who is grading them will have exclusive use of a room that is large, quiet, and uncluttered. In the event that applicants run into any issues, they are free to seek assistance from the invigilators at any time.

Test Structures

Speaking sections are included in both the Academic IELTS and the General IELTS; nonetheless, the substance and organization of these speaking sections are exactly the same. The speaking portion of the IELTS exam is broken down into three key components that are equally important. Each component is responsible for carrying out a different role in relation to the task intake, the candidate output, and the interaction pattern.

The first phase of the examination begins with the examiner establishing the subject's identity and shaking hands with them (the introduction). The examiner will next move on to asking open-ended questions about common topics such as homes and families, employment and education, interests and daily routines, friends, and other people in your life. It shouldn't take more than four or five minutes to finish this section.

The candidate will be handed a cue card to discuss a certain subject during the individual long turn that occurs during the second half of the game. Before they are given the opportunity to speak for a total of up to two minutes, the candidates are given one minute to prepare their remarks. In addition, the examiner will give a whiteboard for you to use, and you are allowed to jot down notes on it. As soon as the allotted amount of time has passed, and only then, will the examinees be interrupted while they are speaking. This particular section lasts somewhere in the range of three to four minutes.

In the third and final part of the test, which is a two-way discussion, the candidate will be questioned about two, three, or more related topics, depending on how the first part of the test was interpreted. At this phase of the examination, the examiner and the candidate may engage in conversation about more conceptual topics and ideas. This section lasts approximately four to five minutes. Furthermore, as stated by Taylor (2007),

candidates are obligated to utilize a diverse range of language functions in section 3, which includes providing explanations, descriptions, and making conjectures.

Rating Scales

The IELTS speaking test's scoring scale, similarly like the score range for the other parts of the IELTS tests, ranges from 1 to 9. The applicants' performances will be graded according to their pronunciation, grammatical accuracy, lexical resource, fluency, and cohesion Seedhouse & Nakatsuhara, (2018). These are the four categories that will be taken into consideration. However, there is not a required minimum score in order to pass the test. The information in its entirety can be found in the appendix (more specifically, the band descriptions in their public version).

The Development of the IELTS Speaking Tests

The IELTS speaking examinations were first included as a section of the overall IELTS examination in the year 1989. The British Council, Cambridge English Language Assessment, and the International Development Program of Australian institutions and colleges are in charge of their administration respectively. After that, in the year 1995, the IELTS went through a process that resulted in major changes. In order to ease some of the stress that is associated with organizing the speaking examination, the candidates were given the option of taking the speaking test on a separate day than the other three modules. Both the content of the test and the way it was evaluated underwent extensive revisions in the year 2001. To be more explicit, there were activities included in the speaking test, and the speaking paper included the examiner instructions and scoring guidelines. Taylor, (2001).

METHOD

The International English Language Testing System (IELTS), which is generally acknowledged as being one of the modern competence exams, has sections that are comparable to those found in examinations from the post-scientific era. It is common knowledge that one of the most important parts of the IELTS exam is the speaking portion, which measures a candidate's ability to communicate in English. This evaluation is meant to determine the level of spoken language abilities that each candidate possesses, and the grading criteria serve as the benchmark for this evaluation. The International English Language Testing System (IELTS) is a requirement for people who want to live, work, or study in countries such as Australia, Canada, New Zealand, or the United Kingdom. In order to meet the requirements of these nations, individuals must have IELTS scores that are acceptable. This information can be obtained by going to www.ielts.org on the internet. The range of possible scores on the IELTS speaking examinations is between 1 and 9, the same as the range of possible scores on the other components of the IELTS exam. The candidates' performances will be judged according to a wide range of criteria, including their fluency and coherence, lexical resources, grammatical accuracy, and pronunciation. These are the four distinct categories that will be taken into consideration in this discussion. Nonetheless, in order to pass the

examination, it is not necessary for the candidate to get some predetermined minimum score. So, it is absolutely necessary to do research on the reliability and plausibility of the speaking test.

