



Pearson
PTE Academic

Score Guide for test takers

Version 15 – April 2021

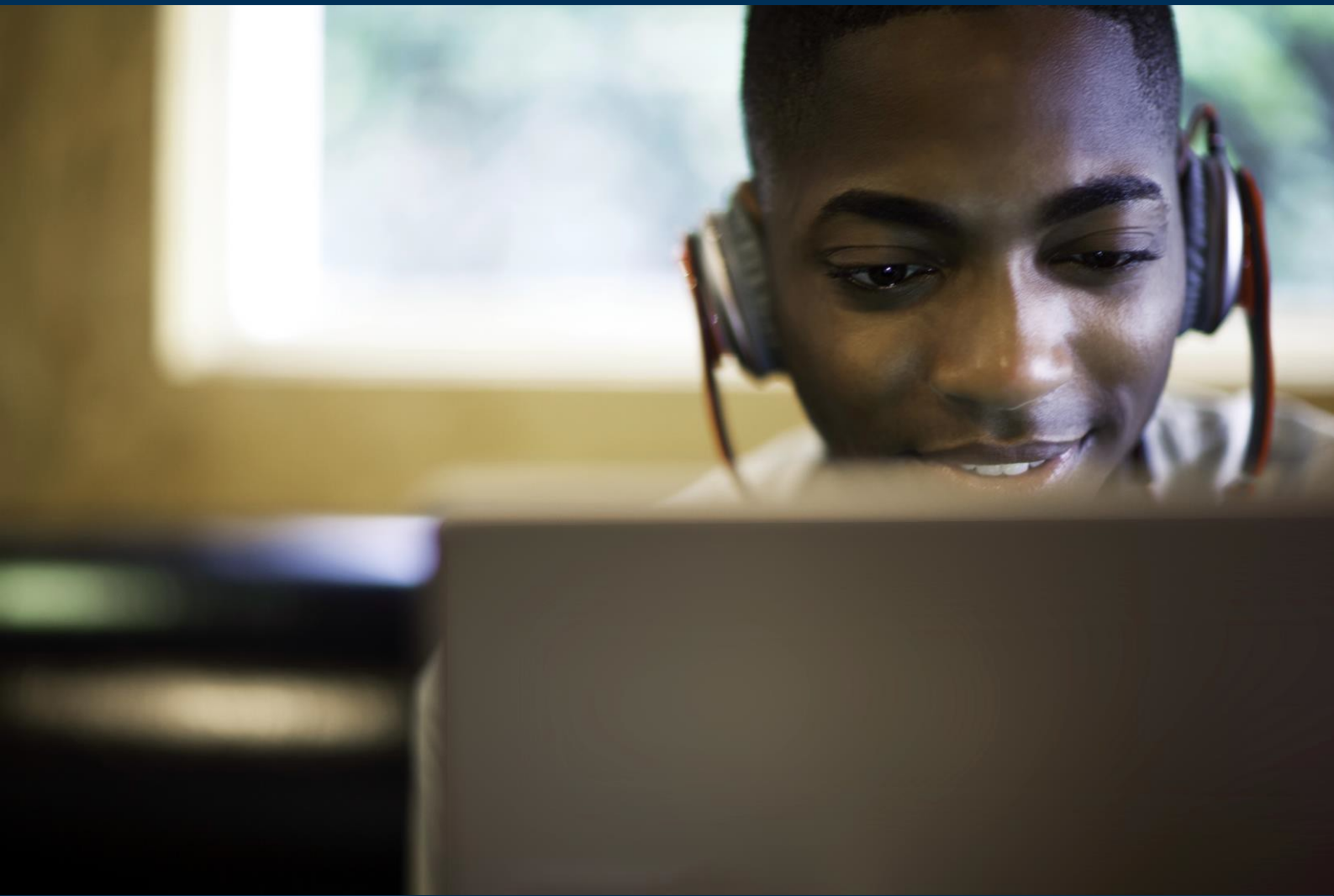


Table of Contents

Introduction	4
1. Reported Scores: An Overview	5
<i>Overall score</i>	6
<i>Communicative skills scores</i>	6
<i>Enabling skills scores</i>	7
2. Item Scoring: An Overview	9
<i>Correct or incorrect</i>	9
<i>Partial credit</i>	9
<i>How all 20 item types are scored</i>	9
Scoring criteria: Pronunciation and Oral fluency.....	17
3. Using PTE Academic Scores	20
<i>How test takers can use PTE Academic scores</i>	20
Overall score and communicative skills scores.....	20
Sending scores to governments or institutions.....	21
4. Estimates of concordance between PTE Academic, CEFR, TOEFL iBT and IELTS	22
<i>The PTE Academic Score Scale and the CEFR</i>	23
<i>What PTE Academic scores mean</i>	25
Estimate of concordance between PTE Academic scores and TOEFL iBT.....	26
Estimate of concordance between PTE Academic and IELTS.....	26
5. Scored Samples	27
<i>Automated scoring</i>	27
Scoring written English skills.....	27
Scoring spoken English skills.....	28
