

TOEFL® RESEARCH SERIES

Beyond Score Correlations



A Content
Comparison
of IELTS
Academic
and TOEFL
iBT® Tests in
the Context
of a Score
Concordance
Study

Sara T. Cushing

TOEFL-RR-107

ETS Research Report
No. RR-26-03

ETS Research Report Series

EIGNOR EXECUTIVE EDITOR

Daniel F. McCaffrey
Lord Chair in Measurement and Statistics

ASSOCIATE EDITORS

Usama Ali
Senior Measurement Scientist

Beata Beigman Klebanov
Principal Research Scientist, Edusoft

Katherine Castellano
Managing Principal Research Scientist

Larry Davis
Director Research

Jamie Mikeska
Managing Senior Research Scientist

Teresa Ober
Research Scientist

Jonathan Schmidgall
Senior Research Scientist

Jesse Sparks
Managing Senior Research Scientist

Zuwei Wang
Senior Measurement Scientist

Klaus Zechner
Senior Research Scientist

Jiyun Zu
Senior Measurement Scientist

PRODUCTION EDITORS

Ayleen Gontz
Manager Editorial Services

Emily Ramsower
Manager Editorial Services

Since its 1947 founding, ETS has conducted and disseminated scientific research to support its products and services, and to advance the measurement and education fields. In keeping with these goals, ETS is committed to making its research freely available to the professional community and to the general public. Published accounts of ETS research, including papers in the ETS Research Report series, undergo a formal peer-review process by ETS staff to ensure that they meet established scientific and professional standards. All such ETS-conducted peer reviews are in addition to any reviews that outside organizations may provide as part of their own publication processes. Peer review notwithstanding, the positions expressed in the ETS Research Report series and other published accounts of ETS research are those of the authors and not necessarily those of the Officers and Trustees of Educational Testing Service.

The Daniel Eignor Editorship is named in honor of Dr. Daniel R. Eignor, who from 2001 until 2011 served the Research and Development division as Editor for the ETS Research Report series. The Eignor Editorship has been created to recognize the pivotal leadership role that Dr. Eignor played in the research publication process at ETS.

Beyond Score Correlations: A Content Comparison of IELTS Academic and TOEFL iBT® Tests in the Context of a Score Concordance Study

Sara T. Cushing, Department of Applied Linguistics & ESL, Georgia State University

Abstract

Score concordance tables are widely used by higher education institutions to compare scores from different English language proficiency tests, yet their validity depends on the extent to which the tests measure comparable constructs. This study examines the content comparability of the International English Language Testing System (IELTS) Academic and TOEFL iBT® tests in the context of a recent co-sponsored score concordance study. Moving beyond score correlations, the analysis compares test content across the four language skills—reading, listening, writing, and speaking—using published research, official test documentation, and publicly available sample materials. The comparison is framed by the validity frameworks adopted by each test provider and focuses on task characteristics, linguistic demands, response formats, and scoring criteria. Results indicate substantial overlap in the constructs assessed by both tests, particularly in reading and writing, where tasks target similar academic language skills and are evaluated using comparable criteria. More pronounced differences emerge in listening and speaking, with TOEFL iBT placing greater emphasis on academic content, integrated skills, and pragmatic inference, while IELTS includes more general listening contexts and examiner-mediated interaction in speaking. Despite these differences, both tests provide multiple opportunities for test takers to engage with extended discourse and demonstrate receptive and productive language abilities. Overall, the findings support the use of score concordance tables between IELTS Academic and TOEFL iBT, while emphasizing the need for cautious interpretation and recognition that scores should not be treated as fully interchangeable.

Keywords: IELTS Academic, TOEFL iBT®, score concordance, test comparability, language test validity, academic English proficiency, high-stakes language testing, reading skills, writing skills, speaking skills, listening skills, higher education admissions

Corresponding author: Sara T. Cushing, email: stcushing@gsu.edu

Introduction

“Good tests are all alike; every bad test is bad in its own way.” This Tolstoyan assertion might be taken at face value by test users who see a concordance table purporting to provide equivalent scores between two tests. As long as both tests are “good” (however that may be defined), it seems reasonable to assume that scores from one test can be easily converted into scores on a second test. However, this assumption rests on an assurance that, in addition to both tests being of high quality, the research behind the concordance table is sound. Unfortunately, this is not always the case; indeed, a recent review of language test concordance studies (Knoch & Fan, 2024) found that providers of language tests often fall short when it comes to fulfilling best practices in conducting concordance studies. Knoch and Fan state, “Concordance is deemed appropriate only when the tests measure related constructs, their content is judged similar, and a strong correlation exists between their test scores” (p. 683). For this reason, it is essential to go beyond correlations when determining whether two tests are indeed sufficiently alike to justify the production and use of concordance tables.

For many years, colleges and universities in English-medium contexts have required prospective international students to demonstrate English language proficiency. The two major tests used for this purpose are the International English Language Testing System (IELTS) Academic—jointly owned by the British Council, IDP IELTS, and Cambridge University Press & Assessment—and TOEFL iBT[®], owned by ETS. The two tests come out of different traditions of test development: the UK placing more emphasis on test content and expert judgment, while the US is more psychometrically based (Geranpayeh, 1994). However, there are strong similarities: both have decades of research into their reliability and validity; both include sections devoted to the traditional four language skills of reading, listening, speaking, and writing; and both target similar populations—thus providing preliminary evidence of construct and content comparability.

Recently, the test providers have co-sponsored a concordance study to provide an up-to-date set of concordance tables between the two tests (Ikeda et al., 2025). This study is based on a sample of 969 study participants who took both tests in a counterbalanced order. Ikeda et al. (2025) state the following about IELTS Academic and TOEFL iBT:

[Both tests] target the same test taker groups and are used for the same purpose (largely the evaluation of academic language ability to study in an English-speaking institution of

higher education). They also evaluate similar constructs, as evidenced by the inclusion of four sections targeting the four language skills (reading, listening, speaking, writing), and the reporting of the same types of subscores. The two tests also have some noticeable design differences, primarily in the way they evaluate speaking skills. (p. 2)

While Ikeda et al. (2025) conducted preliminary comparisons of the constructs of both tests, this report was commissioned to evaluate more thoroughly the claim that the content of both tests is sufficiently congruent to make the concordance tables meaningful and accurate. In the report, I compare the test content across the four skills of reading, speaking, listening, and writing. To complete this comparison, I relied in part on analyses done on two studies I recently completed comparing IELTS Academic and TOEFL iBT, respectively, with the Duolingo English Test (Cushing, 2025; Ren & Cushing, 2022). I also consulted the latest information available on both tests, including research reports, official sample test items on the test websites, and other publications available to the public. While neither test owner currently provides complete sample tests for download, ETS offers a full online practice test, for which registration is required, and IELTS offers downloadable samples of test items in all four skill areas.

Both IELTS and TOEFL iBT have maintained strong programs of test research and validation, but they make use of somewhat different frameworks for organizing and publishing their validation efforts.

The approach taken to validation in IELTS is primarily based on Weir's (2005) socio-cognitive validity framework. This framework is presented in the form of questions to be answered in validity research. The following questions are taken from Taylor (2013, pp. 27–29), specifically with regard to listening but applicable across all skills:

1. Test-taker characteristics: How are the physical/physiological, psychological and experiential characteristics of candidates catered for by this test?
2. Cognitive validity: Are the cognitive processes required to complete the test tasks appropriate and sufficiently comprehensive to be treated as representative of the construct?
3. Context validity: Are the characteristics of the test tasks and their administration appropriate and fair to the candidates who are taking them?
4. Scoring validity: How far can we depend on the scores which result from the test?

5. Consequential validity: What effects do the test and test scores have on various stakeholders?
6. Criterion-related validity: What external evidence is there that the test is measuring the construct of interest?

In contrast, TOEFL iBT relies on an argument-based framework based on the work of Kane (2004, 2013) and fleshed out in Chapelle (2008). A validity argument is structured through claims about test score meaning, which are justified by support for their underlying inferences (Chapelle & Lee, 2021). The TOEFL iBT validity argument is summarized for a lay audience in TOEFL (2025) in a series of propositions, as follows:

- The content of the test is relevant to and representative of the kinds of tasks and written and oral texts that students encounter in college and university settings.
- Tasks and scoring criteria are appropriate for obtaining evidence of test takers' academic language abilities.
- Academic language proficiency is revealed by the linguistic knowledge, processes, and strategies test takers use to respond to test tasks.
- The structure of the test is consistent with theoretical views of the relationships among English language skills.
- Performance on the test is related to other indicators or criteria of academic language proficiency.
- Test scores are reliable and comparable across test forms.
- The test results are used appropriately and have positive consequences.

Although they are expressed differently, both validation frameworks address the content of the test, the knowledge and cognitive processes/strategies required to respond successfully to test tasks, and (primarily for productive skills) the scoring criteria that are used to evaluate responses. These components of test validation are the most central in considering whether the two tests can be considered similar enough to warrant a concordance study. Figure 1 summarizes the most relevant items from the two frameworks to address this question. Note that the TOEFL framework is more explicitly tied to academic language ability; however, this notion is implicit in IELTS's use of "appropriate" to discuss both the test tasks and response processes.