In addition, the method used was qualitative study in the form of content analysis through doing an in-depth examination of the IELTS Speaking Test with the help of these transcripts is facilitated by their utilization. (Khatimah et al., 2022) According to Sandelowski qualitative descriptive study is the process of retrieving information from stakeholders about their own experiences with the institutions in order to reconstruct the actual governmental design of public policies or organizational management system. The purpose of this study was to review the reliability and validity of tests critically, in addition to the past three decades of language testing, which will be discussed in further depth in the following section. The conditions under which the IELTS Speaking test is administered are emphasized throughout this analysis of the benefits and drawbacks of the test, which analyzes the validity, reliability, and usefulness of the test from a variety of points of view. Dependability can be measured by the degree to which one's results remain stable from one series of tests and test tasks to the next. The quality of being dependable is one that is not only essential but also of the utmost relevance. The dependability of test scores is directly proportional to the regularity of their distribution Bachman, L. F., & Palmer, (2010). In addition, Chalhoub-Deville & Turner (2000) define dependability as the degree to which test scores for a set of test-takers are constant across multiple applications of a measurement approach. In other words, dependability is the degree to which a measurement technique can be relied upon. While Validity is a determination of the amount to which empirical facts and theoretical rationale support the appropriateness and sufficiency of conclusions and actions based on test scores. This can be done by comparing the empirical data with the theoretical rationale. The empirical data and theoretical basis for the inferences and deeds are both referred to as this phrase. There will then be a series of questions regarding the types of useful scores, the ways in which these scores can aid in decision-making, and the question of whether or not tests have positive outcomes for those who choose to take them. The result of this is an expansion of our understanding of the factors that constitute validity Fulcher, (2006). A demonstration that is primarily focused on the affective components should not include the testing abilities in order to guarantee that the test scores will be meaningful Bachman, (1990).

In this particular study, the researchers accessed the official IELTS website in order to collect data concerning the structure, content, and grading criteria of the IELTS Speaking Test, as well as recommendations for test takers who are preparing for the exam. The researcher was able to acquire a solid command of general information pertaining to the IELTS Speaking Test as a result of this. The researcher also went online and read a significant number of different series of IELTS Research Reports. This is in addition to the previous step. In order to answer the primary research question, which was "Is the IELTS Speaking Test reliable and valid? ", the researchers reorganized their points of view and reorganized their conclusions based on the solid foundation that was provided by the body of literature. This was accomplished by taking notes and

performing an in-depth analysis of the findings obtained from various types of research conducted in other studies.

FINDINGS AND DISCUSSION

Validity

Content Validity

The content validity of a test refers to the degree to which the content of the test is relevant to a particular behavioral area of interest and the degree to which the content of the items or tasks are typical of the entire test, as assessed by an examination of the relevant literature Hughes, (2003). It is essential to determine the domain specification because doing so permits research into the connections that exist between test performance and performance in a variety of different contexts Hughes, (2003). Content validity on its alone is insufficient to support a claim of validity because there is a lack of data showing how well people do on examinations Bachman, (1990). The speaking portion of the IELTS exam is comprised of a number of different exercises, all of which have been validated in terms of the subject matter that they cover and are designed to correspond with the communicative standards that are included in the test. Speech functions such as comparing, summarizing, explaining, recommending, contrasting, narrating, paraphrasing, and analyzing are common in a candidate's output. This is true even in the absence of an external exam structure. Compare and contrast, narrate, describe, explain, recommend, contrast, and compare and contrast are some examples of what you can do Karim & Haq, (2014). Nonetheless, there is always the possibility that a test will not be completely valid. Your ability to ask questions is not evaluated as part of the speaking portion of the IELTS test Seedhouse & Morales, (2017).