Enabling skill scores and trait scores examples.....	29
<i>Spoken samples</i>	30
Example <i>Describe image</i> item.....	30
Test Taker responses.....	33
<i>Overall performance rating</i>	35
<i>Written samples</i>	36
Example <i>Write essay</i> item 'Tobacco'.....	36
Test Taker Responses.....	39
<i>Overall performance rating</i>	42
6. Glossary	43
Appendix	45

Introduction

Pearson Test of English Academic (PTE Academic) is an international computer-based English language test. It provides a measure of a test taker's language ability to assist education institutions and professional and government organizations that require a standard of academic English language proficiency for admission purposes.

The *Score Guide* is designed for test takers who want to learn more about how the different tasks in PTE Academic are scored. The Guide will help you to understand:

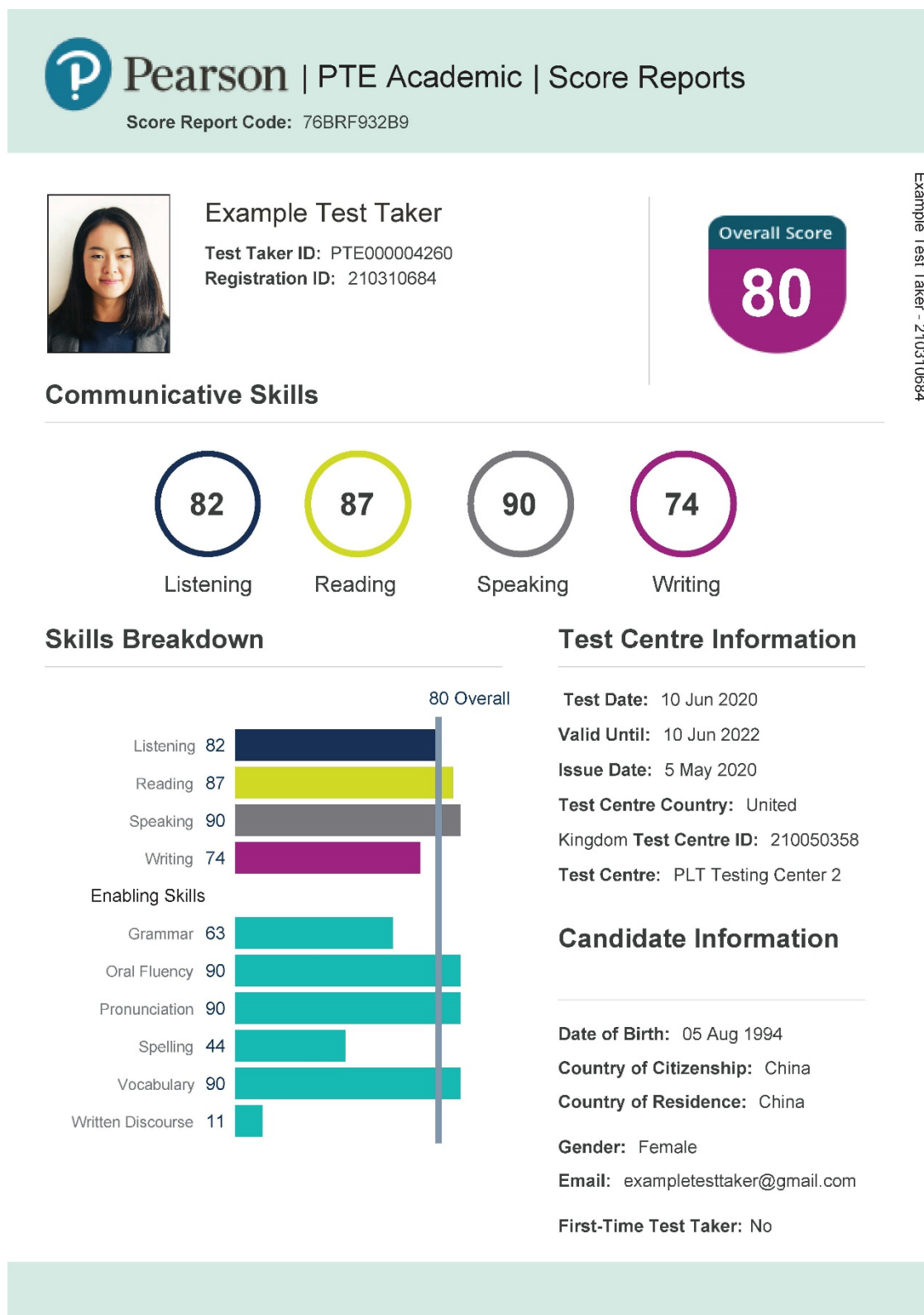
- What test takers are assessed on
- How to use scores reported on the score report
- How to compare PTE Academic scores with scores on other English language tests
- How automated scoring operates

