New Perspectives on IELTS Authenticity: An Evaluation of the Speaking Module

Marzieh Souzandehfar[1]*

ARTICLE INFO	ABSTRACT
	This study represents the inaugural attempt at assessing the authenticity of the tasks encompassed in the IELTS Speaking Module. The evaluation is conducted from the vantage points of applied linguistics and general education, and serves to enhance comprehension of authenticity and authentic assessment. In order to achieve this objective, an analysis was conducted on the Speaking Module tasks using Bachman and Palmer's model of test usefulness from the discipline of applied linguistics, as well as Herrington and Herrington's inventory of the fundamental attributes of genuine assessment from the realm of general education. The results of both task analyses revealed low indices of authenticity of the tasks. The high degree of convergence of the results from the analyses could open new horizons for experts in applied linguistics to exchange ideas about authentic language assessment with those in general education. Furthermore, such characteristics as problem-solving skills, higher-order thinking skills, integrated assessment, and multiple indicators of learning, are not included in Bachman and Palmer's framework.

1. Introduction

The notion of authenticity has always been open to debate within the fields of applied linguistics as well as general education. In applied linguistics, the idea emerged in the late 1970s when communicative methodology was gaining importance and there was a growing interest in teaching and testing 'real-life' language. On the other hand, in general education, the notion was recognized a decade later. Since then, there has been much overlap in the definitions in both fields, yet the debates have remained largely independent of each other (Lewkowicz, 2000).

As a result, Lewkowicz (2000) proposed that there should be a closer connection between discussions about authenticity in applied linguistics and general education. This connection would lead to a more comprehensive understanding of authenticity and authentic assessment. Additionally, he underscored the importance of grounding these discussions in empirical evidence in order to provide practical insights to a debate that has largely relied on theory.

In applied linguistics, authenticity has been considered as a central issue in validating test tasks, especially when it comes to direct speaking assessments (Ismail et al., 2023; Moore & Morton, 1999; Peltekov, 2021). Authentic test tasks refer to assessments that closely resemble the tasks that language users are expected to encounter in the domain of target language use (TLU) (Bachman & Palmer, 1996). In fact, unlike traditional assessment which primarily focuses on evaluating a student's understanding and mastery of knowledge, typically limited to a written test, authentic assessment encompasses various methods beyond writing tests to gauge a student's learning outcomes. In fact,

---

[1] Department of Translation Studies, Faculty of Humanities, Jahrom University, Jahrom, Fars, Iran. souzandeh@jahromu.ac.ir, mzsouzandeh@gmail.com

authentic assessment pertains to evaluating students based on their ability to perform real-world tasks that showcase their meaningful application of essential knowledge and skills.

On the other hand, the majority of prestigious universities globally necessitate students to achieve a particular score on an established English language assessment as a prerequisite for admission. An illustration of this is the IELTS or International English Language Testing System, which is administered to assess the English language abilities of students from non-English speaking nations. The IELTS test is offered in two different formats, namely Academic and General Training. The outcomes of the examination will ascertain the proficiency level of students in the four primary skills of the English language.

As a result, the evaluation of the authenticity of the academic IELTS test, as a key determinant in the admission of international students in a majority of universities worldwide, holds considerable importance. This gains even more significance when it comes to the speaking skill which is considered the most vital skill in an EFL/ESL context, including an academic one (Amelia et al., 2022; Rao, 2019). Consequently, the current study was intended to investigate the degree of correspondence between tasks in the IELTS speaking test and those in the TLU domain, i.e. those that students are required to undertake in university study.

For the first time, this study evaluated the authenticity of the IELTS speaking module tasks by bringing the authenticity discussions within the fields of applied linguistics and general education closer to each other to provide a better comprehension of the notion of authenticity and authentic assessment.

Furthermore, rather than focusing on more specific aspects like predictive validity, washback effect, and impact of the IELTS test on educational programs, the current study took up the issue of authenticity as a more comprehensive and umbrella term to be investigated as its primary objective.

Finally, while the traditional view of the communication needs of international students in tertiary education has given more prominence to reading, writing, and to a lesser extent, listening skills, it is now generally accepted that oral communication skills are as important as literacy skills (Carroll & Ryan, 2005). Since it has been demonstrated that long-term learning depends on the learner actively processing the material that they read and that there are cognitive benefits to verbal participation (Gagne et al., 1994), recent years have seen a shift in teaching methodology within universities to a more participatory style of learning.

As a result, given the increased value placed on interactive communication in the classroom, the present study more specifically focused on the authenticity of the IELTS Speaking Module, which is the same for both academic and general IELTS tests, rather than other modules to see whether the test which is designed for entrance to higher education is adequately assessing relevant interactional skills.

2. Review of Literature

During the late 1970s, in the field of applied linguistics, the concept of authenticity emerged because of the growing importance of communicative methodology and an increased focus on teaching and assessing practical language in real-life situations (Lewkowicz, 2000). However, in the realm of general education, the notion was recognized a decade after it was first introduced. Since then, there has been a considerable amount of agreement in the definitions of these areas. However, the conversations about them have mostly taken place independently

2.1. Authenticity Debate in Applied Linguistics

In the field of applied linguistics, Widdowson (1978) made a distinction between the 'genuineness' and 'authenticity' of language, suggesting that genuineness pertains to the inherent nature of the text and is an objective aspect. Widdoson argued that the relationship between the passage and the reader determines authenticity, which pertains to the reader's appropriate reaction.

In 1985, Breen proposed a new perspective on the concept of "authenticity." Rather than viewing authenticity as a singular idea, Breen argued that it is closely connected to texts, learners' understanding of those texts, the activities they engage in, and the social settings of the language classroom. He suggested that the concept of authenticity is quite intricate and that it is overly simplistic to categorize materials as either authentic or inauthentic.

Bachman developed the concepts proposed by Widdowson and Breen during the early 1990s. He proposed the idea of differentiating between two forms of authenticity: situational authenticity, referring to the alignment of test tasks with the use of the target language, and interactional authenticity, which focuses on the interaction between the test taker and the test task. He asserted that being genuine was not simply about aligning test tasks with target language use (TLU) tasks. He believed that authenticity was the result of the test takers actively participating in the test tasks.

In 1996, Bachman and Palmer took a significant leap by distinguishing the concept of authenticity from that of interactiveness. Bachman and Palmer (1996) described authenticity as the extent to which the features of a language test task align with the traits of a task in a TLU context. The given definition aligns with situational authenticity, while interactiveness has taken the place of what was formerly known as interactional authenticity. The reason for this change was to acknowledge that real-life tasks are always authentically based on specific situations. Therefore, authenticity can only be a characteristic of different tasks, such as those utilized for testing or teaching.

Except for some random and sporadic studies on authentic assessment in applied linguistics (Aryadoust, 2023; Filipi, 2015; Harsol et al., 2022; Ismail et al., 2023; Souzandehfar, 2018; Spottl et al., 2022; Staples et al., 2018), most of the other studies (Ducasse & Brown, 2011; Ingram & Bayliss, 2007; Karim & Haq, 2014; Li, 2019; Manzari, 2023; Moore & Morton, 1999; Razmi et al., 2021; Rea-Dickins et al., 2007; Sarab & Rahmani, 2023; have focused on more specific aspects like predictive validity, washback effect, and impact of second language tests on educational programs. In fact, authentic assessment as one of the most important concepts in language assessment is not paid enough attention to.