Figure 1. Comparison of Validation Frameworks With Respect to Test Content

Comparison

IELTS	TOEFL iBT
Context validity: Are the characteristics of the test tasks and their administration appropriate and fair to the candidates who are taking them?	Tasks and scoring criteria are appropriate for obtaining evidence of test takers' academic language abilities.
Scoring validity: How far can we depend on the scores which result from the test?	The content of the test is relevant to and representative of the kinds of tasks and written and oral texts that students encounter in college and university settings.
Are the cognitive processes required to complete the test tasks appropriate and sufficiently comprehensive to be treated as representative of the construct?	Academic language proficiency is revealed by the linguistic knowledge, processes, and strategies test takers use to respond to test tasks.

Before moving on to the test descriptions, it may be worth discussing why scholars have argued that a test of academic language ability should be distinct from a test of general language ability. Norris et al. (2021) summarized the arguments succinctly in the first chapter of their edited volume *Assessing Academic English for Higher Education Admissions*. Norris et al. (2021) stated that it is essential for students whose first language is different from the language of education “to develop sufficient proficiency in the language such that they can engage successfully with the variety of communication tasks that characterize higher education” (p. 2). These communicative tasks include not only communication events stereotypically associated with higher education such as reading textbooks, listening to lectures, and taking tests, but also tasks associated with “social and navigational dimensions of academic endeavors” (Norris et al., 2021, p. 9), such as understanding a course syllabus or arranging for a study group meeting. In terms of the language itself, academic English is “associated with sophisticated lexical, grammatical, and discourse features, all of which are deployed in receptive, productive, and interactive tasks involving all four language skills” (Norris et al., 2021, p. 9). Norris et al. (2021) went on to state that the “defining characteristic” of high-stakes English language tests for academic English is the reliance on “tasks that represent communicative uses of English” (p. 10) that are relevant to academic contents. Essential academic language tasks range “from reading extended texts and listening to academic lectures to writing essays and speaking on a range of topics” (Norris et al., 2021, p. 11).

This report is organized as follows. First, I provide a general overview of both tests, presented alphabetically with IELTS Academic first and TOEFL iBT second. I consider each skill separately, beginning with the receptive skills of reading and listening followed by the productive skills of writing and speaking. The discussion of receptive skills focuses on the nature of the input to test takers (i.e., reading and listening texts), the response types, and the various subskills targeted by test items. For productive skills, I discuss the comparability of the tasks and the scoring criteria across the two tests. In each section, I discuss points of commonality and points of difference to inform an ultimate verdict of comparability for the purposes of justifying a score concordance study. The last section summarizes the results and provides recommendations.

Overview of the Two Tests

IELTS Academic

The IELTS Guide for Test Takers (IELTS, n.d.-a) states that the IELTS test is “designed to assess and award your ability to study, work, and succeed where English is spoken” (p. 2). IELTS has four sections. The listening, reading, and writing sections are given in one sitting, and the oral interview is given separately. IELTS provides both a general test (i.e., IELTS General Training) and an academic test (i.e., IELTS Academic); for the purposes of this report, I am focusing on the academic test only, which differs from the General Training test in reading and writing but not in listening and speaking.

There are both computer-based and paper-based administrations of IELTS. The test takes about 2 hours 45 minutes and is administered at testing centers worldwide. Scores are reported as an overall band score between 1 and 9, including half-bands, and as band scores for each section.

TOEFL iBT

The TOEFL construct is presented as follows: “TOEFL iBT test scores are interpreted as the ability of the test taker to use and understand English as it is spoken, written, read, and heard in college and university settings” (TOEFL, 2025, p. 4). This statement ties the test explicitly to academic contexts and emphasizes both the receptive (i.e., understanding) and productive (i.e., using) aspects of English.

TOEFL iBT¹ is administered on a computer in testing centers around the world and also at home, using remote proctoring via the TOEFL iBT Home Edition. It also contains four

sections: reading, listening, writing, and speaking. The writing and speaking sections of the test include integrated tasks (i.e., tasks that require test takers to read and/or listen to input before responding). The test, which was shortened substantially in 2023, takes approximately two hours to complete. Scores are reported on a scale from 0 to 120 for the total score and on a scale from 0 to 30 for the subscores in each of the four skills.

Table 1 compares the timing and number of tasks/items in each section for both tests. More details are provided in the skill-specific sections that follow. As the table shows, while both tests have substantial sections for each skill area, TOEFL iBT is nearly an hour shorter than IELTS, primarily due to shortened times for reading and writing implemented in 2023. Slightly more time is devoted to listening and speaking on TOEFL iBT than on IELTS, however.

Table 1. Comparison of IELTS Academic and TOEFL iBT Structure

Test section	IELTS Academic		TOEFL iBT	
	Estimated timing	Question/ Task	Estimated timing	Questions/ Tasks
Reading	60 minutes	3 reading texts, 40 items	35 minutes	2 reading texts, 20 items
Listening	30 minutes	4 listening texts, 40 items	36 minutes	5 listening texts, 28 items
Writing	60 minutes	2 tasks	29 minutes	2 tasks
Speaking	11–14 minutes	3-part interview with an examiner	16 minutes	4 discrete tasks delivered by computer
Total	2 hours, 44–55 minutes		1 hour, 56 minutes	

The following sections discuss the four language skills as they are assessed in both tests.

Reading

Reading is an essential academic skill and has been a central focus of both IELTS and TOEFL programs since their inception. Reading is both the most logistically straightforward skill to test, and the most common modality through which to test foundational language skills such as grammar and vocabulary. Numerous scholars have put forward various models of reading for academic testing; good summaries can be found in Schedl et al. (2021) and in Liu and Read

(2023). While these models differ in their details, the general consensus is that academic reading ability requires as a foundation a basic knowledge of core academic vocabulary and the ability to parse and comprehend complex sentences. Beyond these fundamental skills, readers need to be able to shift their reading strategies to accomplish a variety of reading purposes, including both careful and expeditious reading at the local and global level. Good readers are able to follow an author's overall argument in a text, make connections to previous knowledge, and infer the author's purpose and stance toward information in a text. An even fuller construct of academic reading also includes synthesizing information across text and critically evaluating reading materials.

Given the constraints of a timed administration, it is not feasible for a test to cover all aspects of academic reading, but a few general observations can be made. First, the reading passage(s) should be long enough to allow for testing some of the higher level skills such as integrating information from different parts of a text, discerning the overall organization of a text, or scanning quickly to find a specific piece of information. In terms of text types, texts should be similar to those assigned for classroom use. Texts are often either specifically academic or journalistic, but aimed at a college-educated audience, and are typically informational rather than literary. Finally, test items and response options must be carefully designed, so that higher level reading skills are required to arrive at a correct answer.

Test Content

IELTS Academic

As noted previously, the IELTS Academic reading section consists of three reading passages and a total of 40 items. The reading passages average approximately 800 words in length and are taken from books, magazines, and other publications on topics that are “of general interest to students at undergraduate or postgraduate level” (IELTS, n.d.-b). The texts vary in genre, and some may contain graphs, diagrams, or illustrations, but there is at least one text that presents a “detailed logical argument” in every form of the test.

IELTS has 11 different formats for reading items, although candidates will not encounter all of them on any given form of the test. These item formats are as follows:

- Multiple choice (i.e., factual questions)
- Identifying information (i.e., true/false/not given)

- Identifying writer’s views/claims (i.e., yes/no/not given)
- Matching information
- Matching headings
- Matching features
- Matching sentence endings
- Sentence completion
- Summary/note/table-flow-chart completion
- Diagram label completion
- Short-answer questions

The first seven item types involve selected responses, while the rest involve inserting a short answer taken directly from the reading text—typically in the form of a limited number of words (two or three). For limited-response items, answers must be spelled correctly to be counted as correct, a rule intended to facilitate high inter-rater reliability. The number of items for each item type varies across passages. These item types are intended to test the following reading skills:

- Detailed understanding of specific points
- General understanding of main points
- Scanning to find specific information
- Distinguishing main from supporting ideas
- Recognizing relationships/connections between facts in a text
- Recognizing opinions and theories

TOEFL iBT

The reading section of the TOEFL iBT consists of two reading passages taken from college-level textbooks, with 10 questions each. Questions are all selected-response or drag-and-drop; there are no limited-response items. Reading passages are approximately 700 words long, and the question types for each passage come from the following (ETS, n.d.-a):

- Factual information questions: Recognize information explicitly stated in the text
- Inference questions: Identify or understand information not explicitly stated in the text
- Vocabulary: Identify the meaning of words or phrases as they are used in a specific reading passage

- Sentence simplification: Choose a sentence that means the same as a sentence from the reading passage
- Insert a sentence: Demonstrate understanding of passage organization by inserting a sentence somewhere in a paragraph of the passage
- Prose summary: Choose three statements that express the most important ideas in a passage

Content Comparison

The content of IELTS Academic and TOEFL iBT Reading are compared in Table 2.