A succession of questions and answers that are adjacency pair based on a particular topic barely gives the applicant with a topic change or introduction. To put it another way, applicants have very limited opportunities to demonstrate their capacity to oversee the development of topics and take turns speaking Seedhouse & Morales, (2017). The applicants at all levels have the option of taking the speaking test either for general or academic purposes, depending on their preferences. Nevertheless, regardless of which option they choose, both the content and the structure of the speaking test will remain the same. It is true that younger candidates who take the IELTS exam for the purpose of higher education come from a very different background than older candidates who may be in this country for immigration reasons. Therefore, improper subjects such as a task connected to business might not be acceptable for adolescents Seedhouse & Morales, (2017).

Predictive validity

The predictive validity of an analysis is directly proportional to how accurate the forecast is (Fulcher, 2006). The anticipated actions of a future-oriented criterion are factored into the calculation of a score for that criterion. The relationship between the test's apparent measurement of a skill and the projected performance is something that should be taken into consideration when analyzing a test's predictive power Karim &

Haq, (2014). However, according to Quaid (2018), the development of language in the IELTS tests at the micro level might not always be an indicator of overall language proficiency (macro level). It is much simpler to evaluate general speaking abilities in a general setting than it is in an academic one due to the fact that the oral examinations used in general and academic settings are same in both format and content (Karim & Haq, 2014). In addition, throughout the entirety of the IELTS evaluation process, examiners are looking for L2 classroom discourse that is interactive and as near as possible to serving a goal that is relevant to the actual world (Seedhouse and Morales, (2011). There are times when this mode of institutional discourse is unable to correctly predict the future professional achievement of individuals. The predictive validity is unaffected by the hypothetical inferences that can be drawn from the results of the speaking exam (Seedhouse & Morales, (2017).

Face Credibility

Face validity is something that is defined by Hughes (2003) as the credibility or public acceptance of a test when it is first looked at. To put it another way, if a test has face validity, it gives the impression that it measures what it is designed to assess. One way of looking at it is to say that the IELTS speaking test satisfies the requirements for face validity. According to the findings of the research conducted, the two most essential elements in evaluating whether or not a test can be considered to have face validity are the familiarity of the test format and the authenticity of the test task (Karim & Haq, (2014). As a consequence of this, the inquiry of the apparent validity of the IELTS speaking test will follow this phase, as will be detailed in further detail later on. The IELTS speaking test features an easy-to-follow structure and a reliable basis for its questions (Taylor, (2001). The context reveals that there are three parts, which are the introduction, the extensive individual turn, and the two-way debate. Participants in the examination have the ability to comprehend the testing flow, testing methods, testing materials, and testing structures in distinct segments.

Therefore, the IELTS speaking test "appears" to be dependable overall. Second, applicants, educators, and even parents are able to quickly acquire examination prescriptions and other instructional resources linked to the exam (Fernandez, 2018). IELTS-specifically, the official website, which can be accessed at <http://www.elts.org/>, provides details about what the IELTS is, the format of the test, the rules that apply, as well as information on instruction and research. In addition, the material is not fixed but rather adjusts to new circumstances and develops throughout the course of time. The IELTS Research Note, which can be found at <http://www.cambridgeenglish.org>, is a compilation of research that is relevant to the IELTS test and aims to evaluate it for the purpose of improving it.

Reliability

The International English Language Testing System (IELTS), which is not accredited, does not provide candidates with a pass or fail grade. According to IELTS (2017), experimental and generalizable research indicate that the IELTS speaking test has a relatively high correlation coefficient despite the lack of scoring reliability or statistical

significance. This is the case despite the fact that the test does not have statistical significance. The parts that follow will focus on analyzing two factors that have an impact on the reliability of the test:

Marker Variability

Because the IELTS speaking test is subjectively scored, there will inevitably be some difference in the marking of the test. This is an unavoidable result of the rating system. Two types of grading that are carried out by raters are known as inter-rater and intra-rater grading Karim & Haq, (2014). Intra-rater reliability refers to the consistency with which the same rater repeatedly evaluates the same performance based on the same criteria, whereas inter-rater reliability refers to the consistency with which many raters agree on the same performances. The IELTS Speaking test is administered by a single rater, which means that the interviewer's subjectivity and unpredictability determine a candidate's stated proficiency level. This level may not correspond to the candidate's inherent ability due to the fact that the IELTS Speaking test is administered by a single rater Karim & Haq, (2014). Test raters might be classified as either "the difficult interviewers" or "the easy interviews," according to Bijani, H., & Khabiri (2017). In particular, difficult interviewers place an emphasis on the complex abilities of speculating and defending ideas. This is because they want to see how well candidates can think on their feet. For instance, they could stop candidates in the middle of the previous task to ask them another important question before they finish the assignment. On the other hand, interviewers who are easy on the eyes have a propensity to ask questions in a style that is uncomplicated and economical, and they don't bother test-takers too often with argumentative inquiries. Additionally, students may select open-ended questions and demonstrate scaffolding behavior. Cultural expectations under test condition may also play a significant effect in the test's dependability. Examiners from various cultures may place greater emphasis on distinct parts of language creation (Fulcher, 2006).

Conditions for Test

There is a possibility that the conditions of the test will have an impact on the results as well as the reliability of the test. The key factors that determine the accuracy of a test are the conditions of the testing environment, the compatibility of the testing partner, and the testing procedure Karim & Haq, (2014). In general, the impact of the physical setting on the IELTS speaking test can have both a beneficial and a negative effect on the candidate's performance. The testing environment that takes place indoors is not affected by the weather conditions outside, which eliminates one potential source of error. In addition, taking the exam in a room that is closed off from the rest of the building provides test takers with an atmosphere that is free from the distractingly loud sounds that can be heard outside Bachman, L. F., & Palmer, (2010). However, the face-to-face examination settings in real time may rise to stress and anxiety between an interlocutor and a candidate. This is especially true in situations where applicants cannot hear from the examiner at first or find it difficult to complete demanding tasks Nakatsuhara et al., (2017). When conducting face-to-face examinations, it is essential to

keep eye contact with the examiner at all times and to have a respectfully close social distance. Examiners who view candidates who prefer to stand at a distance or who are too timid to look directly at the examination as dishonest or disrespectful may have an impact on a large number of candidates who prefer to do either of these things. Given how the IELTS speaking test is administered, the interview format may be the only method that can accurately evaluate a candidate's ability to communicate orally. In a setting as formal and constrained as this one, it is not possible for everyone to perform well while also feeling at ease. For instance, the same test-taker may perform admirably in the format of role-playing assignments but not so well in the format of multiple-choice questions with fixed answers. According to Hughes (2003), "the addition of additional items will result in an increase in the test's reliability."

Practicality

According to Hasrol, S. B., Zakaria, A., & Aryadoust (2022), "utility" should be understood as a superordinate concept that incorporates dependability, authenticity, and interactivity, in addition to usefulness. The practicability of a test is an essential component in its overall evaluation because it is one of these characteristics. The practical considerations that are of the utmost importance include effectiveness, management, scoring, and the interpretation of findings Karim & Haq, (2014). In addition, practicability refers to the process of giving the test, not the analysis of the results of the test Fulcher & Davidson, (2007). A test is also guaranteed to be useful by using a balanced cycle model, which ensures that its reliability, authenticity, interactivity, and usability all have the same impact on the test's results Quaid, (2018). There is no denying that the ease with which the IELTS speaking test can be administered contributes to the test's reputation for being relatively realistic. Due to the fact that the IELTS exam has been given for more than three decades, the examination procedure and safety precautions, examiner training system, and evaluation researches have been widely and methodically implemented. The IELTS speaking test has a short testing period and straightforward procedures, both of which contribute to the test's high level of practicability. Participants will maintain their energy and experience less fatigue if the duration of the speaking test is reduced to between 11 and 14 minutes, which is the new maximum. However, there is no test that is completely representative of the real world. The continued use of IELTS oral proficiency interviews is a waste of time, money, and resources consisting of both people and materials. Candidates have a favorable view of computer-based speaking tests as being less intimidating and stressful than oral proficiency interviews for the IELTS Quaid, (2018). In conclusion, despite the fact that it has some inherent drawbacks, the IELTS speaking test provides a relatively high level of practicability due to the fact that it lasts for such a short amount of time and is so easy to administer.