2.2. Authenticity Debate in General Education

The most widely recognized guidelines for defining authentic assessment in general education were developed by Herrington and Herrington (1998; 2006), prominent figures in the field of authentic assessment. The guidelines were divided into four categories: context, student factors, task factors, and indicators. The initial requirement for authentic assessment is that the context should accurately represent the real-life conditions in which the performance will take place. The student factor or student's role necessitates students to effectively apply their acquired knowledge and produce refined, impressive performances or products. In addition, it necessitates considerable dedication and teamwork from students. In regards to authentic activity or task factors, the test questions should encompass intricate and ambiguous obstacles that call for decision-making and a wide range of tasks. Furthermore, this standard necessitates that the evaluation is smoothly incorporated into the task. The final factor, namely indicators is concerned with multiple indicators of learning. It also requires achieving validity and reliability with appropriate criteria for scoring varied products.

Likewise, in an online setting, Herrington et al., (2002) devised ten standards for identifying an authentic task. Their work largely embodies the characteristics recognized by Herrington and Herrington (1998; 2006). It specifically focuses on the applicability of knowledge outside the classroom, the variety of results, intricate tasks, and integration with evaluation.

The third approach is a framework created by Gulikers et al. (2006) that has five dimensions. The first and second approaches have already encompassed these dimensions and they do not signify any extra characteristics of genuine assessment.

The next approach is built upon the findings of Frey and Schmidt (2007), who identified various factors that characterize authentic assessment. These factors include the type of stimuli used, the level of complexity involved, the conditions and resources available, the resulting consequences, and whether the tasks are determined by an assessor or the student.

Keyser and Howell (2008) have implemented another approach which presents the characteristics of authentic evaluation in a different way. Their method identifies the specific features emphasized in previous approaches, even though they use slightly different terminology.

Burkill et al., (2009) present the sixth and final approach which gives equal importance to both the product and the process. It focuses on developing practical skills and advanced mental abilities such as analyzing, combining, and assessing information. It combines various skills to complete a comprehensive project, while also fostering creativity and generating original thoughts and

reactions. These characteristics mostly align with the features that were identified in the previous approaches.

As indicated by the previous overview, there have been two separate discussions on the topic of authenticity that have mostly overlooked each other. One has taken place in applied linguistics, while the other has occurred in general education. Lewkowicz (2000) proposes the need for closer integration of authenticity discussions in applied linguistics and general education to enhance our understanding of authenticity and authentic assessment. In addition, he highlights the importance of grounding these conversations in empirical evidence in order to add practical knowledge to what has been primarily a theoretical discussion. Consequently, this study was conducted in reaction to this request.

The purpose of this study was to examine the authenticity of the tasks in the IELTS Speaking Module from the perspective of both applied linguistics and general education. This was done by incorporating the model of test usefulness by Bachman and Palmer (1996; 2010) from the field of applied linguistics and the essential characteristics of authentic assessment by Herrington and Herrington (1998; 2006) from general education field. More specifically, this study attempts to address the following research questions in more detail:

1.   To what extent do the characteristics of the IELTS Speaking Module tasks correspond to those of TLU tasks, based on Bachman and Palmer's (1996; 2010) model of test usefulness?

2.   To what extent do the characteristics of the IELTS Speaking Module tasks correspond to Herrington and Herrington's (1998; 2006) essential characteristics of authentic assessment?

3.   Method

This study involved two different analyses of IELTS Speaking Module tasks. One analysis was based on Bachman and Palmer's (1996; 2010) model of test usefulness in applied linguistics, while the other analysis was based on Herrington and Herrington's (1998; 2006) authenticity criteria in general education.

3.1. Participants

In addition to the researcher, three Ph.D. students (two females and one male) in TEFL at Shiraz University were invited to cooperate with the task analyses and ratings. The students had all passed language testing courses and were completely familiar with Bachman and Palmer's (1996; 2010) model of test usefulness. As for Herrington and Herrington's (1998; 2006) authenticity criteria from the realm of general education, all the raters were introduced to the characteristics of an authentic test in detail.

3.2. Instruments
3.2.1. Bachman and Palmer's (1996; 2010) Model of Test Usefulness

For the first task analysis, Bachman and Palmer's (1996; 2010) model of test usefulness was utilized. In 1996, Bachman and Palmer proposed a model of test usefulness that includes six test qualities – reliability, construct validity, authenticity, interactiveness, impact, and practicality. Bachman and Palmer (1996) separated the notion of authenticity from that of interactiveness, defining authenticity as 'the degree of correspondence of the characteristics of a given language test task to the features of a TLU task' (Bachman and Palmer, 1996, p. 23). To find the degree of correspondence between test and TLU tasks – that is, to determine the authenticity of test tasks – Bachman and Palmer (1996) proposed a framework of task characteristics. This framework provides a systematic way of matching tasks in terms of their setting, the test rubrics, test input, the outcome the tasks are expected to give rise to, and the relationship between input and response (see Appendix A). More specifically, authenticity evaluation, in Bachman and Palmer's (1996) framework, involves two stages:

1.   providing a rich description of the TLU domain, and
2.   finding the degree of correspondence of tasks in the TLU domain to test tasks.

Bachman and Palmer (1996; 2010) argue that if the description of the TLU domain is not complete and detailed, this information cannot be used to evaluate the degree of correspondence between TLU tasks and test tasks.

3.2.2 Herrington & Herrington's (1998; 2006) List of Essential Characteristics Of Authentic Assessment. For the second task analysis, Herrington & Herrington's (1998; 2006) list of essential characteristics of authentic assessment was utilized from the field of general education. The list consists of four categories: context, the student's role, authentic activity, and indicators. Using these guidelines, the assessment is most likely to be authentic if it satisfies the following criteria:

Context:
 • Requires fidelity of context to reflect the conditions under which the performance will occur (rather than contrived, artificial, or decontextualized conditions) (Meyer, 1992; Reeves & Okey, 1996; Wiggins, 1993)

Student's role
 • Requires students to be effective performers with acquired knowledge, and to craft polished, performances or products (Wiggins, 2019; Wiggins, 1993; Wiggins, 1989)
 • Requires significant student time and effort in collaboration with others (Linn et al., 1991; Kroll et al., 1992)

Authentic activity
 • Involves complex, ill-structured challenges that require judgment, and a full array of tasks (Wiggins, 2019; 1993; 1989; Linn, et al., 1991; Torrance, 1995)
 • Requires the assessment to be seamlessly integrated with the activity (Reeves & Okey, 1996; Young, 1995)

Indicators
 • Provides multiple indicators of learning (Lajoie, 1991; Linn, et al., 1991)
 • Achieves validity and reliability with appropriate criteria for scoring varied products (Wiggins, 2019; Lajoie, 1991; Resnick & Resnick, 1992).