The Guide has been bookmarked and linked so that you can access sections quickly from the 'Contents' page and dip into the topics you want to know more about.

The contents of this Guide, along with those published on our [website](#), are the only official and valid information about PTE Academic.

1. Reported Scores: An Overview

PTE Academic reports an overall score, communicative skills scores and enabling skills scores. Below is a picture of a full Test Taker Score Report. On the next page, there is a picture showing the score details, with an explanation of what they mean.



Overall score

The overall score is based on performance on all test items (tasks in the test consisting of instructions, questions or prompts, answer opportunities and scoring rules). Each test taker will answer between 71 and 83 items in any given test and there are 20 different item types. For each item, the score given contributes to the overall score. The score range is 10–90 points.

Note: the overall score is not an average calculation of the communicative scores.

Communicative skills scores

The communicative skills measured are **listening**, **reading**, **speaking** and **writing**. These are the scores you will be judged on for visa/HEI entry. Items testing these communicative skills also test specific subskills. For items requiring integrated skills (those assessing reading and speaking, listening and speaking, reading and writing, listening and writing, or listening and reading) the item score contributes to the score for the communicative skills that the item assesses. The score range for each skill is 10–90 points.

Enabling skills scores

Enabling skills are reported so that test takers can understand areas of strength and areas for improvement in their language ability. The enabling skills are **grammar, oral fluency, pronunciation, spelling, vocabulary, and written discourse**. Enabling skill scores are summaries of a test taker's performance on different aspects of language that are scored in different item types.

Some speaking and writing items are scored against criteria relating to traits in the spoken or written responses. Traits are characteristics of the response relating to features of language ability. The traits listed in the table below contribute to enabling skills scores, as well as to overall and communicative skills scores.

Enabling skill	Enabling skill definition	Traits used in scoring
Oral fluency	Smooth, effortless and natural-paced delivery of speech	Oral fluency
Pronunciation	Ability to produce speech sounds in a way that is easily understandable to most regular speakers of the language. Regional or national pronunciation variants are considered correct to the degree that they are understandable to most regular speakers of the language.	Pronunciation
Written discourse	Correct and communicatively efficient production of written language at the textual level. Written discourse skills are manifested in the structure of a written text, its internal coherence, logical development, and the range of linguistic resources used to express meaning precisely.	Development, structure and coherence AND General linguistic range
Vocabulary	Appropriate choice of words used to express meaning precisely in written and spoken English, as well as lexical range	Vocabulary
Grammar	Correct use of language with respect to word form and word order at the sentence level	Grammar
Spelling	Writing of words according to the spelling rules of the language. All national variations in spelling are considered correct.	Spelling

Some speaking and writing items are also scored for Content. Content scores identify how appropriately the content of a spoken or written response addresses the item prompt. Some writing items are scored for Form. Form scores are based on formal characteristics of the response such as the number of words. Content and Form scores contribute to overall and communicative skills scores, but do not contribute to any enabling skills scores.