Suggestion

The Consequence of an Intervention

According to the analysis in Section 4, the Speaking component of the IELTS exam can in fact be completely authentic. However, language functions such as

comparing, summarizing, expressing, proposing, contrasting, narrating, imitating, and analyzing are frequently used in a candidate's test result. Questioning skills, on the other hand, are rarely used in any of the three testing elements Seedhouse & Morales, (2017). In this particular scenario, it is suggested that a two-minute interruption be added, as well as the addition of a fourth segment following the two-way discussion. There are a variety of entry points for candidates' inquiries directed toward examiners to choose from. For instance, the candidate could ask the examiner questions regarding topics that are covered in Sections 2 and 3. Participants in the exam have the option of asking a follow-up question based on the information presented in the first paragraph of the section. Due to the fact that the questioning context of the IELTS speaking test and university group settings are so comparable, the additional section's objective is to provide candidates who sign up for the IELTS primarily for the purpose of obtaining a higher education with the opportunity to play a more active role Nakatsuhara et al., (2017). In addition, the manner in which the candidates were questioned might have resulted in a more natural and reciprocal conversation in the preceding part of the interview. In addition, by asking candidates questions, raters are able to comprehensively validate their grade decisions Seedhouse & Morales, (2017).

Video-conferencing Transmission

It is possible that performance characteristics such as eye contact and close proximity influenced the performance of test takers and, as a consequence, the conclusions of the evaluation component that came before it. Physical testing environments can cause emotional factors such as anxiety and tension, which can lead to unexpected results, particularly for individuals who are not accustomed to handling emergency situations. This is especially true for individuals who are not accustomed to handling emergency situations Bachman, L. F., & Palmer, (2010). Participants in an aptitude test who live in remote areas are required to make an investment of effort, time, and money in order to attend the examination that has been predetermined. One possible solution to this problem is to communicate with people all over the world using video conferencing. According to the findings of a study conducted by Nakatsuhara et al., (2017) the operational outcome, examiner interviewing, and rating behaviors are different depending on whether the delivery mode is face-to-face or video-conferencing. The participants' interest in and comprehension of the video-conferencing style of communication is evidenced, for example, by their enhanced use of negotiating and signaling. Test takers are able to communicate with raters in real time by utilizing audio and video in two or more sites through the use of video-conferencing communication, which has the added benefit of bringing together applicants and raters who are located on different continents.

Double-Marking Procedures

As was shown in the section on evaluation, one of the most important aspects that goes into deciding how reliable a test is is the inescapable marker variability that is inherent in the structure of face-to-face tests Karim & Haq, (2014). Inter-rater dependability plays a crucial role in the IELTS speaking exam because the test is

administered by a single rater. The unreliability of the exam is exacerbated by the fact that it is scored by a single person, each of whom brings their own set of personal preferences to the task. After getting a "actual" score during the test sessions, it is recommended that the examiner give a second examiner the opportunity to double-mark the audio or video recording of the testing session. When compared to the operationalization of large-scale exams and the enormous expenditures of double marking several times in the real world, this type of double-mark-record method appears to be more practicable. In addition, the rapid development of computer technology in the present day facilitates the collection and transmission of recorded performances Karim & Haq, (2014). The importance of non-live double-marking techniques is that they allow testers to uncover both positive and bad traits that would have gone unnoticed during the single interview. In addition, the test's trustworthiness is enhanced by the fact that its numerous markers are scored by multiple scorers Karim and Haq, (2014). Consequently, systems that include double marking contribute somewhat to an increase in the reliability of scores.