3.2.3. Sample IELTS Speaking Module. In order to find the degree of correspondence of tasks in the TLU domain to test tasks in the IELTS Speaking Module, a sample IELTS Speaking Module was selected from the Internet (see Appendix B).

The test consists of three parts as follows:
Part 1: Introduction (4-5 mins)
General questions on familiar topics, e.g. home, family, studies, work, interests, future plans, past experiences

Part 2: Long-turn (3-4 mins)
Talk about a given topic for 1-2 minutes. Prompts are provided on a task card and 1 minute preparation time is given for you to make notes and think about what you want to say.

Part 3: Discussion (4-5 mins)
Discussion type questions related to the topic of Part 2 where you are expected to express your opinion about more abstract issues and ideas.

3.3. Data Collection
 Two of the raters had already had the experience of taking an IELTS test, which was very helpful in collecting data about the actual test characteristics and conditions. In addition, the researcher provided all the raters with a rich description of the IELTS Speaking Module tasks and in particular the actual conditions of the Speaking Module in the real context gathered from the IELTS

website (https://www.ielts.org/). In fact, the IELTS speaking test is an in-person interview conducted by a single examiner, divided into three sections, and takes approximately 11 to 14 minutes to complete. After the test is recorded, the examiner evaluates the candidate's performance and determines their band score immediately upon completion of the test.

The performance will be evaluated based on four distinct categories, including fluency and cohesiveness, vocabulary and language skills, grammar usage and accuracy, and pronunciation. The candidate will receive a band score ranging from 1 to 9.

Furthermore, a rich description of TLU was provided based on Bachman and Palmer's (1996) definition of the TLU domain as the "situation or context in which the test taker will be using the language outside of the test itself" (p., 18). Here in the current study, we specifically focused on the academic context of a university where international students who passed the academic version of the IELTS test will attend. As a result, through a discussion with the three raters, every possible academic context at a real university was detected and then based on Bachmann and Palmer's (1996; 2010) model of test usefulness, characteristics of the setting, input, expected response, and relationship between input and response in each context were fully described.

3.4. Data Analysis

3.4.1. Task Analysis 1. According to Bachman and Palmer's (1996; 2010) framework, authenticity evaluation involves two major stages. The first stage is to provide a rich description of the TLU domain, and the second step is to find the degree of correspondence of tasks in the TLU domain to test tasks. In order to analyze the first task and determine the similarity between the characteristics of the tasks in the IELTS Speaking Module and TLU tasks, a thorough description of both sets of tasks was given using the literature and Bachman and Palmer's (1996; 2010) framework of task characteristics. In Bachman and Palmer's (1996) study, it was found that there is often a lack of alignment between language use tasks and test tasks in terms of the test rubric. This is due to the fact that in everyday language, this characteristic is usually understood without being explicitly stated, whereas in a test, it is necessary to clearly and explicitly express it. Therefore, the examination rubric is excluded from the list of task characteristics in task analyses, such as the one conducted in this particular study.

In the subsequent phase, the researcher and three other Ph.D. students in TEFL were involved in the process. Students received training to analyze and compare the different qualities of TLU tasks and the speaking module of the IELTS exam, using detailed descriptions as a basis. Despite the group's prior knowledge of Bachman and Palmer's (1996; 2010) test usefulness model, the researcher made an effort to provide a comprehensive explanation of the task characteristics framework. Furthermore, they were all made familiar with IELTS Speaking Module tasks and their administration procedures and conditions. Finally, they were provided with a full description of the communication skills that are required of students in the real academic context.

In this comparison, the raters were asked to compare the two domains in terms of each characteristic and put "1" for shared characteristics and "0" for non-shared ones. This comparison provided an indicator of the degree of authenticity of IELTS Speaking Module tasks (Table 1). More specifically, the level of authenticity was stated quantitatively in terms of "the percentage of distinctive characteristics shared" (Bachman & Palmer, 1966). Furthermore, using Fleiss's Kappa, inter-coder reliability among the ratings was calculated to confirm the consistency of ratings.

3.4.2. Task Analysis 2. In the second task analysis, IELTS speaking tasks were analyzed based on Herrington and Herrington's (1998; 2006) authenticity criteria from the realm of general education. To this end, a comprehensive list of all the criteria was prepared and given to the raters. Since this framework was taken from the field of general education, more clarification was needed to be made for the raters regarding each criterion and characteristic. Using these guidelines, the same four experts tried to determine whether IELTS speaking tasks satisfy the criteria. In this analysis, the raters were asked to put "1" in cases where the test tasks fulfilled a particular criterion and "0" wherever the test tasks did not tap the essential characteristics of an authentic assessment. Finally, the extent to which test tasks satisfied the criteria of an authentic assessment was calculated in terms of

percentages for each rater (Table 2). Furthermore, using Fleiss's Kappa, inter-coder reliability among the ratings was calculated to confirm the consistency of ratings.

4. Results
4.1. Task Analysis 1

To answer the first research question, i.e. the extent to which the characteristics of the IELTS Speaking Module tasks correspond to those of TLU tasks, based on Bachman and Palmer's (1996; 2010) model of test usefulness, a rich description of the TLU domain and the test tasks was provided (Table 1) for the sake of comparison and finding out the correspondence between the two domains.

Table 1
Characteristics of TLU and Test Tasks

	Characteristics of TLU task	**Characteristics of IELTS speaking module tasks**
1 Characteristics of the setting		
Physical characteristics	*Location:* Class, conference hall, department, instructor's office, laboratories, campus, out-of-university contexts Physical conditions: quite variable	*Location:* a room Physical conditions: quiet, well lit, non-distracting
	Materials and equipment and degree of familiarity: classroom equipment, conference hall facilities, laboratory equipment, etc. most of which are familiar to lg. user	*Materials and equipment and degree of familiarity:* a chair, a desk, a pen/pencil and a piece of paper, all of which are familiar to the test taker
Participants	*In campus:* instructors, peers, university employees *Out of campus:* landlord, shopkeeper, neighbors, etc.	Interviewer and interviewee
Time of task	Day or night	Mostly Daytime

2   Characteristics of the input

Format		
Channel (aural, visual)	Aural and visual	Mostly aural
Form (language, non-language, both)	Both language and non-language	language
Language (native, target, both)	Mostly target	target
Length	Variable lengths	Relatively short
Type (item, prompt)	Item and prompt	prompt
Degree of speededness	Moderate (also depending on the situation)	Moderate
Vehicle (live, reproduced, both)	Both	Live
Language of input		
Language characteristics		
Organizational characteristics		
- Grammatical (vocabulary, syntax, phonology, graphology)	*Vocabulary:* wide range of general and specialized vocabulary *Morphology and syntax:* wide range of structures	*Vocabulary:* relatively narrow range of only general vocabulary *Morphology and syntax:* narrow range of organized structures
- Textual (cohesion, rhetorical/ conversational organization)	*Cohesion:* might be cohesive or not. *Rhetorical characteristics:* wide range in academic and non-academic contexts	*Cohesion:* Mostly Cohesive *Rhetorical characteristics:* narrow range
Pragmatic characteristics		
- Functional (ideational, manipulative, heuristic, imaginative)	*Functional:* all functions	*Functional:* mostly ideational
- Sociolinguistic (dialect/variety, register, naturalness, cultural references and figurative language)	*Sociolinguistic Variety:* standard & local *Register:* formal & informal *Cultural references & figurative lg.:* quite possible	*Sociolinguistic Variety:* standard *Register:* mostly formal *Cultural references & figurative lg.:* minimal
Topical characteristics	Wide range of academic/specialized & general topics	Narrow range of general topics