For some speaking and listening items, when the content of a response to an item is scored as not dealing appropriately with an item prompt, or if formal requirements of the response are not met, no score points for the response will be given. Here are two examples giving descriptions of essay responses that will not receive any score points:

1. An essay which addresses a completely different topic from the item prompt (content)
2. An essay which is less than 120 words (form)

Please see Section 2: Item Scoring for further information on how items are scored.

2. Item Scoring: An Overview

All items in PTE Academic are machine scored. Scores for some item types are based on correctness alone, while others are based on correctness, formal aspects and the quality of the response.

Formal aspects refer to the form of the response: for example, whether it is over or under the word limit for a particular item type.

The quality of the response is represented in the enabling skills. For example, in the item type *Re-tell lecture* the response is scored on enabling skills such as oral fluency and pronunciation.

Scores for item types assessing speaking and writing skills are generated by automated scoring systems. There are two types of scoring:

Correct or incorrect

Some item types are scored as either correct or incorrect. If responses are correct, 1 score point will be given, but if they are incorrect, no score points are awarded.

Partial credit

Other item types are scored as correct, partially correct or incorrect. If responses to these items are correct, the maximum score points available for each item type will be received. If they are partly correct, some score points will be given, but less than the maximum available for the item type. If responses are incorrect, no score points will be received.

How all 20 item types are scored

The tables that follow give an overview of how the 20 item types in the three parts of PTE Academic (Part 1 Speaking and Writing, Part 2 Reading, Part 3 Listening) are scored. They also show timings, the number of items in any given test, and the communicative skills, enabling skills and other elements scored.

Part 1 Speaking and Writing (approx. 72–93 minutes)					
Item type	Time allowed	Number of items	Scoring	Communicative skills scored	Traits scored
Read aloud	30-35 minutes	6-7	Partial credit	Reading and speaking	Content, pronunciation, oral fluency Content: Each replacement, omission or insertion of a word counts as one error. Maximum score: depends on the length of the item prompt
					Pronunciation: 5 Native-like 4 Advanced 3 Good 2 Intermediate 1 Intrusive 0 Non-English (See detailed criteria on page 20.)
					Oral fluency: 5 Native-like 4 Advanced 3 Good 2 Intermediate 1 Limited 0 Disfluent (See detailed criteria on page 20.)

Part 1 Speaking and Writing (approx. 72–93 minutes)

Item type	Time allowed	Number of items	Scoring	Communicative skills scored	Traits scored
Repeat sentence		10-12	Partial credit	Listening and speaking	<p>Content, pronunciation, oral fluency</p> <p>Content: Errors = replacements, omissions and insertions only Hesitations, filled or unfilled pauses, leading or trailing material are ignored in the scoring of content 3 All words in the response from the prompt in the correct sequence 2 At least 50% of words in the response from the prompt in the correct sequence 1 Less than 50% of words in the response from the prompt in the correct sequence 0 Almost nothing from the prompt in the response</p> <hr/> <p>Pronunciation: 5 Native-like (See detailed criteria on page 20.) 4 Advanced 3 Good 2 Intermediate 1 Intrusive 0 Non-English</p> <hr/> <p>Oral fluency: 5 Native-like (See detailed criteria on page 20.) 4 Advanced 3 Good 2 Intermediate 1 Limited 0 Disfluent</p>

Part 1 Speaking and Writing (approx. 72–93 minutes)

Item type	Time allowed	Number of items	Scoring	Communicative skills scored	Traits scored
Describe image		6-7	Partial credit	Speaking	<p>Content, pronunciation, oral fluency</p> <p>Content:</p> <p>5 Describes all elements of the image and their relationships, possible development and conclusion or implications</p> <p>4 Describes all the key elements of the image and their relations, referring to their implications or conclusions</p> <p>3 Deals with most key elements of the image and refers to their implications or conclusions</p> <p>2 Deals with only one key element in the image and refers to an implication or conclusion. Shows basic understanding of several core elements of the image</p> <p>1 Describes some basic elements of the image, but does not make clear their interrelations or implications</p> <p>0 Mentions some disjointed elements of the presentation</p>
					<p>Pronunciation:</p> <p>5 Native-like</p> <p>4 Advanced (See detailed criteria on page 20.)</p> <p>3 Good</p> <p>2 Intermediate</p> <p>1 Intrusive</p> <p>0 Non-English</p>
					<p>Oral fluency:</p> <p>5 Native-like (See detailed criteria on page 20.)</p> <p>4 Advanced</p> <p>3 Good</p> <p>2 Intermediate</p> <p>1 Limited</p> <p>0 Disfluent</p>