Table 2. Comparison of IELTS Academic and TOEFL iBT Reading

Characteristic	IELTS Academic	TOEFL iBT
Timing	60 minutes	35 minutes
Reading passages	Three reading passages (700–900 words each) Around 13 items per passage (40 total)	Two reading passages (approximately 700 words each) 10 items per passage (20 total)
Test items	11 item types: selected-response and limited-response items (1–3 words from the text)	Six item types, all selected-response or drag-and-drop
Skills focus	Items are targeted at the following skills: <ul style="list-style-type: none"> • Read for the general sense of a passage • Read for main ideas and details • Understand inferences and implied meanings • Recognize writer’s opinions, attitudes, and purposes • Follow the development of an argument 	Items are targeted at the following skills: <ul style="list-style-type: none"> • Recognize factual information • Recognize implied information • Infer writer’s purpose • Identify the meaning of words as they are used in the text • Understand the logical order of ideas • Recognize the major ideas and relative importance of information in the text
Weighting	All items weighted equally	All items weighted equally

Note. Bulleted wording in the Skills Focus row was taken from information on the test for the general public (TOEFL, n.d.-b; IELTS, n.d.-b).

As noted previously, neither test producer currently allows the general public to download complete sample tests.² Instead, this analysis used two passages from a TOEFL iBT sample test downloaded in 2024, along with three passages or passage excerpts from the 2025

IELTS Academic Reading Sample Tasks, available at IELTS (n.d.-c). These passages were analyzed using Text Inspector (<https://textinspector.com>): an online text analysis tool created by Stephen Bax, which provides a variety of readability and vocabulary statistics for texts. Selected results of this analysis are presented in Table 3.

As Table 3 shows, the analysis, using numerous metrics, indicates that all five texts are estimated to be at the C1 level of proficiency or higher, based on the Common European Framework of Reference for Languages (CEFR). Note that one TOEFL iBT passage was rated at D1, which is not a CEFR level but is defined on the Text Inspector website as “university undergraduate language use.” Some of the vocabulary metrics include the percentage of words from the Academic Word List (Coxhead, 2000) and from CEFR Levels B2 and C1, according to the English Vocabulary Profile: an online interactive resource that describes words typically mastered at various levels of proficiency (Capel, 2015). These metrics show that the reading level of the texts appear to be roughly similar and at an advanced level.

Table 3. Comparison of IELTS Academic and TOEFL iBT Reading Passages Metrics

Metric	IELTS 1 (Rockets)	IELTS 2 (Traffic)	IELTS 3 (Dung beetles)	TOEFL 1 (By-catch)	TOEFL 2 (Rome)
CEFR level	C2	C2+	C1+	D1	C1+
Total words	550 ^a	842	332 ^a	714	688
Words/sentence	23.08	30.39	19.94	26.52	18.73
Flesch-Kincaid reading grade	12.45	16.15	11.19	15.53	11.98
Academic word list (%)	12.45	16.15	11.19	15.53	11.98
B2 type (%)	11.07	14.83	10.38	14.79	15.32
C1 type (%)	2.77	4.44	2.19	2.67	4.94

^a The complete text is not publicly available; only an extract was included in the sample materials.

The difficulty of a reading test depends not only on the input text itself, but on the test items. TOEFL iBT relies on multiple choice with some drag-and-drop items, while IELTS provides a wider range of response options, including completing tables or diagrams with information from the text. The example test items that follow demonstrate that both tests include several item types that require comprehension at a level beyond an individual sentence, and thus tap into higher level reading skills.

Figure 2 is an example of a sentence insertion item from TOEFL iBT. Test takers need to determine where in the passage a sentence would best fit, which requires understanding of the flow of ideas throughout the paragraph. Specifically, finding the correct response requires realizing that “this wider possibility” has to refer back to something in the paragraph that is not presented as fact and then also recognizing that “could be provided” in the last sentence expresses a hypothetical situation.

Figure 2. Sample TOEFL iBT Item

The screenshot shows a digital test interface. On the left, a text box contains the following instructions and a target sentence:

Look at the four squares [■] that indicate where the following sentence can be added to the passage.

Some companies in the power industry are aware of this wider possibility and are planning sizable wind-farm projects in states other than California.

Where would the sentence best fit?

On the right, a passage titled "Electricity from Wind" is displayed. The passage text is as follows:

Since 1980, the use of wind to produce electricity has been growing rapidly. ■ In 1994 there were nearly 20,000 wind turbines worldwide, most grouped in clusters called wind farms that collectively produced 3,000 megawatts of electricity. ■ Most were in Denmark (which got 3 percent of its electricity from wind turbines) and California (where 17,000 machines produced 1 percent of the state's electricity, enough to meet the residential needs of a city as large as San Francisco). ■ In principle, all the power needs of the United States could be provided by exploiting the wind potential of just three states—North Dakota, South Dakota, and Texas. ■

Large wind farms can be built in six months to a year and then easily expanded as needed. With a moderate to fairly high net energy yield, these systems emit no heat-trapping carbon dioxide or other air pollutants and need no water for cooling; manufacturing them produces little water pollution. The land under wind turbines can be used for grazing cattle and other purposes, and leasing land for wind turbines can provide extra income for farmers and ranchers.

Wind power has a significant cost advantage over nuclear power and has become competitive with coal-fired power plants in many places. With

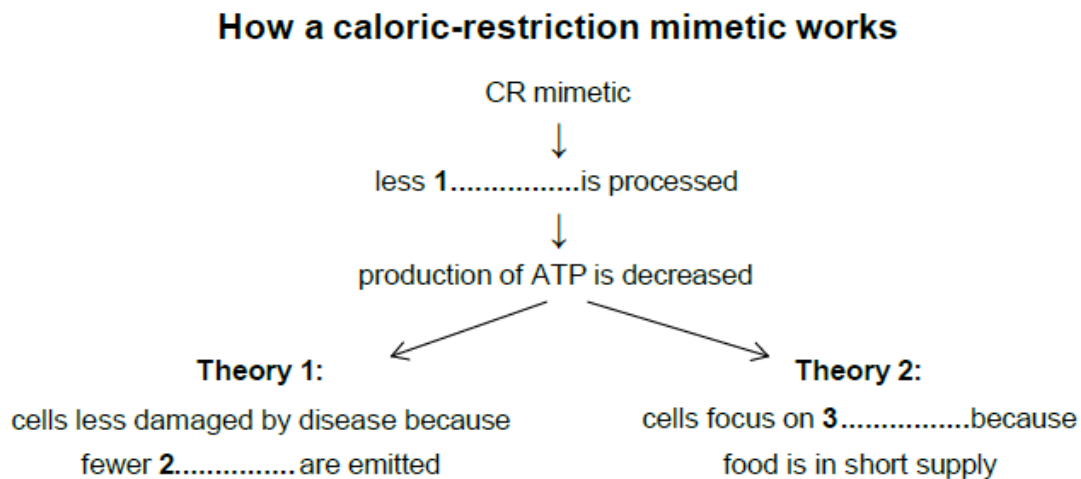
The interface includes a video player control bar at the bottom with a timestamp of 00:33 / 02:14.

Figure 3 presents an IELTS Academic item that is structured differently but similarly requires the reader to understand the flow of ideas in the passage. In this item type, test takers have to complete a flow chart or table that summarizes some of the main ideas. Note that in this particular instance, the test taker cannot simply scan the text for the correct answers. For example, the reader has to read both paragraphs to understand that 2DG is an example of a CR mimetic. In the third sentence of the second paragraph, the reader has to be able to equate “the drug” with “2DG” and realize that “it” refers back to “glucose.” These are just some examples of how readers have to make use of their knowledge of vocabulary, complex grammar, and cohesive devices such as pronouns and synonyms to make sense of academic reading passages.

Figure 3. IELTS Academic Flow Chart Item

The best-studied candidate for a caloric-restriction mimetic, 2DG (2-deoxy-D-glucose), works by interfering with the way cells process glucose. It has proved toxic at some doses in animals and so cannot be used in humans. But it has demonstrated that chemicals can replicate the effects of caloric restriction; the trick is finding the right one.

Cells use the glucose from food to generate ATP (adenosine triphosphate), the molecule that powers many activities in the body. By limiting food intake, caloric restriction minimizes the amount of glucose entering cells and decreases ATP generation. When 2DG is administered to animals that eat normally, glucose reaches cells in abundance but the drug prevents most of it from being processed and thus reduces ATP synthesis. Researchers have proposed several explanations for why interruption of glucose processing and ATP production might retard aging. One possibility relates to the ATP-making machinery's emission of free radicals, which are thought to contribute to aging and to such age-related diseases as cancer by damaging cells. Reduced operation of the machinery should limit their production and thereby constrain the damage. Another hypothesis suggests that decreased processing of glucose could indicate to cells that food is scarce (even if it isn't) and induce them to shift into an anti-aging mode that emphasizes preservation of the organism over such 'luxuries' as growth and reproduction.