3 Characteristics of the expected response

Format

Channel (aural, visual)	Mostly Aural	Aural
Form (language, non-language, both)	Mostly language; some non-language gestures or figures might be expected as well in particular contexts	Language
Language (native, target, both)	Native	Native
Length	Wide range of lengths, depending on the task	Limited by time
Type (item, prompt)	Mostly prompt; item is also possible	Prompt
Degree of speededness	Moderate (also depends on the context)	Moderate
Vehicle (live, reproduced, both)	Mostly live	Live

Language of expected response
Language characteristics
Organizational characteristics

- Grammatical (vocabulary, syntax, phonology, graphology)	*Vocabulary:* wide range of general & specialized voc. in academic & non-academic contexts	*Vocabulary:* wide enough to discuss topics at length and make meaning clear
	Morphology & syntax: wide range but mostly organized in academic contexts	*Morphology & syntax:* a mix of simple & complex structures with some flexibility
- Textual (cohesion, rhetorical/ conversational organization)	*Cohesion:* cohesive, esp. in academic context	*Cohesion:* cohesive

Pragmatic characteristics

Functional (ideational, manipulative, heuristic, imaginative)	*Functional:* all functions	*Functional:* mostly ideational
Sociolinguistic (dialect/variety, register, naturalness,	*Sociolinguistic Variety:* standard or local	*Sociolinguistic Variety:* mostly standard
		Register: mostly

cultural references and figurative language)	*Register:* formal or informal (depending on the context) *Cultural references:* quite possible	formal *Cultural references:* minimal	
Topical characteristics	Wide range of academic and general topics	Limited general topics	

4   Relationship between input and response

Reactivity (reciprocal, non-reciprocal, adaptive)	Mostly reciprocal (non—reciprocal & adaptive are also possible)	Tasks 1 & 3 are relatively reciprocal and task 2 is non-reciprocal	
Scope of relationship (broad, narrow)	Both broad & narrow are possible	Mostly broad (narrow is also possible)	
Directness of relationship (direct, indirect)	Mostly indirect (direct is quite possible)	indirect	

Based on Table 1, the raters made a comparison between the characteristics of the tasks in the two domains. The result of this comparison has been illustrated in detail in terms of the percentage of the shared features in Table 2. Furthermore, results of the Fleiss Kappa inter-coder reliability (K = 0.6) illustrated substantial agreement among the raters.

Table 2
Degree of Correspondence (Percentage of Shared Features) between TLU and IELT Speaking Tasks

		Rater 1	Rater 2	Rater 3	Rater 4
1	**Characteristics of the setting**				
	Physical characteristics	0	0	0	0
	Participants	0	0	0	0
	Time of task	0	0	1	0
2	**Characteristics of the input**				
	Format				
	Channel (aural, visual)	0	0	0	1
	Form (language, non-language, both)	0	1	1	1
	Language (native, target, both)	1	1	1	0
	Length	1	0	0	0
	Type (item, prompt)	1	0	0	1
	Degree of speededness	0	0	0	1
	Vehicle (live, reproduced, both)	1	0	1	0

Language of input
    Language characteristics
       Organizational characteristics

- Grammatical (vocabulary, syntax, phonology, graphology)	0	0	1	0
- Textual (cohesion, hetorical/conversational organization)	0	0	1	0
Pragmatic characteristics				
- Functional (ideational, manipulative, heuristic, imaginative)	0	0	0	0
- Sociolinguistic (dialect/variety, register, naturalness, cultural references and figurative language)	0	0	0	0
Topical characteristics	0	0	0	0

3   Characteristics of the expected response

Format

Channel (aural, visual)	0	0	1	1
Form (language, non-language, both)	1	1	1	0
Language (native, target, both)	1	1	1	0
Length	0	0	0	0
Type (item, prompt)	0	1	0	1
Degree of speededness	0	0	1	0
Vehicle (live, reproduced, both)	1	1	0	1

Language of expected response
    Language characteristics
       Organizational characteristics

- Grammatical (vocabulary, syntax, phonology, graphology)	0	1	0	0
- Textual (cohesion, rhetorical/conversational organization)	0	1	0	0
Pragmatic characteristics				
- Functional (ideational, manipulative, heuristic, imaginative)	0	0	0	0
- Sociolinguistic (dialect/variety, register, naturalness, cultural references and figurative language)	0	0	0	0

	Rater 1	Rater 2	Rater 3	Rater 4
Topical characteristics	0	0	0	0
4 Relationship between input and response				
Reactivity (reciprocal, non-reciprocal, adaptive)	1	1	0	1
Scope of relationship (broad, narrow)	0	0	1	1
Directness of relationship (direct, indirect)	1	0	1	0
Degree of correspondence (Percentage)	**34.61%**	**30.76%**	**42.30%**	**34.61%**
Mean (Percentage)	**35.57%**			

As it is evident in Table 2, the degree of correspondence between the TLU tasks and IELTS speaking tasks, estimated by each rater in terms of percentages, are relatively low. They are all below 50%, with 42.30 % as the highest estimation. Furthermore, the mean of the percentages is 35.57% which is again below 50%. This shows that based on the raters' comparisons, there is only 35.57% correspondence between the IELTS Speaking Module tasks with those of TLU.

4.2. Task Analysis 2

The results of the Fleiss Kappa inter-coder reliability (K = 0.8) for the second coding revealed an even stronger agreement among the raters. This might be due to the fact that the raters became more familiar with the analysis and coding process in a more homogeneous way. With respect to the second question of the study, i.e. the extent to which the characteristics of the IELTS speaking module tasks correspond to Herrington and Herrington's (1998; 2006) essential characteristics of authentic assessment from the field of general education, the raters again found low correspondence (32.81%) between the test characteristics and the essential characteristics of an authentic test from the perspective of general education. This has been shown in terms of the percentage of the shared features in Table 3.

Table 3
Degree of Correspondence (Percentage of Shared Features) between Authenticity Criteria and IELTS Speaking Tasks

	Rater 1	Rater 2	Rater 3	Rater 4
1 Context				
fidelity of the task conditions and connectedness to the real world	0	0	0	0
2 Student factors				
problem-solving skills	0	0	1	0
higher order thinking	0	0	1	0
production of knowledge rather than the reproduction of knowledge	1	1	1	1
significant student time	1	0	0	0

collaboration	0	1	0	0
effective performers	1	1	0	1
polished products	0	0	1	0
depth of knowledge	1	0	0	1
3 Task factors				
wide range of responses	1	1	0	1
complexity	0	0	0	1
ill-structured	0	0	0	0
judgments and multiple steps	0	1	0	0
integrated assessment	0	0	0	0
4 Indicators				
multiple indicators of learning	0	0	0	0
validity & reliability through appropriate criteria for scoring	0	1	1	0
Degree of authenticity (percentage)	**31.25%**	**37.5%**	**31.25%**	**31.25%**
Mean (Percentage)	**32.81**			

Like the first task analysis, the percentages indicating the correspondence between the characteristics of the IELTS speaking tasks and Herrington and Herrington's (1998; 2006) criteria for an authentic assessment are low. This index (32.81) shows even lower correspondence than the one (35.57%) presented in Table 2.