Part 1 Speaking and Writing (approx. 72–93 minutes)

Item type	Time allowed	Number of items	Scoring	Communicative skills scored	Traits scored
<p>Re-tell lecture</p>		<p>3-4</p>	<p>Partial credit</p>	<p>Listening and speaking</p>	<p>Pronunciation, oral fluency</p> <p>Pronunciation:</p> <p>5 Native-like</p> <p>4 Advanced</p> <p>3 Good</p> <p>2 Intermediate</p> <p>1 Intrusive</p> <p>0 Non-English</p> <p>(See detailed criteria on page 20.)</p>
					<p>Oral fluency:</p> <p>5 Native-like</p> <p>4 Advanced</p> <p>3 Good</p> <p>2 Intermediate</p> <p>1 Limited</p> <p>0 Disfluent</p> <p>(See detailed criteria on page 20.)</p>
					<p>Content:</p> <p>5 Re-tells all points of the presentation and describes characters, aspects and actions, their relationships, the underlying development, implications and conclusions</p> <p>4 Describes all key points of the presentation and their relations, referring to their implications and conclusions</p> <p>3 Deals with most points in the presentation and refers to their implications and conclusions</p>

					<p>2 Deals with only one key point and refers to an implication or conclusion. Shows basic understanding of several core elements of the presentation</p> <p>1 Describes some basic elements of the presentation but does not make clear their interrelations or implications</p> <p>0 Mentions some disjointed elements of the presentation</p>
Part 1 Speaking and Writing (approx. 72–93 minutes)					
Item type	Time allowed	Number of items	Scoring	Communicative skills scored	Traits scored
Answer short question		10-12	Correct/incorrect	Listening and speaking	<p>Vocabulary</p> <p>1 Appropriate word choice in response</p> <p>0 Inappropriate word choice in response</p>
Summarize written text	20-30 minutes	2-3	Partial credit	Reading and writing	<p>Form, grammar, vocabulary</p> <p>Form:</p> <p>1 Is written in one, single, complete sentence</p> <p>0 Not written in one single, complete sentence or contains fewer than 5 or more than 75 words. Summary is written in capital letters.</p>
					<p>Content:</p> <p>2 Provides a good summary of the text. All relevant aspects mentioned</p> <p>1 Provides a fair summary of the text but misses one or two aspects</p> <p>0 Omits or misrepresents the main aspects of the text</p>
					<p>Grammar:</p> <p>2 Has correct grammatical structure</p>

					<p>1 Contains grammatical errors but with no hindrance to communication</p> <p>0 Has defective grammatical structure which could hinder communication</p>
					<p>Vocabulary:</p> <p>2 Has appropriate choice of words</p> <p>1 Contains lexical errors but with no hindrances to communication</p> <p>0 Has defective word choice which could hinder communication</p>

Part 1 Speaking and Writing (approx. 72–93 minutes)					
Item type	Time allowed	Number of items	Scoring	Communicative skills scored	Traits scored
Write essay	20-40 minutes	1-2	Partial credit	Writing	Content; form; development, structure, coherence; grammar; general linguistic range; vocabulary range; spelling Content: 3 Adequately deals with the prompt 2 Deals with the prompt but does not deal with one minor aspect 1 Deals with the prompt but omits a major aspect or more than one minor aspect 0 Does not deal properly with the prompt
					Form: 2 Length is between 200 and 300 words 1 Length is between 120 and 199 or between 301 and 380 words 0 Length is less than 120 or more than 380 words. Essay is written in capital letters, contains no punctuation or only consists of bullet points or very short sentences
					Development, structure and coherence: 2 Shows good development and logical structure 1 Is incidentally less well structured, and some elements or paragraphs are poorly linked 0 Lacks coherence and mainly consists of lists or loose elements

Part 1 Speaking and Writing (approx. 72–93 minutes)