Discussion

A useful framework for comparing the skills and cognitive processes for the reading sections of TOEFL iBT and IELTS Academic is found in Liu and Read (2023). As noted previously, Liu and Read argued that academic reading involves core academic language—both vocabulary and sentence structure—as well as both careful reading within and between texts and expeditious reading. This framework is based on an analysis of reading needs for university

study that comprised both the importance and the perceived difficulty of specific reading skills, as well as the feasibility of assessing each skill on a test. Another advantage of Liu and Read's framework is that it was not created with either IELTS or TOEFL iBT in mind and can thus be considered test-neutral. The results are summarized in Table 4.

Table 4. Comparison of IELTS Academic and TOEFL iBT Based on Skills and Cognitive Processes

Cognitive process	IELTS Academic Reading	TOEFL iBT Reading
Core academic language knowledge		
Understanding general academic vocabulary	X	X
Careful reading for intra-textual model building		
Integrating textual information across sentences	X	X
Inferring the situation (environment, event and relationship) implied in a text	X	X
Understanding author's point of view (such as attitudes, beliefs, and opinion)	X	X
Inferring the contextual meaning of figurative language	X	X
Careful reading for intertextual model building		
Understanding the relationships between multiple texts		(X)
Drawing implications/conclusions based on multiple texts		(X)
Expeditious reading		
Searching for specific meaning	X	X
Skimming for general idea	X	X

Note. This table uses a framework from Liu and Read's (2023) work. (X) indicates that the integrated speaking and writing tasks draw on multiple texts, but this skill is not assessed in the reading section of TOEFL iBT.

As the table shows, the reading sections from both tests assess the construct of academic reading equally thoroughly, with the possible exception of careful reading for intertextual model building. TOEFL iBT does include tasks in speaking and writing that require synthesizing information from reading and listening texts, but this skill is not specifically assessed and does not contribute to the reading score.

In summary, in the area of reading, both tests address a similar construct of academic reading in terms of the reading input and the test items themselves. The readings on both tests are from either academic sources or are similar in terms of topic and genre as well as linguistic characteristics. The items on both tests, though somewhat different in format, tap into similar reading skills. Thus, in the area of reading, a concordance appears to be warranted.

Listening

Listening is one of the most important academic skills; in fact, of the four skills, listening is the communication skill that college students engage in the most (Janusik & Wolvin, 2009). Listening is also a skill that many international students find particularly challenging, as they may have had limited previous experience with academic English (Rodgers & Webb, 2016). In addition, some research suggests that, for international students, conversational difficulties can be primarily attributed to a lack of listening competence rather than of speaking ability (Papageorgiou et al., 2021).

While listening is an essential academic skill, it can be difficult to assess for several reasons. Logistically, a listening test requires more technology than a reading test to ensure that all test takers can hear the listening input clearly and without distraction. The listening process itself is invisible, so that listening comprehension can only be inferred through the use of other language skills, such as reading to select the correct answers to multiple-choice comprehension questions or summarizing what is heard through writing or speaking. Developing a listening test also requires considering many more variables and complex decisions than developing a reading test, including the characteristics of the speaker, such as accent and gender; the type of text (i.e., monologic versus dialogic); and aspects of the text presentation, such as whether or not to include visual information, when the test items are presented with respect to the listening input, and how many times a test taker can listen to the input.

For practicality reasons, most language tests focus on one-way listening (i.e., listening to a monologue or a dialogue without interacting with the speakers), but it is important to note that two-way listening is also an important aspect of listening in academic contexts. For example, students need to participate in small group discussions or tutorial sessions, for which two-way listening is essential (Lynch, 2011). Assessing listening during interaction in real time brings with it a different set of challenges (Lam, 2021; Wagner, 2022). The use of human interlocutors to provide real-time listening opportunities introduces an element of variability that can reduce test reliability, while simulated interactions in a computer-delivered test may lack authenticity. This is one reason why many large-scale tests assess listening through one-way communication only.

As with reading, academic listening has been conceptualized in different ways by various scholars. Field (2013) provides a good overview of the cognitive processes in listening, and

Papageorgiou et al. (2021) provides a thorough review of listening for academic purposes, concluding with a construct definition for assessing English for Academic Purposes (EAP) listening that comprises three considerations: domains of listening, communication goals, and foundational abilities. As discussed previously, the three subdomains that are considered relevant to academic listening in this formulation are social-interpersonal, academic-navigational, and academic-content. The main communicative goals of listening are to understand main ideas, supporting details, relationships among ideas, inferences, opinions, speaker purpose, and speaker attitude. To accomplish these goals, listeners need to be able to process extended spoken information in real time, which involves making use of phonological information, including intonation, stress, and pause, lexical and grammatical information, and pragmatic information to comprehend meaning. At a higher level, to understand the connections between statements and ideas, listeners need to understand and process various organizational devices such as discourse markers, cohesive devices, exemplifications, and so on.

It follows from this brief discussion that a test of academic listening should include content that is representative of the various relevant subdomains and extended listening texts, which allow for test questions that tap into higher level skills such as distinguishing between main ideas and supporting points or following a speaker's line of argument. The test items themselves should be constructed in such a way that a listener cannot guess the correct answer from their background knowledge of a topic but must use appropriate listening skills and strategies to arrive at a correct answer.

Test Content

IELTS

Unlike reading, IELTS listening is the same for both the General Training and Academic tests. The listening section contains four listening texts, two dealing with everyday social situations and two with "education and training situations" (IELTS, n.d.-d) with one monologue and one dialogue for each type. There are 40 questions altogether, including multiple choice, matching, plan/map/diagram labelling, form/note/table/flow chart/summary completion, sentence completion, and short-answer questions, similar to reading as described previously. These items are intended to measure a variety of skills, including the following:

- Detailed understanding of specific points

- General understanding of the main point
- Understanding of directions and/or physical descriptions of places
- Ability to identify important points in a listening text
- Ability to discern specific pieces of information such as prices, places, or times

TOEFL iBT

The TOEFL iBT Listening section comprises five listening texts, three lectures, and two conversations representing social-interactive or navigational contexts. The lecturers frequently include one or more brief interactions, typically questions from the lecturer and answers from a student. Test takers are allowed to take notes on the listening content and use those notes to respond to questions, which are not displayed to test takers until after the listening passage. Most monologic passages are accompanied by context visuals (e.g., a photograph representing the lecturer) and frequently one or more content visuals, such as an unfamiliar word written on a board. There are 28 questions in total: six on each lecture and five on each conversation.

Content Comparison

Table 5 compares the listening sections for IELTS and TOEFL iBT.

Table 5. Comparison of IELTS and TOEFL iBT Listening Sections

Characteristic	IELTS	TOEFL iBT
Task description	Four passages with 10 questions each Uses a variety of question types	Test taker listens to a text and answers written comprehension questions
Number of tasks	4 (2 dialogues, 2 monologues)	5 (2 conversations, 3 lectures)
Number of items per task	10	5–6
Item type	Selected and constructed response	Selected-response comprehension questions (multiple choice)
Length of listening passage	Listening passages are approximately 269 words in length on average Passages about 2 minutes each	Conversations: 500–600 words; around 3 minutes Lectures: 700–800 words; around 5 minutes
Total time for section	30 minutes	36 minutes
Total listening time	12 minutes	16 minutes
Weighting	All items weighted equally	All items weighted equally

As discussed previously, the design of listening tests requires a number of considerations beyond those involved in reading tests. These can be divided into three categories: task variables, speaker variables, and linguistic demands of the input and output (Elliott & Wilson, 2013, p. 153). For the purposes of this report, I highlight some of these variables as they relate to IELTS and TOEFL iBT in Table 6. These include the inclusion of visual support, times heard, and timing of item presentation (i.e., task variables), gender and accents (i.e., speaker variables), and text domain/purpose (i.e., input variable). Among these variables, the main differences are the timing of items and the subdomains. On IELTS, the test questions are visible to test takers as they listen, albeit with some scrolling, giving them a target for directing their attention. On TOEFL iBT, test takers may take notes as they listen, but they do not see the test questions until after they have listened to the input. The topics and genres of the listening passages are also different, with the TOEFL iBT topics being more academic in nature.

Table 6. Comparison of Contextual Variables in Listening

Contextual variable	IELTS	TOEFL iBT
Transmission	Audio. Context and topic provided in instructions; some items have diagrams or charts to complete.	Audio. Some content and context visuals.
Times heard	Once	Once
Timing of item presentation	During listening	Post listening; notetaking allowed during listening
Gender of speakers	Mix of male and female voices	Mix of male and female voices
Accent of speakers	Range of accents: British, North American, Australian/New Zealand	Range of accents: North American, British, Australian/New Zealand
Text purpose/subdomain	Transactional/business Social-interpersonal Academic-navigational	Academic-content Academic-navigational Social-interpersonal

To provide a comparison of the linguistic demands of the listening passages, I selected the three longest monologues or monologue excerpts from the IELTS sample items, and the two lectures from the downloaded TOEFL exam described previously, removing any segments that involve interaction with another person, and analyzed them with Text Inspector as described in the Reading section. Selected results are presented in Table 7.