5. Discussion

As for the first research question, the degree of correspondence between the characteristics of the IELTS Speaking Module tasks and those of TLU tasks, based on Bachman and Palmer's (1996; 2010) model of test usefulness, Table 2 showed a mean percentage of 35.57%, which is lower than 50%. According to Bachman and Palmer (1996), if the TLU domain "is very broad, or is varied, then it may be realistic to expect only a moderate level of authenticity" (p. 136). As a result, if we consider the TLU domain here in this study merely as the academic context, the level of authenticity is expected to be higher than moderate; i.e., something above 50%. But, if we include the non-academic context out of the university as well, the TLU domain will become very broad and consequently, the moderate level of authenticity will be expected. However, the results of the task analysis here are not satisfying in either case.

Having a closer look at the table 2, one can recognize several crucial characteristics upon which all the raters are agreed for lack of correspondence between the TLU tasks and test tasks. The first of these characteristics, in order of appearance in the table, is the physical characteristics. All the raters believed that there is a great difference between the TLU tasks and IELTS speaking tasks regarding the location, physical conditions, and degree of familiarity. In fact, in real life academic contexts, there is a wide range of settings like classrooms, conference halls, laboratories, departments,

instructors' offices, etc. with a great amount of variability in the their physical conditions and degree of familiarity. Furthermore, if one considers the non-academic contexts out of the university, such as the streets, shops, dorms, etc. this lack of correspondence will be felt even more.

The second important characteristic on which no correspondence was noticed by any of the raters was the participants. In real academic contexts, there are different participants involved, like instructors, peers, other university employees, etc. that can influence the performance of the students pragmalinguistically and sociopragmatically. However, in the IELTS test situation, the only participants are the interviewer and the interviewee. According to Rea-Dickins et al. (2007), the presence of different other participants in the real academic context can have a great influence on the students' communicative functions. Additionally, Ducasse and Brown (2011) argue that tests that are led by interviewers, like the IELTS speaking test, have faced criticism for not affording students the opportunity to showcase a wide range of interactional abilities because of the inherent hierarchical dynamic between examiners and candidates.

It is perhaps no exaggeration to say that lack of correspondence between TLU tasks and IELTS speaking tasks in pragmatic characteristics of the input and expected response is one of the greatest shortcomings of IELTS Speaking Module which is marked as "0" by all the raters in Table 2. Basically, the IELTS Speaking Module does not test the examinee's pragmatic competence. And if occasionally the appropriateness of language use is taken into account, it is done as a peripheral aspect to other abilities, such as fluency and coherence, lexical resource, grammatical range and accuracy, and pronunciation. Jarvis and Stakounis (2010) emphasized the role of pragmatic awareness and speaking in social contexts for EAP students. In fact, they came to the conclusion that "the reality did not match what EAP students had expected, and this led to the majority feeling somewhat disappointed or dissatisfied with both their lack of linguistic improvement and lack of contact with native people and culture" (p. 9).

Although literature has shown the difficulty of testing pragmatics (e.g., Roever, 2011), but the exclusion of this crucial aspect of language can lead to more difficulties regarding the validity of the tests. Furthermore, Roever (2011) and Tsutagawa (2012) emphasized testing learners' pragmatic competence to be used in social interactions rather than testing isolated aspects of this competence.

'Topical characteristics' is another area for which none of the raters found any correspondence between TLU tasks and IELTS speaking tasks. In the real academic context, most of the topics the students will be going to deal with are specialized and technical topics that are related to their field of study. This is in sharp contrast with the general and everyday topics that the examinees encounter in IELTS speaking tasks. This is exactly in line with Kerstjen and Nery (2000) who argued that the predictive power of the IELTS Speaking scores relied on the test takers' field of study. Consequently, as Rea-Dickins et al. (2007) stated, the IELTS Speaking Module is a poor predictor of test takers' future academic performance.

Furthermore, in other characteristics like, time of task, form of input, length of input, degree of the speededness of input and output, and language of input and output, most of the raters did not find any correspondence between the TLU domain and IELTS speaking tasks.

Regarding the second research question, the extent to which the characteristics of the IELTS Speaking Module tasks correspond to Herrington and Herrington's (1998; 2006) essential characteristics of authentic assessment, Table 3 showed even a lower percentage mean (32.81) compare to the one in Table 2. One main reason for this lower index could be the fact that comparing to Backmann and Palmer's (1996; 2010) model of test usefulness, there are some extra important characteristics of authentic assessment in Herrington and Herrington's (1998; 2006) list which was not found in IELTS speaking tasks by any of the raters. The first of these characteristics is the "context"; i.e. the fidelity of the task conditions and connectedness to the real world. According to Herrington and Herrington (1998; 2006), this characteristic is considered as the most crucial one in the evaluation of an authentic assessment. All of the raters were agreed upon the fact that there is little connectedness between the IELTS speaking tasks and those in the real academic context. In fact, the whole first task analysis could be put under this comprehensive title of the 'fidelity of the task conditions and connectedness to the real world'. In the first task analysis, the most obvious infidelities were observed in physical characteristics, participants, pragmatic characteristics, and topical characteristics.

The second characteristic which all the raters believed that the IELTS speaking tasks lack is the "ill-structured property". In fact, none of the IELTS speaking tasks are presented as ill-structured and ambiguous tasks which have unstated goals and without any predictable solution. This is while Herrington and Herrington (1998; 2006) consider this characteristic as one of the essential task factors in an authentic assessment. According to Herrington and Herrington (1998; 2006), it is recommended that authentic activities should be challenging and offer various possible interpretations instead of being easily solved using pre-existing algorithms. In order to successfully accomplish the main task, individuals are required to recognize and define their own distinct tasks and sub-tasks. Having a close look at the IELTS speaking tasks, one can easily recognize that all the tasks are well-structured with predetermined goals to be fulfilled by the examinees. According to Ducasse and Brown (2011), although informational functions, like explaining, informing, describing, etc. - that comprise most of the well-structured tasks in the IELTS Speaking Module - are important in academic situations, interactional functions, such as arguing, analyzing, questioning, negotiating, challenging and defending – as more ill-structured tasks - are more crucial for active participation in group work.

Furthermore, all of the raters agreed that the test was not "seamlessly integrated with the activity itself". It is obvious that the IELTS speaking module is administered as a test separate from the speaking activity. In fact, Herrington and Herrington (1996; 2006) emphasize that for a test to be authentic, it should not be separate from the activity itself in such a way that students do not feel that they are being tested. Although this factor is one of the essential criteria for a test to be authentic, whether this can be taken into account for high-stake tests such as IELTS is open to debate (Reeves & Okey, 1996; Young, 1995; Herrington & Herrington, 1998).