Item type	Time allowed	Number of items	Scoring	Communicative skills scored	Traits scored
Write essay					<p>Grammar: 2 Shows consistent grammatical control of complex language. Errors are rare and difficult to spot. 1 Shows a relatively high degree of grammatical control. No mistakes which would lead to misunderstandings 0 Contains mainly simple structures and/or several basic mistakes</p> <p>General linguistic range: 2 Exhibits smooth mastery of a wide range of language to formulate thoughts precisely, give emphasis, differentiate and eliminate ambiguity. No sign that the test taker is restricted in what they want to communicate 1 Sufficient range of language to provide clear descriptions, express viewpoints and develop arguments 0 Contains mainly basic language and lacks precision</p> <p>Vocabulary range: 2 Good command of a broad lexical repertoire, idiomatic expressions and colloquialisms 1 Shows a good range of vocabulary for matters connected to general academic topics. Lexical shortcomings lead to circumlocution or some imprecision 0 Contains mainly basic vocabulary insufficient to deal with the topic at the required level</p> <p>Spelling: 2 Correct spelling 1 One spelling error 0 More than one spelling error</p>

Part 2 Reading (approximately 31–41 minutes)					
Item type	Time allowed	Number of items	Scoring	Communicative skills scored	Traits scored
Reading and writing: Fill in the blanks	32-41 minutes	5-6	Partial credit (for each correctly completed blank)	Reading and writing 1 Each correctly completed blank 0 Minimum score	
Multiple-choice, choose multiple answers		2-3	Partial credit (for each correct response. Points deducted for incorrect options chosen)	Reading 1 Each correct response - 1 Each incorrect response 0 Minimum score	
Re-order paragraphs		2-3	Partial credit (for each correctly ordered, adjacent pair)	Reading 1 Each pair of correct adjacent textboxes 0 Minimum score	
Reading: Fill in the blanks		4-5	Partial credit (for each correctly completed blank)	Reading 1 Each correctly completed blank 0 Minimum score	
Multiple-choice, choose single answer		2-3	Correct/incorrect	Reading 1 Each correctly completed blank 0 Minimum score	

Part 3 Listening (approx. 44–61 minutes)					
Item type	Time allowed	Number of items	Scoring	Communicative skills scored	Traits scored
Summarize spoken text	20-30 minutes	2-3	Partial credit	Listening and writing	<p>Content, form, grammar, vocabulary, spelling</p> <p>Content:</p> <p>2 Provides a good summary of the text. All relevant aspects are mentioned.</p> <p>1 Provides a fair summary of the text, but one or two aspects are missing</p> <p>0 Omits or misrepresents the main aspects</p>
					<p>Form:</p> <p>2 Contains 50-70 words</p> <p>1 Contains 40-49 words or 71-100 words</p> <p>0 Contains less than 40 words or more than 100 words. Summary is written in capital letters, contains no punctuation or consists only of bullet points or very short sentences.</p>
					<p>Grammar:</p> <p>2 Correct grammatical structures</p> <p>1 Contains grammatical errors with no hindrance to communication</p> <p>0 Defective grammatical structure which could hinder communications</p>
					<p>Vocabulary:</p> <p>2 Appropriate choice of words</p> <p>1 Some lexical errors but with no hindrance to communication</p> <p>0 Defective word choice which could hinder communication</p>
					<p>Spelling:</p> <p>2 Correct spelling</p> <p>1 One spelling error</p> <p>0 More than one spelling errors</p>

Part 3 Listening (approx. 44–61 minutes)					
Item type	Time allowed	Number of items	Scoring	Communicative skills scored	Traits scored
Multiple-choice, choose multiple answers	23-28 minutes	2-3	Partial credit (for each correct response. Points deducted for incorrect options chosen)	Listening 1 Each correct response - 1 Each incorrect response 0 Minimum score	
Fill in the blanks		2-3	Partial credit (for each correct word spelled correctly)	Listening and writing 1 Each correct word spelled correctly 0 Minimum score	
Highlight correct summary		2-3	Correct/incorrect	Listening and reading 1 Correct response 0 Incorrect response	
Multiple-choice, choose single answer		2-3	Correct/incorrect	Listening 1 Correct response 0 Incorrect response	
Select missing word		2-3	Correct/incorrect	Listening 1 Correct response 0 Incorrect response	
Highlight incorrect words		2-3	Partial credit (For each word. Points deducted for incorrect options chosen)	Listening and reading 1 Each correct word - 1 Each incorrect word 0 Minimum score	
Write from dictation		3-4	Partial credit (for each word spelled correctly)	Listening and writing 1 Each correct word spelled correctly 0 Each incorrect or misspelled word	

Scoring criteria: Pronunciation and Oral fluency

The following scoring criteria apply to the speaking item types that are scored on pronunciation and oral fluency in PTE Academic.