CONCLUSION

The speaking portion of the IELTS test is generally considered to be authentic and reliable. In particular, the test is valid with regard to both its face validity and its substance validity. The content validity of the examination is ensured thanks to a brief complementarity that exists between three academic disciplines that are not communicatively related to one another. Since both the format and the information that pertains to the test are easily accessible, the appearance of the test can be trusted. It is possible for a candidate to take the IELTS speaking test if they are effectively preparing for it by making use of tools and resources. However, there are some ways in which a test can be considered invalid. Due to the lack of questioning ability, it is difficult to produce a topic shift or introduction, which has implications for the validity of the content. Furthermore, the IELTS speaking test, which is administered in the same format and consists of the same material for students of all levels, is not appropriate for all situations. In addition, the institution's discourse frequently fails to accurately predict how candidates will perform professionally in the future due to the examiners' tendency to draw speculative conclusions. There is a scope for improvement in both the test circumstances and the indicator variability of the IELTS speaking test. The analysis comes to a close with a discussion of three potential recommendations based on the findings of the evaluation. To begin, there is the possibility of inserting a two-minute intervention directly after the portion of the fourth section that is dedicated to a two-way discussion. Second, one might take into account videoconferencing as a method of service delivery. Thirdly, for audio or visual techniques, double-marking can improve validity and dependability. Henceforth, the deep analysis of the research can be an upright insight to present the better IELTS speaking test.

REFERENCES

Ameri-Golestan, A., & Nezakat-Alhossaini, M. (2013). Efficacy of Explicit Teaching

- and Consciousness-Raising in IELTS Writing. *Procedia - Social and Behavioral Sciences*, 70, 853–858. <https://doi.org/10.1016/j.sbspro.2013.01.130>
- Bachman, L. F., & Palmer, A. S. (2010). Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics, 2010, Vol. 10, No. 2 Book Review. *Language Testing*, 10(2), 41–46.
- Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford.
- Bijani, H., & Khabiri, M. (2017). Investigating the Effect of Training on Raters' Bias toward Test Takers in Oral Proficiency Assessment: A FACETS analysis. *Journal of Asia TEFL*, 14(4), 687.
- Brown A. (2006). Candidate discourse in the revised IELTS speaking test. *IELTS Research Report*, 6, 71–90.
- Carey & Mannell. (2009). The contribution of interlanguage phonology accommodation to inter-examiner variation in the rating of pronunciation in oral proficiency interviews. *IELTS Research Reports*, 9, 217–236.
- Chalhoub-Deville, M., & Turner, C. E. (2000). What to look for in ESL admission tests: Cambridge certificate exams, IELTS, and TOEFL. *System*, 28(4), 523–539. [https://doi.org/10.1016/S0346-251X\(00\)00036-1](https://doi.org/10.1016/S0346-251X(00)00036-1)
- Fernandez, C. J. (2018). Behind a spoken performance: test takers' strategic reactions in a simulated part 3 of the IELTS speaking test. *Language Testing in Asia*, 8(1). <https://doi.org/10.1186/s40468-018-0073-4>
- Fulcher, G. (2006). Language testing. *Language Teaching*, 39(4), 291–294. <https://doi.org/10.1017/s0261444806243854>
- Fulcher, G., & Davidson, F. (2007). Language testing. In *Language Teaching* (Vol. 40, Issue 1). <https://doi.org/10.1017/s0261444806244112>
- Green A. & Hawkey R. (2012). Re-fitting for a different purpose: A case study of item writer practices in adapting source texts for a test of academic reading. *Language Testing*, 29(1), 109–129.
- Hasrol, S. B., Zakaria, A., & Aryadoust, V. (2022). A systematic review of authenticity in second language assessment. *Research Methods in Applied Linguistics*, 1(3).
- Himelfarb, I. (2019). A primer on standardized testing: History, measurement, classical test theory, item response theory, and equating. *Journal of Chiropractic Education*, 33(2), 151–163.
- Hughes, A. (2003). *Testing for Language Teachers*. Cambridge University Press. <https://doi.org/https://doi.org/10.1017/CBO9780511732980>
- Im, G. H., Shin, D., & Cheng, L. (2019). Critical review of validation models and practices in language testing: their limitations and future directions for validation