Table 7. Comparison of IELTS and TOEFL iBT Listening Metrics

Metric	IELTS 1 Life in UK	IELTS 2 Library Map	IELTS 3 Arts Center	TOEFL 1 Bat Acoustics	TOEFL 2 Habitat Selection
CEFR level	C1	B2+	C1+	C2	C1+
Words/sentence	18.79	25.4	22.67	27.67	18.95
Flesch-Kincaid reading grade	7.93	10.64	9.82	11.95	10.06
Academic word list (%)	7.33	2.42	5.12	6.67	8.77
B2 type (%)	2.72	3.73	5.09	9.66	10.05
C1 type (%)	0	0.62	0.54	3.86	3.31

The overall CEFR levels of these passages, based on a combination of metrics, show some overlap, although these specific TOEFL passages are a higher level on average: the IELTS monologues are rated between B2+ and C1+, while the TOEFL lectures come in at C1+ and C2. The reading grade levels for IELTS passages are also slightly lower on average than the TOEFL ones. One striking difference between the two tests is the vocabulary levels. The TOEFL passages include 6.67% to 8.77% academic words, while the range for these IELTS passages is from 2.42% to 7.33%. There are about twice as many B2 words in the TOEFL lectures than in the IELTS passages, and while C1 words are rare in the IELTS passages, they comprise 3% or more of the words in the TOEFL iBT lectures.

The difference in vocabulary between the two tests is mitigated somewhat by the fact that in the TOEFL iBT lectures, technical terms are typically defined and accompanied by a picture of the word. For example, in a lecture about bats, the term “echolocation” is introduced after one of the speakers mentions the concept. The professor introduces the term while a visual of the word is presented, after which the professor provides a definition (see Figure 4).

Figure 4. Excerpt from TOEFL iBT Lecture

- Student: Well, bats—since they’re all blind, bats have to use sound for—uh, y’know—to keep from flying into things.
- Professor: That’s echolocation.
- Professor: Using echoes—reflected sound waves—to locate things...



The difference in the vocabulary statistics between TOEFL iBT and IELTS reflects the fact that the IELTS listening section is used for both the General Training and Academic

versions of the test. Of the six sample listening scripts available for download from the IELTS website as of April 2025, none represent a lecture or other academic content, and only one can be construed as social/navigational: a conversation about applying to the Open University. The other listening texts are more representative of daily life tasks, such as mailing a package or finding out information at a tourist office. There are two short monologues represented in the sample items; one is a brief talk about staying in the UK, and the other is a tour of a library.

As discussed briefly, listening tests traditionally involve some reading, at least in terms of test questions to assess comprehension. Turning now to the test questions, I consider both the amount of reading involved and the subskills assessed.

Listening items on TOEFL iBT are predominantly multiple choice with four options; some questions ask for test takers to choose two options. The TOEFL iBT items tend to be formulated as questions or incomplete sentences and are often complex syntactically, as follows:

1. What is the main purpose of the lecture?
 - a. To compare active habitat selection with passive habitat selection
 - b. To show that most habitat preferences in animals are learned
 - c. To compare the habitat requirements of several bird species
 - d. To examine the consequences of habitat selection by animals

The listening items on IELTS involve much less reading, as they often involve filling in tables or charts rather than reading sentences. Multiple-choice questions, when they occur, have three options rather than four and are not always presented as complete sentences, as in Figure 5.

Figure 5. IELTS Listening Multiple-Choice Item

Choose the correct letter, A, B or C.

- 9** Type of insurance chosen
- | | |
|----------|----------|
| A | Economy |
| B | Standard |
| C | Premium |
- 10** Customer wants goods delivered to
- | | |
|----------|-------|
| A | port |
| B | home |
| C | depot |

Papageorgiou et al. (2021) provide a list of communication goals and foundational skills for assessing listening in academic settings. Using this framework, Table 8 summarizes the intended subskills assessed by the two tests as presented in their online information for test takers. Note that IELTS test items are categorized on the website by item type, with skills assessed by each item type listed beneath the item type description, while TOEFL iBT Listening items are categorized by skills targeted. The table lists the item types as they are labeled by each test provider.

Table 8. Comparison of Listening Items on IELTS and TOEFL iBT

Listening subskills	IELTS	TOEFL iBT
Communication goals		
Main ideas & supporting details	Multiple choice	Gist
	Matching	Detail
	Summary completion	
Relationships among ideas	Matching	Organization
	Sentence completion	Connecting content
	Plan/map/diagram labeling	
	Summary completion	
Inferences & opinions	N/A	Inference
Speaker purpose & attitude	N/A	Function, Attitude
Foundational skills		
Process extended speech in real time	All item types	All item types
Make use of phonological information	All item types	All item types
Make use of lexical and grammatical meaning	All item types	All item types
Make use of pragmatic information	N/A	Inference, Function, Attitude
Process organizational devices	All item types	Organization, Connecting content

Note. This table is based on a framework from Papageorgiou et al.'s (2021) work.

As the table shows, IELTS items tend to focus on ideas and relationships among ideas and not as much on inferences and opinions or speaker purpose and attitudes, whereas TOEFL iBT includes items that specifically assess these areas.

Discussion

As the tables show, both tests have dedicated listening sections that require test takers to listen to extended texts and answer questions targeted at understanding main ideas and details.

Both tests include both dialogues and monologues, and both include a variety of accents. Both tests require use of phonological, lexical, and grammatical information to process sustained aural input. In that sense, the two tests are quite comparable, and a concordance is warranted.

However, TOEFL iBT Listening is more representative of academic contexts, in that the listening texts are both more situationally authentic (e.g., TOEFL iBT allows note taking, an important academic skill) and use more academic language than IELTS listening. Furthermore, TOEFL iBT includes items specifically designed to assess pragmatic understanding in terms of making inferences about a speaker's opinions. Papageorgiou et al. (2021) argue that pragmatic understanding "might arguably be one of the key listening competences in higher education contexts where listening spans multiple registers and contexts, and where the ability to draw inferences about, for example, an interlocutor's stance, is of vital importance" (p. 73).

For this reason, test users should be cautious about interpreting IELTS Academic and TOEFL iBT Listening scores as being fully interchangeable, particularly for decisions about test takers that implicate these specific academic listening skills. As an example, Wagner (2016) found that both TOEFL iBT Speaking and Listening scores had moderate correlations with local tests for international teaching assistants (ITAs), but the listening test was a better predictor of ratings of teaching competence than the speaking score. It might not be wise to assume that IELTS Listening scores would have the same relationship with teaching ratings as TOEFL iBT Listening scores, given these differences.

Writing

A recent open letter to graduate students begins thusly: "Let me begin with a hard truth. As scientists, writing is our chief activity" (Hazelett, 2025, p. 447). This statement highlights the centrality of writing in academics, where the ability to write is critical to success at all levels of education and in academic careers. It is thus reasonable to expect a test of academic language to include a writing section in which test takers must "produce coherent, comprehensible texts" (Cumming et al., 2021, p. 108) rather than assessing discrete subskills such as error correction or reformulating sentences.

To produce a coherent text, Shaw and Weir (2007) describe six main cognitive processes for writing, as follows:

1. **Macro-planning**, which involves coming up with ideas and identifying the parameters of the writing task such as genre, purpose, and audience
2. **Organization**, which involves identifying relationships among ideas, ordering them appropriately, and prioritizing them in terms of their importance to the central thesis
3. **Micro-planning**, or planning out the language to be used to form sentences and paragraphs
4. **Translation** from abstract ideas into words, clauses and sentences
5. **Monitoring**, or evaluating the text for correctness, flow, and adherence to the writer's intentions
6. **Revising**, or making corrections to the text as a result of monitoring

This framework suggests that writing tasks should be complex enough and provide sufficient writing time for test takers to demonstrate their ability to plan, organize, draft, and revise their writing within the limitations of the testing period, and to be of a sufficient length that raters can reliably assess them on such aspects of writing as development of content, organization, and use of language.

A distinguishing characteristic of academic writing is that authentic writing tasks are nearly always written in response to source texts. As Cumming et al. (2021) stated, "Most students' writing for academic purposes involves them displaying (and ideally, also showing evidence of their transforming) their knowledge in direct relation to the content and contexts they have been studying, reading, hearing about, and discussing in academic courses" (p. 113).

Another important consideration in assessing writing is the number of writing tasks to include. There is always a trade-off between allowing sufficient time for candidates to write a fully developed essay (often 30 minutes or longer) and including multiple writing tasks so that raters' evaluations of writing is not based on a single written product (Weigle, 2002, p. 102).

Test Content

IELTS Academic

The IELTS Academic writing section comprises two tasks which need to be completed within 60 minutes. The first task requires test takers to describe visual information (typically a chart, diagram, or graph) in at least 150 words, with a 20-minute recommended time. The second

task requires test takers to discuss a point of view, argument, or problem in 250 words or more. Task 2 is weighted twice as heavily in scoring. The scoring criteria include task achievement (Task 1) or task response (Task 2), coherence and cohesion, lexical resource, and grammatical range and accuracy. IELTS writing is scored locally by a single examiner, and a sample of scripts is scored centrally by senior examiners to estimate rater reliability (Uysal, 2010). The IELTS Academic writing rubrics were recently revised based on examiner feedback, with minimal effects on scoring (see Clark et al., 2023, for details).