Finally, except for the examinees' score on the IELTS speaking module, there is no other indicator of their speaking ability. In other words, 'multiple indicators' is another characteristic of an authentic test which all the raters believed that the IELTS test lacks. However, if one considers the three tasks in the Speaking Module as three indicators of the examinees' speaking ability, one can decide that this characteristic of an authentic assessment is satisfied as well.

In addition to the aforementioned criteria, there were some characteristics on which most of the raters had reached consensus that the IELTS speaking tasks lacked. These characteristics were problem-solving skills, higher order thinking, significant student time, collaboration, polished product, complexity, and judgments. These are all the characteristics of group work problem-solving or consensus building activities which, according to Ducasse and Brown (2011), are frequent in university classrooms. In these activities, students actively engage in sharing information, ideas, and opinions with their peers. They not only respond to each other's interactive actions but also seek further explanations or justifications, provide support, disagree or counter others' views, and ask for clarifications or confirmation. Rea-Dickins et al. (2007) found that students who meet the admission requirements for a program may still lack critical thinking and evaluative skills, despite scoring well on the test's sub-skills.

Moreover, Ducasse and Brown (2011) argued that there is sufficient proof of tasks that necessitate students to engage in cooperative efforts within small groups to exchange thoughts and information, as well as strive towards mutual agreement. However, based on the findings of the task analysis, there is a lack of observable instances where participants in IELTS interviews engage in the process of exchanging ideas and seeking consensus. As candidates only need to answer specific questions or propositions, there is no way to assess their skill in actively engaging in oral discussions, expressing their own ideas and knowledge, and critiquing or supporting others' contributions.
However, Ducasse and Brown (2011) questioned the belief that a test like IELTS should encompass all or most of the language features that are present in the specific context it is meant to evaluate, such as university classes. They further mentioned that any endeavor of that kind would pose challenges in terms of feasibility and comprehensiveness.

Consequently, it seems that the results of the two task analyses converge to a large extent. One interpretation is that the characteristics of an authentic assessment from the two fields are closely connected to the necessary characteristics that exist in TLU tasks but not in the IELTS Speaking Module tasks. Furthermore, since the literature confirms the existence of these characristics in the real-life academic context, it can be concluded that the criteria which are proposed for an authentic assessment in the field of general education are complementary to those in applied linguistics. This

aligns perfectly with the idea proposed by Lewkowicz (2000) that the integration of discussions in applied linguistics and general education can enhance our understanding of authenticity and authentic assessment. In fact, such empirical studies, like the present one, can provide a more plausible basis for the experts from both fields to build upon and find more logical solutions to the existing problems of authentic assessment which is, according to Ingram (2003), the most significant goal of language testing. More specifically, the high degree of the convergence of the results from the analyses could open new horizons to experts in applied linguistics to exchange ideas about authentic language assessment with those in general education in order to develop more authentic language tests.

6. Conclusion and Implications

The findings of the task analyses, utilizing Bachman and Palmer's (1996; 2010) model of test usefulness from applied linguistics, and Herrington and Herrington's (1998; 2006) authentic assessment criteria from general education, indicated a notable lack of authenticity in the IELTS Speaking Module tasks.

Based on Bachman and Palmer's (1996; 2010) authenticity framework and regarding the lack of correspondence between the IELTS Academic Speaking Module tasks in terms of physical characteristics, participants, pragmatic characteristics, and topical characteristics, those who are in charge of the development and administration of these tests, can provide the examinees with a more authentic situations to compensate for the limitations and deficiencies that exist in such tests. Two solutions can be offered in this respect. In the first solution, the IELTS Academic Speaking Module can be conducted in a simulated academic context like what the ones in real universities (Souzandehfar & Soozandehfar, 2019). These simulated academic contexts could be a small classroom, a conference hall, etc. each of which with its own relevant physical characteristics, participants, and topical characteristics available to the examinee. In this way the pragmatic characteristics can also have more correspondence to those of real-life academic contexts. Although this change needs more budget and time than the present IELTS Speaking tests, creating such authentic contexts in small scales and utilizing more sophisticated methods of observation and scoring can result in a more valid test.

Another solution is that with the aid of recent developments in computer technology, like Artificial Intelligence and Virtual Reality, language test developers can provide the examinees with a virtual academic context through intelligent software that can place them in a semi-real context with similar characteristics to the real academic ones. This software can be designed in such a way as if the examinee were, for example, in the real academic context of a classroom, conference hall, or instructor's office with physical conditions, participants, and topical characteristics similar to those in the TLU domain. This idea can solve the problem of testing the pragmatic aspect of language to some extent.

Furthermore, the comparison of the results of the two task analyses can provide the experts in the field of applied linguistics with an opportunity to exchange ideas with those in general education and make better decisions in developing more authentic language assessments. In fact, there are some essential characteristics for an authentic test in the field of general education, such as problem-solving skills, higher-order thinking skills, integrated assessment, and multiple indicators of learning, that are not paid much attention to in Bachman and Palmer's (1996; 2010) authenticity framework. Taking these characteristics into account can lead language test developers to more authentic and consequently more valid tests.

7. Limitations of the Study

Although four raters, with high inter-coder reliability, cooperated with each other to do the analyses, it should be admitted that the task analyses were highly subjective. Finding more objective methods for such task analyses is highly recommended in future studies.

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author received no financial support for the research, authorship, and publication of this article.

References
Amelia, D., Afrianto, A., Samanik, S., Suprayogi, S., Pranoto, B. E., & Gulo, I. (2022). Improving public speaking ability through speech. Journal of Social Sciences and Technology for Community Service, 3(2), 322-330. https://doi.org/10.33365/jsstcs.v3i2.2231

Aryadoust, V. (2023). The vexing problem of validity and the future of second language assessment. Language Testing, 40(1), 8-14. https://doi.org/10.1177/02655322221125204

Bachman, L. (1990). Fundamental consideration in language testing. Oxford: Oxford University Press. https://doi.org/10.1191/0265532203lt268xx

Bachman, L. & Palmer A. (1996). Language testing in practice. Oxford: Oxford University Press. https://books.google.com/books?id=E0yH0NdySrQC&printsec=frontcover#v=onepage&q&f=false

Bachman, L. F., & Palmer, A. S. (2010). Language assessment in practice: Developing language assessments and justifying their use in the real world. Oxford: Oxford University Press. https://doi.org/10.1177/0265532211400870

Breen, M. (1985). Authenticity in the language classroom. Applied Linguistics, 6(1), 60–70. https://doi.org/10.1093/applin/6.1.60

Burkill, S., Dunne, L., Filer, T. & Zanstra, R. (2009). Authentic voices: Collaborating with students in refining assessment practices. In Assessment in different dimensions, proceedings of the ATN assessment conference 2009, 19-20 November 2009 (pp. 84-93). https://doi.org/10.53300/001c.5654