- research. *Language Testing in Asia*, 9(1), 1-26., 9(1), 1–26.
- Iwashita N. & Vasquez C. (2015). Examination of discourse competence at different proficiency levels in IELTS speaking part 2. *IELTS Research Reports*, 5, 1–44.
- Karim, S., & Haq, N. (2014). An Assessment of IELTS Speaking Test. *International Journal of Evaluation and Research in Education (IJERE)*, 3(3), 152–157. <https://doi.org/10.11591/ijere.v3i3.6047>
- Khabbazzbashi N. (2017). Topic and background knowledge effects on performance in speaking assessment. *Language Testing*, 34(1), 23–48.
- Khatimah, H., Yanelda, S., & Muslem. (2022). Delving Debaters' Self-Development Strategies in Enhancing Their Speaking Performance. *JL3T (Journal of Linguistics, Literature and Language Teaching)*, 8(1), 39–47. <https://doi.org/10.32505/jl3t.v8i1.3982>
- Li, J. (2019). An evaluation of IELTS speaking test. *Open Access Library Journal*, 6(12), 1–17.
- Liao, L. (2021). *Examining IELTS score consistency across testing sites The question of test fairness (49-56)* . April.
- Moland, C. W., & Oetting, J. B. (2021). Comparison of the Diagnostic Evaluation of Language Variation–Screening Test Risk subtest to two other screeners for low-income prekindergartners who speak African American English and live in the urban South. *American Journal of Speech-Lang*, 30(6), 2528-2541.v.
- Nakatsuhara, F., Inoue, C., Berry, V., & Galaczi, E. (2017). *IELTS Partnership Research Papers. Phase 2*, 1–67.
- Pearson, W. S. (2019). ‘Remark or retake’? A study of candidate performance in IELTS and perceptions towards test failure. *Language Testing in Asia*, 9(1), 17.
- Peltekov, P. (2021). The International English Language Testing System (IELTS): A Critical Review. *Journal of English Language Teaching and Linguistics*, 6(2), 395. <https://doi.org/10.21462/jeltl.v6i2.581>
- Quaid, E. D. (2018). Reviewing the IELTS speaking test in East Asia: theoretical and practice-based insights. *Language Testing in Asia*, 8(1). <https://doi.org/10.1186/s40468-018-0056-5>
- Read, J. (2022). Test Review: The International English Language Testing System (IELTS). *Language Testing*, 39(4), 679–694.
- Seedhouse, P., & Nakatsuhara, F. (2018). *The discourse of the IELTS speaking test: interactional design and practice*. Cambridge University Press.
- Seedhouse, P., & Morales, S. (2017). Candidates questioning examiners in the IELTS Speaking Test: An intervention study. *IELTS Research Reports Online Series*.

- Shums'kyi, O. L. (2020). USE OF IT FOR FUTURE FOREIGN LANGUAGE TEACHERS' LINGUISTIC SELF-EDUCATION. *Information Technologies and Learning Tools*, 75(1), 269–278.
- Spolsky, B. (1997). Language Testing: Art or Science in Nickel, G (ED). *Proceedings of the Fourth International Congress of Applied Linguistics*, 3.
- Taylor, L. (2001). *Revising the IELTS Speaking Test: Developments in Test Format and Task Design*.
- Taylor L. (2007). *The impact of joint-funded research studies on IELTS speaking module*. Cambridge University Press.
- Tukhtabaeva, Z. (2020). Analyzing Testing Tools According to the Five Principles of Language Assessment. *Bulletin of Science and Practice*, 6(6), 298–304. <https://doi.org/10.33619/2414-2948/55/39>
- Wigglesworth & Elder. (2010). An investigation of the effectiveness and validity of planning time in speaking test tasks. *Language Assessment Quarterly*, 7(1), 1–24.