TOEFL iBT

TOEFL iBT also includes two writing tasks. The first task is a 20-minute Integrated Writing task, for which test takers read a short passage and listen to a brief lecture on the same topic. The writing task requires the test takers to summarize the lecture and relate it to points in the reading passage. There is no minimum length requirement, but effective responses are typically between 150 and 225 words, according to ETS’s published material (ETS, n.d.-d)

The writing section on TOEFL iBT test was revised in 2023, replacing a 30-minute independent essay task that was similar to Task 2 in IELTS Academic with a 10-minute Writing for an Academic Discussion (WAD) task. The WAD task is intended to replicate an online discussion, where a professor posts a question about a topic and one or two students have responded; the task requires the test taker to contribute to the discussion by posting their own opinion in 100 words or so. Both writing tasks are scored using holistic scales from 0–5. TOEFL iBT Writing is double scored, with one human rater and a proprietary automated scoring system based on the TOEFL iBT rubric, with discrepancies resolved by another human rater.

Content Comparison

The writing sections from the two tests are summarized in Table 9.

Table 9. Comparison of Writing on IELTS Academic and TOEFL iBT

Characteristic	IELTS Academic	TOEFL iBT
Number of tasks	2	2
Task description	<p>Task 1: Describe or explain information presented in a chart, graph, or table</p> <p>Task 2: Write an essay in response to a point of view, argument, or problem</p>	<p>Integrated Writing: Summarize information from reading and listening to short texts</p> <p>WAD: State and support an opinion in an online classroom discussion</p>

Characteristic	IELTS Academic	TOEFL iBT
Timing	60 minutes total (tasks not timed separately)	30 minutes total (tasks timed separately)
	Recommended: 20 minutes on Task 1, 40 minutes on Task 2	20 minutes for Integrated Writing, 10 minutes for WAD
Functions elicited	Describe visual information Support an opinion	Synthesize information from reading and listening Support an opinion
Text length of expected response	Task 1: At least 150 words; Task 2: at least 250 words	Integrated Writing: 150–225 words WAD: At least 100 words
Scoring	0–9 (analytic)	0–5 (holistic)

Scoring Criteria

The test construct is also reflected in the rating criteria used to score writing, as the characteristics of writing that are valued are those that are scored. In Table 10, the descriptors from the highest band for each test are presented side by side.

Table 10. Comparison of the Scoring Criteria for IELTS and TOEFL iBT (Highest Possible Score)

Criterion	IELTS Academic Band 9	TOEFL iBT Score 5
Task response	<p>T1: All the requirements of the task are fully and appropriately satisfied.</p> <p>There may be extremely rare lapses in content.</p> <p>T2: The prompt is appropriately addressed and explored in depth.</p> <p>A clear and fully developed position is presented which directly answers the question/s.</p> <p>Ideas are relevant, fully extended and well supported.</p> <p>Any lapses in content or support are extremely rare.</p>	<p>IW: A response at this level successfully selects the important information from the lecture and coherently and accurately presents this information in relation to the relevant information presented in the reading. The response is well organized, and occasional language errors that are present do not result in inaccurate or imprecise presentation of content or connections.</p> <p>WAD: The response is a relevant and very clearly expressed contribution to the online discussion, and it demonstrates consistent facility in the use of language.</p>
Coherence and cohesion/topic development	<p>The message can be followed effortlessly.</p> <p>Cohesion is used in such a way that it very rarely attracts attention.</p> <p>Any lapses in coherence or cohesion are minimal.</p> <p>Paragraphing is skillfully managed.</p>	<p>IW: The response is well organized, and occasional language errors that are present do not result in inaccurate or imprecise presentation of content or connections.</p> <p>WAD: Relevant and well-elaborated explanations, exemplifications, and/or details.</p>

Criterion	IELTS Academic Band 9	TOEFL iBT Score 5
Lexical resource	<p>T1: Full flexibility and precise use are evident within the scope of the task.</p> <p>T2: Full flexibility and precise use are widely evident.</p> <p>A wide range of vocabulary is used accurately and appropriately with very natural and sophisticated control of lexical features.</p> <p>Minor errors in spelling and word formation are extremely rare and have minimal impact on communication.</p>	<p>WAD: Precise, idiomatic word choice.</p> <p>Almost no lexical or grammatical errors other than those expected from a competent writer writing under timed conditions (e.g., common typos or common misspellings or substitutions like there/their).</p>
Grammatical range and accuracy	<p>A wide range of structures is used with full flexibility and control.</p> <p>Punctuation and grammar are used appropriately throughout.</p> <p>Minor errors are extremely rare and have minimal impact on communication.</p>	<p>IW: Occasional language errors that are present do not result in inaccurate or imprecise presentation of content or connections.</p> <p>WAD: [The response] demonstrates consistent facility in the use of language.</p> <p>Effective use of a variety of syntactic structures.</p> <p>Almost no lexical or grammatical errors other than those expected from a competent writer writing under timed conditions (e.g., common typos or common misspellings or substitutions like there/their).</p>

Note. T1 = Task 1; T2 = Task 2; IW = Integrated Writing task; WAD = Writing for an Academic Discussion task. Bold face represents relevant portions of a descriptor.

Discussion

Both tests include two tasks, at least one of which relies on source material, an important consideration for academic writing. Both tests also include a task that requires arguing a point of view. The essay task on IELTS Academic is longer than the WAD task on TOEFL iBT, both in suggested time (i.e., 40 versus 10 minutes) and in length (i.e., suggested length at least 250 versus 100 words), and thus provides the opportunity for test takers to develop an argument more thoroughly than does the WAD task in the TOEFL iBT test. On the other hand, the traditional argument essay as found in IELTS Academic and, previous to 2023, on TOEFL iBT has been criticized by writing scholars for promoting the mechanistic “five paragraph essay” (e.g., Kim, 2017). Davis and Norris (2023) argued that the WAD task simulates a common academic genre and also provides context that provides an audience and purpose for writing. Davis and Norris

also found that scores on the traditional Independent task were similar to those given to the WAD task, and that many of the linguistic features were similar across the two tasks.

Despite the differences in tasks and scoring processes, criteria for evaluating writing on the two tasks include very similar considerations, including a thorough, well-organized response to the demands of the task with minimal errors and flexible and precise use of linguistic resources. Both tests also include two different tasks that vary in their cognitive demands. The tasks of describing data, summarizing and reporting on data from sources, and expressing and supporting opinions with evidence are all relevant to academic writing. These considerations suggest that concordances across the writing scores are warranted.

Speaking

When asked to compare IELTS Academic and TOEFL iBT, many international students will say that speaking is the area that differs the most. Superficially, the main difference is that the IELTS speaking section consists of a one-on-one interaction with a trained examiner, while the TOEFL iBT Speaking section is administered on the computer and thus does not involve real-time interaction with another human being. To determine whether a concordance between the two tests is warranted, however, it is important to look beyond this difference at the actual requirements of the speaking tasks, the nature of the language elicited, and the criteria by which the speaking output is scored.

In a review of literature on academic speaking in university contexts, Xi et al. (2021) report that there is a wide range of expectations for speaking across different disciplines, with students in the humanities expected to participate more in class discussions and small group work than those in the natural and computational sciences. As discussed previously, the main subdomains of oral interaction for academic purposes include academic content, navigational, and social-interpersonal. Speaking events can be broadly divided into non-reciprocal (i.e., monologic) and reciprocal (i.e., dialogic or group interactions).

Test Content

IELTS

The IELTS speaking section is the same for both Academic and General Training tests. The three IELTS speaking tasks are embedded within a tightly scripted oral interview with a trained examiner. In the first task, the interviewer asks questions about familiar topics: in the

official practice materials provided by IELTS (IELTS, n.d.-c), the Task 1 topic is the test taker's hometown and living accommodations. In the second task, the interviewer hands the test taker a card with a topic to discuss for one to two minutes. The test taker is given one minute to prepare and may take notes. The example task is to describe something the test taker owns that is important to them. The third task is a two-way discussion that is related to the topic of Task 2. In the example provided in the sample IELTS materials, the interviewer asks questions about the types of things that provide status in the test taker's home country, whether this has changed over the generations, and the role of advertising in influencing purchases. As these examples show, the tasks are loosely related to each other. The interview as a whole is scored by the examiner on fluency and coherence, lexical resource, grammatical range and accuracy, and pronunciation on a 0–9 scale.

TOEFL iBT

The TOEFL iBT Speaking section consists of four tasks, one independent and three integrated tasks, which involve summarizing information from an input text such as a brief reading passage (i.e., less than 100 words) or a short excerpt from a lecture or conversation. The integrated tasks involve academic or campus-related topics. The independent task involves responding to a question about a personal opinion. All prompts are delivered by computer.

In the sample tasks available online as of April 2025 (TOEFL, n.d.-c), the independent task asks the test takers to express a preference for meeting up with friends in restaurants or cafés versus meeting in homes. In the reading/writing integrated task on a campus topic, the test taker reads a short passage about a plan to replace an old auditorium with a newer structure, and then listens to a short conversation about the plan, in which one speaker expresses an opinion. The test taker's task is to summarize the speaker's opinion. In the two more academic tasks, the test taker reads a short passage—typically the definition of an important concept in an academic field—and then listens to a lecture excerpt that expands on the reading. The final task consists of summarizing a brief lecture without any supporting reading material. Tasks are scored on a scale from 0–4, with slightly different versions of the scale used for the independent and the integrated tasks. The main factors in scoring include task completion, delivery, language use, and development.