Carroll, J. & Ryan, J. (2005). Teaching international students: improving learning for all, Routledge, London and New York. https://doi.org/10.32674/jis.v1i2.560

Ducasse, A. M. & Brown, A. (2011). The role of interactive communication in IELTS speaking and its relation to candidates' preparedness for study or training contexts. IELTS Research Report, 12(1), 1-26. https://www.ielts.org/-/media/research-reports/ielts-rr-volume-12-report-3.ashx

Filipi, A. (2015). Authentic interaction and examiner accommodation in the IELTS speaking test: A discussion. Papers in Language Testing and Assessment, 4(2), 1-17. https://doi.org/10.58379/kdbk9824

Frey, B. & Schmitt, V. (2007). Coming to terms with classroom assessment. Journal of Advanced Academics, 18(3), 402-423. https://doi.org/10.4219/jaa-2007-495

Gagne, ED., Yekovich, CW. & Yekovich, FR. (1994). The cognitive psychology of school learning (2nd ed.), Harper Collins, New York. https://doi.org/10.2307/1167182

Gulikers, J., Bastiaens, T. & Kirschner, P. (2006). Authentic assessment, student teacher perceptions: The practical value of the five-dimensional framework. Journal of Vocational Education and Training, 58(3), 337-357. https://doi.org/10.1080/13636820600955443

Hasrol, S. B., Zakaria, A., & Aryadoust, V. (2022). A systematic review of authenticity in second language assessment. Research Methods in Applied Linguistics, 1(3), 100-123. https://doi.org/10.1016/j.rmal.2022.100023

Herrington, J. & Herrington, A. (1998). Authentic assessment and multimedia: How university students respond to a model of authentic assessment. Higher Education Research and Development, 17(3), 305-322. https://doi.org/10.1080/0729436980170304

Herrington, J. & Herrington, A. (2006). Authentic conditions for authentic assessment: Aligning task and assessment, in Critical Visions, Proceedings of the 29th Higher Education Research and Development Society of Australasia (HERDSA) Annual Conference, Western Australia, 10-12 July 2006. (pp. 146-151). https://ro.uow.edu.au/edupapers/849/

Herrington, J., Oliver, R., & Reeves, T.C. (2002) Patterns of engagement in authentic online learning environments, in Winds of Changing in the Sea of Learning, Proceedings of the 19th Annual Conference of the Australian Society for Computers in Tertiary Education (ASCILITE), Auckland, New Zealand, 8-11 December 2002. (pp. 279-286). https://doi.org/10.14742/ajet.1701

Ingram D.E. (2003). Methodology in the new millennium: Towards more authenticity in language learning and assessment. Paper to the First International Conference on pedagogies and learning, New meanings for the new millennium, University of Southern Queensland, Toowoomba, 1-4 October, 2003. https://research.usq.edu.au/item/9zw5x/multiple-representations-in-multimedia-and-e-learning-materials-an-issue-of-literacy

Ingram, D. & Bayliss, A. (2007). IELTS as a predictor of academic language performance, Part 1, in IELTS Research Reports Volume 7, IELTS Australia, Canberra and British Council, London, 137-203. https://www.ielts.org/-/media/research-reports/ielts_rr_volume07_report3.ashx

Ismail, S. M., Nikpoo, I., & Prasad, K. D. V. (2023). Promoting self-regulated learning, autonomy, and self-efficacy of EFL learners through authentic assessment in EFL classrooms. Language Testing in Asia, 13(1), 1-20. https://doi.org/10.1186/s40468-023-00239-z

Jarvis, H. & Stakounis, H. (2010). Speaking in social contexts: Issues for pre-sessional EAP students, TESL-EJ, 14(3), 1-14. https://tesl-ej.org/wordpress/issues/volume14/ej55/ej55a1/

Karim, S., & Haq, N. (2014). An Assessment of IELTS speaking test. International Journal of Evaluation and Research in Education, 3(3), 152-157. https://doi.org/10.11591/ijere.v3i3.6047

Kerstjen, M., & Nery, C. (2000). Predictive validity in the IELTS test. In R. Tulloh (Ed.), IELTS Research Reports (Vol. 6), (pp. 85-108). Canberra: IELTS Australia, Pty Ltd. https://www.ielts.org/-/media/research-reports/ielts_rr_volume03_report4.ashx

Keyser, S. & Howell, S. (2008). The state of authentic assessment. Education Resources Information Center (ERIC) database. (ERIC Document No ED503679) (10 March 2010). https://files.eric.ed.gov/fulltext/ED503679.pdf

Kroll, D.L., Masingila, J.O., & Mau, S.T. (1992). Grading cooperative problem solving. Mathematics Teacher, 85(8), 619-627. https://doi.org/10.5951/mt.85.8.0619

Lajoie, S. (1991). A framework for authentic assessment in mathematics. NCRMSE Research Review: The Teaching and Learning of Mathematics, 1(1), 6-12. http://www.fi.uu.nl/publicaties/literatuur/6279.pdf

Lewkowicz, J. A. (2000). Authenticity in language testing: some outstanding questions. Language Testing, 17(1), 43-64. https://doi.org/10.1177/026553220001700102

Li, J. (2019). An evaluation of IELTS speaking test. Open Access Library Journal, 6(12), 1-17. https://doi.org/10.4236/oalib.1105935

Linn, R.L., Baker, E.L., & Dunbar, S.B. (1991). Complex, performance-based assessment: Expectations and validation criteria. Educational Researcher, 20(8), 15-21. https://doi.org/10.3102/0013189x020008015

Manzari, A. (2023). Reliability and validity of self-assessments among Iranian EFL university students. International Journal of Language Testing, 13(1), 225-235. https://doi.org/10.22034/IJLT.2023.363420.1199

Meyer, C.A. (1992). What's the difference between authentic and performance assessment? Educational Leadership, 49(8), 39-40. https://files.ascd.org/staticfiles/ascd/pdf/journals/ed_lead/el_199205_meyer.pdf

Moore, T. & Morton, J. (1999). Authenticity in the IELTS Academic Module Writing Test: A comparative study of Task 2 items and university assignments. EA Journal, 2(1), 64-106. https://core.ac.uk/download/pdf/162201268.pdf

Peltekov, P. (2021). The International English Language Testing System (IELTS): A critical review. Journal of English Language Teaching and Linguistics, 6(2), 395-406. https://doi.org/10.21462/jeltl.v6i2.581

Rao, P. S. (2019). The importance of speaking skills in English classrooms. Alford Council of International English & Literature Journal, 2(2), 6-18. https://www.acielj.com/Papers/vol2issue2/1.ACIELJ%20-Srinu%20sir%20(1-12)%20OK.pdf

Razmi, M. H., Khabir, M., & Tilwani, S. A. (2021). A meta-analysis on the predictive validity of graduate record examination (GRE) general test. International Journal of Language Testing, 11(2), 51-63. https://files.eric.ed.gov/fulltext/EJ1318846.pdf

Rea-Dickins, P., Keily, R., & Yu, G. (2007). Student identity, learning and progression: The affective and academic impact of IELTS on 'successful' candidates. In P. McGovern & S. Walsh,