Content Comparison

Table 11 provides an overview of the speaking sections for both tests.

Table 11. Comparison of Speaking on IELTS and TOEFL iBT

Characteristic	IELTS	TOEFL iBT
Number of tasks	3	4
Task description	All tasks are embedded within an oral interview with an examiner. Task 1: Introduction and interview (general questions on familiar topics). Task 2: Long turn (candidate is asked to speak for 1–2 minutes on a topic). Task 3: Discussion (elaborate on issues related to Task 2).	All tasks are computer-delivered: 1 is an independent/personal opinion task, and 3 are integrated, based on listening and/or reading
Functions elicited	Providing personal information; expressing and justifying opinions; explaining; suggesting; speculating; expressing preferences; comparing; summarizing; narrating	Expressing and justifying opinions; summarizing; synthesizing information from multiple sources
Timing	11–14 minutes total, including 1 minute of preparation time (Task 2)	Approximately 15 minutes
Scoring	0–9 (analytic)	0–4 (holistic)

Despite the superficial appearance of differences across the two tests, the speaking tasks are not dissimilar, in that the test taker is prompted to provide a sustained response to questions provided by either the computer or the examiner. Because of the intention to standardize the input test takers receive, the examiner’s script (referred to as the “interlocutor frame”) is fairly constrained, and the discussion does not seem to be intended to resemble normal conversation. For example, the transcript of the two-way discussion provided in the sample IELTS materials begins like this:

We’ve been talking about things we own. I’d like to discuss with you one or two more general questions relating to this topic. First, let’s consider values and the way they can change. In Switzerland, what kind of possessions do you think give status to people?

As this example shows, the examiner is in control of the interaction, providing a prompt for the test taker to respond to rather than initiating a genuine back and forth about a topic that is interesting to both parties. In this sense, the IELTS examiner fulfills the same function as the spoken instructions on the TOEFL iBT Speaking, which is to prompt the test taker to provide a speech sample that is coherent and long enough to be rated. One difference is that the examiner has the option to provide several follow-up questions when the test taker pauses, which is not the case in TOEFL iBT. On the other hand, the use of computer prompts on TOEFL iBT means that there is more consistency in test administration, potentially leading to higher reliability.

Scoring Criteria

As is the case with writing, IELTS speaking is scored locally by a human examiner, with quality control done centrally, while TOEFL iBT combines human and machine scoring for each test. Turning to the scoring criteria, Table 12 presents a side-by-side comparison of the descriptors at the highest level of each test in scoring.

Table 12. Comparison of the Scoring Criteria for TOEFL iBT and IELTS (Highest Possible Score)

Criterion	IELTS Band 9	TOEFL iBT Score 4 (Independent)	TOEFL iBT Score 4 (Integrated)
General description (TOEFL iBT only)	N/A	The response fulfills the demands of the task, with at most minor lapses in completeness. It is highly intelligible and exhibits sustained, coherent discourse. A response at this level is characterized by all of the following:	
Fluency and coherence	Fluent with only very occasional repetition or self-correction. Any hesitation that occurs is used only to prepare the content of the next utterance and not to find words or grammar. Speech is situationally appropriate and cohesive features are fully acceptable. Topic development is fully coherent and appropriately extended.	Generally well-paced flow (fluid expression). Speech is clear. Response is sustained and sufficient to the task. It is generally well developed and coherent; relationships between ideas are clear (or there is a clear progression of ideas).	Speech is generally clear, fluid and sustained.

Lexical resource	Total flexibility and precise use in all contexts. Sustained use of accurate and idiomatic language.	The response demonstrates effective use of grammar and vocabulary. It exhibits a fairly high degree of automaticity with good control of basic and complex structures (as appropriate). Some minor (or systematic) errors are noticeable but do not obscure meaning.	The response demonstrates good control of basic and complex grammatical structures that allow for coherent, efficient (automatic) expression of relevant ideas. Contains generally effective word choice. Though some minor (or systematic) errors or imprecise use may be noticeable, they do not require listener effort (or obscure meaning).
Grammatical range and accuracy	Structures are precise and accurate at all times, apart from “mistakes” characteristic of native speaker speech.		
Pronunciation	Uses a full range of phonological features to convey precise and/or subtle meaning. Flexible use of features of connected speech is sustained throughout. Can be effortlessly understood throughout. Accent has no effect on intelligibility.	It may include minor lapses, or minor difficulties with pronunciation or intonation patterns, which do not affect overall intelligibility.	It may include minor lapses or minor difficulties with pronunciation or intonation. Pace may vary at times as the speaker attempts to recall information. Overall intelligibility remains high.

Note that IELTS does not provide separate scores for each task within the interview, but scores for fluency and coherence, lexical resource, grammatical range and accuracy, and pronunciation. On TOEFL iBT, in contrast, each task is rated separately on a holistic scale: There is one scale for independent tasks and one for integrated tasks. The descriptors on the TOEFL scales are placed within the relevant IELTS categories in the table, again with some repetition when descriptors fit more than one category.

As with writing, the scoring criteria for both tests are similar, including such factors as overall success in fulfilling the task, fluent and coherent speech, appropriate use of vocabulary, and structure. This suggests that, despite differences in scoring procedures on the two tests, similar criteria are used to score the spoken output.

Discussion

Despite the differences in how the speaking sections of TOEFL iBT and IELTS are delivered, both tests provide opportunities for test takers to speak at length on a provided topic, with time to prepare, and similar criteria are used to evaluate the spoken output. The TOEFL

integrated speaking tasks may be considered somewhat more academic than the IELTS tasks, in that they require comprehension of extensive reading and/or listening input. On the other hand, IELTS speaking tasks, particularly the discussion task, involve more interaction with the examiner, and thus represent an aspect of speaking that can be considered important for academic success that is not covered by TOEFL iBT. However, the interactivity of the IELTS speaking test is limited by the fact that the examiner controls the discourse throughout the test and there are few or no opportunities for the test taker to demonstrate other important speaking functions such as introducing a topic or asking questions of the examiner.

In summary, both tests provide opportunities for students to provide extended speech samples that are evaluated similarly by raters. For this reason, a concordance is warranted with the caveat that the tests cover somewhat different aspects of the speaking construct, with TOEFL iBT slightly favoring presentational discourse and IELTS favoring interactional discourse.

Conclusion

In this report, I have examined the content of IELTS Academic and TOEFL iBT to determine whether their content is sufficiently similar to warrant a concordance study. Table 13 summarizes the main findings of this report to address the question of overall similarity.

Table 13. Summary Table of Content Comparison

Criterion	IELTS Academic	TOEFL iBT
Reading input	Extended, academic	Extended, academic
Listening input	Extended, general	Extended, academic
Reading test items	Variety of high-level reading skills	Variety of high-level reading skills
Listening items	Listen for gist and details	Listen for gist and details, and author's purpose
Speaking tasks	Elicit extended monologues and responses to follow-up questions with a variety of relevant functions	Elicit extended monologues with a variety of relevant functions; three tasks are related to academic content
Writing tasks	Elicit extended prose with relevant functions; integrated with reading/interpreting graph or chart	Elicit extended prose with relevant functions; integrated with listening and reading
Scoring criteria (speaking and writing)	Task fulfillment; organization/coherence; appropriate use of language	Task fulfillment; organization/coherence; appropriate use of language

The two tests are not identical by any means; some important differences are the format of test items for listening and reading, the test content for listening and speaking, and the delivery of the speaking test. However, there is no requirement that two tests cover identical constructs to justify a concordance; they simply need to be sufficiently similar.

As the table shows, IELTS Academic and TOEFL iBT have many commonalities, in many important ways. Both tests provide numerous opportunities for test takers to engage with and produce extended stretches of authentic discourse across all four skill areas and on a variety of topic areas. The degree of congruence differs somewhat across skills, with reading being the most congruent and listening perhaps the least. However, in my view, there is enough similarity to justify the use of concordance tables to compare scores across the two tests, with the caveat that scores should never be considered fully interchangeable.

Ikeda et al. (2025) reminded us that

Score concordance tables are a useful instrument for score users, for example, institutions who use certain test scores for decisions about test takers, and who need to set comparable score requirements across different tests. Score users are advised that score comparisons across tests, while based on empirical research, are estimates only and should be treated with caution. (p. 18)

It is my hope that this report provides insights into the TOEFL iBT and IELTS Academic test that will be helpful to test users in determining how to interpret scores from both tests.

References

- Capel, A. (2015). The English vocabulary profile. In J. Harrison & F. Barker (Eds.), *English Profile in Practice* (pp. 9–27). UCLES/Cambridge University Press.
- Chapelle, C. A. (2008). The TOEFL® validity argument. In C. A. Chapelle, M. K. Enright, & J. M. Jamieson (Eds.), *Building a validity argument for the Test of English as a Foreign Language* (pp. 319–352). Routledge. <https://doi.org/10.4324/9780203937891>
- Chapelle, C. A., & Lee, H. W. (2021). Conceptions of validity. In G. Fulcher & L. Harding (Eds.), *The Routledge handbook of language testing* (pp. 17–31). Routledge. <https://doi.org/10.4324/9781003220756-3>
- Clark, T., Tasviri, R., & Bruce, E. (2023). *IELTS writing scales review and update—summary overview*. IELTS. <https://ielts.org/researchers/our-research/research-reports/ielts-writing-scales-review-and-update-summary-overview>
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213–238. <https://doi.org/10.2307/3587951>
- Cumming, A., Cho, Y., Burstein, J., Everson, P., & Kantor, R. (2021). Assessing academic writing. In X. Xi & J. M. Norris (Eds.), *Assessing academic English for higher education admissions* (pp. 107–151). Routledge. <https://doi.org/10.4324/9781351142403>
- Cushing, S. T. (2025). *Testing academic language proficiency: Comparing the TOEFL iBT® test with the Duolingo English Test* (TOEFL Research Report No. RR-104). ETS. <https://doi.org/10.64634/0437kw56>
- Elliott, M., & Wilson, J. (2013). Context validity. In A. Geranpayeh & L. Taylor (Eds.), *Examining listening: Research and practice in assessing second language listening* (pp. 152–241). UCLES/Cambridge University Press.
- Field, J. (2013). Cognitive validity. In A. Geranpayeh & L. Taylor (Eds.), *Examining listening: Research and practice in assessing second language listening* (pp. 267–290). UCLES/Cambridge University Press.
- Geranpayeh, A. (1994). Are score comparisons across language proficiency test batteries justified? An IELTS-TOEFL comparability study. *Edinburgh Working Papers in Applied Linguistics*, 5, 50–65.
- Hazelett, D. J. (2025). An open letter to graduate students and other procrastinators: It's time to write. *Nature Biotechnology*, 43, 447–450. <https://doi.org/10.1038/s41587-025-02584-1>

- IELTS. (n.d.-a). *The IELTS guide for test takers*. <https://ielts.org/cdn/ielts-guides/ielts-guide-for-test-takers.pdf>
- IELTS. (n.d.-b). *IELTS Academic format: Reading*. <https://ielts.org/take-a-test/test-types/ielts-academic-test/ielts-academic-format-reading>
- IELTS. (n.d.-c). *IELTS Academic test—sample test questions*. <https://ielts.org/take-a-test/preparation-resources/sample-test-questions/academic-test>
- IELTS. (n.d.-d). *IELTS Academic format: Listening*. <https://ielts.org/take-a-test/test-types/ielts-academic-test/ielts-academic-format-listening>
- Ikeda, N., Clark, T., Papageorgiou, S., Gu, L., Ohta, R., Blackhurst, A., & Bruce, E. (2025). *Aligning scores of language proficiency tests: A score concordance study between IELTS Academic and TOEFL iBT®* (TOEFL Research Report No. RR-105). ETS. <https://doi.org/10.64634/tn27a897>
- Janusik, L. A., & Wolvin, A. D. (2009). 24 hours in a day: A listening update to the time studies. *International Journal of Listening*, 23(2), 104–120. <https://doi.org/10.1080/10904010903014442>
- Kane, M. T. (2004). Certification testing as an illustration of argument-based validation. *Measurement: Interdisciplinary Research and Perspectives*, 2(3), 135–170. https://doi.org/10.1207/s15366359mea0203_1
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1–73. <https://doi.org/10.1111/jedm.12000>
- Knoch, U., & Fan, J. (2024). Test score comparison tables: How well are they serving test users? *Language Testing*, 41(3), 681–693. <https://doi.org/10.1177/0265532224123>
- Lam, D. M. K. (2021). Don't turn a deaf ear: A case for assessing interactive listening. *Applied Linguistics*, 42(4), 740–764. <https://doi.org/10.1093/applin/amaa064>
- Liu, X., & Read, J. (2023). Designing a new diagnostic reading assessment for a local post-admission assessment program: A needs-driven approach. In X. Yan, S. Dimova, & A. Ginther (Eds.), *Local language testing: Practice across contexts* (pp. 83–102). Springer Cham. https://doi.org/10.1007/978-3-031-33541-9_5
- Lynch, T. (2011). Academic listening in the 21st century: Reviewing a decade of research. *Journal of English for Academic Purposes*, 10(2), 79–88. <https://doi.org/10.1016/j.jeap.2011.03.001>

- Manna, V. F., Li, S., Papageorgiou, S., & Gu, L. (2025). *TOEFL iBT® technical manual* (TOEFL Research Report No. RR-106). ETS. <https://doi.org/10.64634/eje8f497>
- Norris, J. M., Davis, J. M., & Xi, X. (2021). Framing the assessment of academic English for admissions decisions. In X. Xi & J. M. Norris (Eds.), *Assessing academic English for higher education admissions* (pp. 1–21). Routledge.
- Papageorgiou, S., Schmidgall, J., Harding, L., Nissan, S., & French, R. (2021). Assessing academic listening. In X. Xi & J. M. Norris (Eds.), *Assessing academic English for higher education admissions* (pp. 61–106). Routledge. <https://doi.org/10.4324/9781351142403-3>
- Ren, H., & Cushing, S. (2022). *Comparison of IELTS Academic and Duolingo English Test*. Cambridge Assessment. <https://ielts.org/researchers/our-research/research-reports/comparison-of-ielts-academic-and-duolingo-english-test>
- Rodgers, M. P., & Webb, S. (2016). Listening to lectures. In K. Hyland (Ed.), *The Routledge handbook of English for academic purposes* (pp. 165–176). Routledge. <https://doi.org/10.4324/9781315657455>
- Schedl, M., O'Reilly, T., Grabe, W., & Schoonen, R. (2021). Assessing academic reading. In X. Xi & J. M. Norris (Eds.), *Assessing academic English for higher education admissions* (pp. 22–60). Routledge. <https://doi.org/10.4324/9781351142403-2>
- Shaw, S. D., & Weir, C. J. (2007). *Examining writing: Research and practice in assessing second language writing*. UCLES/Cambridge University Press.
- Taylor, L. (2013). Introduction. In A. Geranpayeh & L. Taylor (Eds.), *Examining listening: Research and practice in assessing second language listening* (pp. 1–35). UCLES/Cambridge University Press.
- TOEFL. (n.d.-a). *TOEFL IBT® Listening section*. ETS. <https://www.ets.org/toefl/test-takers/ibt/about/content/listening.html>
- TOEFL. (n.d.-b). *TOEFL IBT® Reading section*. ETS. <https://www.ets.org/toefl/test-takers/ibt/about/content/reading.html>
- TOEFL. (n.d.-c). *TOEFL IBT® Speaking section*. ETS. <https://www.ets.org/toefl/test-takers/ibt/about/content/speaking.html>
- TOEFL. (n.d.-d). *TOEFL IBT® Writing section*. ETS. <https://www.ets.org/toefl/test-takers/ibt/about/content/writing.html>

- TOEFL. (2025). *TOEFL® research insight series: Volume 4. Validity evidence supporting the interpretation and use of TOEFL iBT® scores*. <https://www.ets.org/pdfs/toefl/toefl-ibt-insight-s1v4.pdf>
- Uysal, H. H. (2010). A critical review of the IELTS writing test. *ELT Journal*, 64(3), 314–320. <https://doi.org/10.1093/elt/ccp026>
- Wagner, E. (2016). *A study of the use of TOEFL iBT® Speaking and Listening scores for international teaching assistant screening* (TOEFL iBT Research Report No. 27). <http://dx.doi.org/10.1002/ets2.12104>
- Wagner, E. (2022). L2 listening comprehension: Theory and research. In E. H. Jeon & Y. In'nami (Eds.), *Bilingual processing and acquisition* (pp. 213–233). John Benjamins.
- Weigle, S. C. (2002). *Assessing writing*. Cambridge University Press.

Notes

¹ Note that this description was accurate as of the date of the concordance study (2024) and does not incorporate changes to the test since that time (see Manna et al., 2025).

² TOEFL iBT offers a subscription-based sample online test through TestReady: <https://www.ets.org/toefl/test-takers/ibt/prepare/toefl-testready.html>

Suggested Citation

Cushing, S. T. (2026). *Beyond score correlations: A content comparison of IELTS Academic and TOEFL iBT® tests in the context of a score concordance study* (TOEFL Research Report No. RR-107). ETS. <https://doi.org/10.64634/qercg225>

This report is also available from IELTS.

Funding

This research was funded by the Cambridge University Press & Assessment on behalf of the IELTS partners, as well as Educational Testing Service (ETS), owner of the TOEFL® test.

Action Editor: Spiros Papageorgiou

Reviewers: Ching-Ni Hsieh and Saerhim Oh

ETS, the ETS logo, TOEFL, and TOEFL IBT are registered trademarks of Educational Testing Service (ETS). All other trademarks are property of their respective owners.

Cover image by Kampus Production, Pexels

Find other ETS-published reports by searching the [ETS ReSEARCHER](#) database.



www.ets.org