IELTS Research Reports, (Vol. 6) (pp. 56-136). Canberra: IELTS Australia, Pty Ltd. https://www.ielts.org/teaching-and-research/research-reports/volume-07-report-2

Reeves, T.C., & Okey, J.R. (1996). Alternative assessment for constructivist learning environments. In B.G. Wilson (Ed.), Constructivist learning environments: Case studies in instructional design (pp. 191-202). Englewood Cliffs, NJ: Educational Technology Publications. https://lib.ugent.be/catalog/rug01:002029648

Resnick, L.B., & Resnick, D.P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B.R. Gifford & M.C. O'Connor (Eds.), Changing assessment: Alternative views of aptitude, achievement and instruction (pp. 37-75). Boston: Kluwer. https://doi.org/10.1007/978-94-011-2968-8_3

Roever, C. (2011). Testing of second language pragmatics: Past and future. Language Testing, 28(4), 463-481. https://doi.org/10.1177/0265532210394633

Sarab, M. R. A., & Rahmani, S. (2023). Development and validation of a scenario-based teacher language assessment literacy test. International Journal of Language Testing, 13(1), 67-103. https://doi.org/10.22034/IJLT.2022.354972.1187

Souzandehfar, M. (2018). Authenticity evaluation of TOEFL iBT speaking module from the perspective of applied linguistics and general education. Iranian Journal of Applied Linguistics, 21(1). 215-261. http://ijal.khu.ac.ir/article-1-2857-en.html

Souzandehfar, M., & Soozandehfar, S. M. A. (2019). Authenticity of "Language Town" as an innovation in assessing learners' speaking ability: Moving towards a virtual language town (VLT). Two Quarterly Journal of English Language Teaching and Learning, 11(24), 289-302. https://elt.tabrizu.ac.ir/article_9638_27e519727c3f5530b80fc623cda807aa.pdf

Spöttl, C., Konrad, E., Schwarz, V., Guggenbichler, E., & Kremmel, B. (2022). Authenticity in language assessment: Revisiting the implementation of a key principle. In L. Will, W. Stadler, & I. Eloff (Eds.), Authenticity across Languages and cultures: Themes of Identity in Foreign Language Teaching and Learning (pp. 157-168). British Library and Cataloguing. https://doi.org/10.21832/9781800411050-007

Staples, S., Biber, D., & Reppen, R. (2018). Using corpus-based register analysis to explore the authenticity of high-stakes language exams: A register comparison of TOEFL iBT and disciplinary writing tasks. The Modern Language Journal, 102(2), 310-332. https://doi.org/10.1111/modl.12465

Torrance, H. (1995). Introduction. In H. Torrance (Ed.), Evaluating authentic assessment: Problems and possibilities in new approaches to assessment (pp. 1-8). Buckingham: Open University Press. https://doi.org/10.1093/acprof:oso/9780199657834.003.0001

Tsutagawa, F. S. (2012). Future directions in pragmatics assessment. Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics, 2(2), 43-45. https://doi.org/10.7916/D8JS9Q1G

Widdowson, H. (1978). Teaching language as communication. Oxford: Oxford University Press. https://www.amazon.com/Teaching-Language-Communication-Applied-Linguistics/dp/0194370771

Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. Phi Delta Kappan, 70(9), 703-713. https://doi.org/10.1177/003172171109200721

Wiggins, G. (2019). The case for authentic assessment. Practical Assessment, Research, and Evaluation, 2(2), 22-38. https://doi.org/10.7275/ffb1-mm19

Wiggins, G. (1993). Assessing student performance: Exploring the purpose and limits of testing. San Francisco: Jossey-Bass. https://doi.org/10.5860/choice.31-5581

Young, M.F. (1995). Assessment of situated learning using computer environments. Journal of Science Education and Technology, 4(1), 89-96. https://doi.org/10.1007/bf02211586

Appendix A

Task characteristics

Characteristics of the setting

    Physical characteristics
    Participants
    Time of task

Characteristics of the test rubrics

    Instructions
    Language (native, target)
    Channel
    Specification of procedures and tasks

    Structure
    Number of parts/tasks
    Salience of parts/tasks
    Sequence of parts/tasks
    Relative importance of parts/tasks
    Number of tasks/items per part

    Time allotment
    Scoring method
    Criteria for correctness
    Procedures for scoring the response
    Explicitness of criteria and procedures

Characteristics of the input

    Format
    Channel (aural, visual)
    Form (language, non-language, both)
    Language (native, target, both)
    Length
    Type (item, prompt)
    Degree of speededness
    Vehicle (live, reproduced, both)

    Language of input
    Language characteristics
    Organizational characteristics
    Grammatical (vocabulary, syntax, phonology, graphology)
    Textual (cohesion, rhetorical/conversational organization)
    Pragmatic characteristics
    Functional (ideational, manipulative, heuristic, imaginative)
    Sociolinguistic (dialect/variety, register, naturalness, cultural
    references and figurative language)

    Topical characteristics

Characteristics of the expected response
> *Format*
>> Channel (aural, visual)
>> Form (language, non-language, both)
>> Language (native, target, both)
>> Length
>> Type (item, prompt)
>> Degree of speededness
>> Vehicle (live, reproduced, both)

> *Language of expected response*
>> Language characteristics
>>> Organizational characteristics
>>>> Grammatical (vocabulary, syntax, phonology, graphology)
>>>> Textual (cohesion, rhetorical/conversational organization)
>>> Pragmatic characteristics
>>>> Functional (ideational, manipulative, heuristic, imaginative)
>>>> Sociolinguistic (dialect/variety, register, naturalness, cultural
>>>>> references and figurative language)

> *Topical characteristics*

> *Relationship between input and response*

> *Reactivity (reciprocal, non-reciprocal, adaptive)*
> *Scope of relationship (broad, narrow)*
> *Directness of relationship (direct, indirect)*

Appendix B
Speaking sample task – Part 1

Part 1 Introduction and Interview
[This part of the test begins with the examiner introducing himself or herself and checking the candidate's identification. It then continues as an interview.]

Let's talk about your hometown or village.
• What kind of place is it?
• What's the most interesting part of your town/village?
• What kind of jobs do the people in your town/village do?
• Would you say it's a good place to live? (Why?)

Let's move on to talk about accommodation.
• Tell me about the kind of accommodation you live in?
• How long have you lived there?
• What do you like about living there?
• What sort of accommodation would you most like to live in?

Speaking sample task – Part 2

Part 2 – Individual long turn

Candidate Task Card

Describe something you own which is very important to you.   You should say: where you got it from   how long you have had it   what you use it for   and explain why it is important to you

You will have to talk about the topic for 1 to 2 minutes.
You have one minute to think about what you're going to say.
You can make some notes to help you if you wish.

Rounding off questions
• Is it valuable in terms of money?
• Would it be easy to replace?

Speaking sample task – Part 3

Part 3 – Two-way discussion

Let's consider first of all how people's values have changed.

• What kind of things give status to people in your country?
• Have things changed since your parents' time?
Finally, let's talk about the role of advertising.
• Do you think advertising influences what people buy?