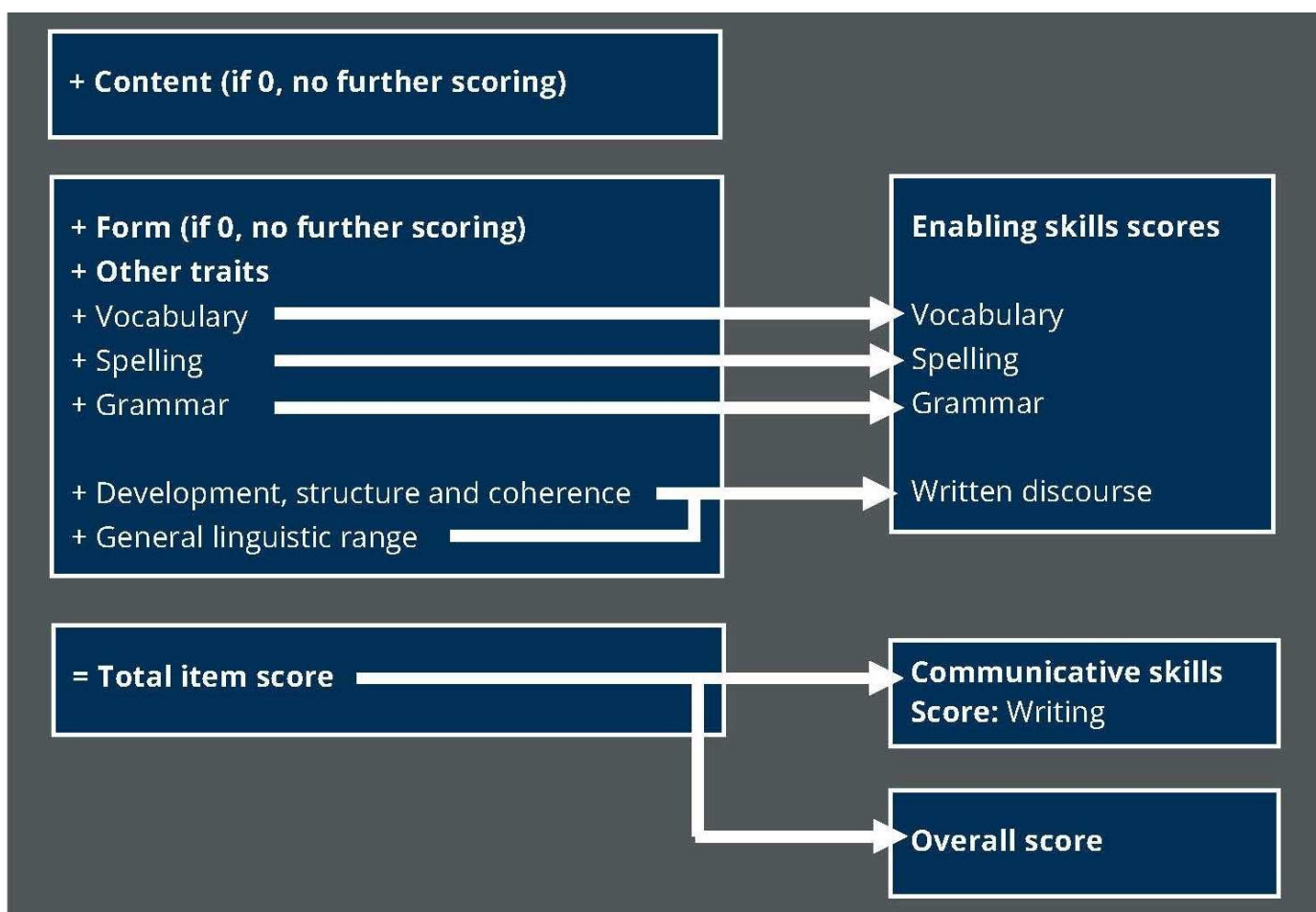
Pronunciation	
5 Native-like	All vowels and consonants are produced in a manner that is easily understood by regular speakers of the language. The speaker uses assimilation and deletions appropriate to continuous speech. Stress is placed correctly in all words and sentence-level stress is fully appropriate.
4 Advanced	Vowels and consonants are pronounced clearly and unambiguously. A few minor consonant, vowel or stress distortions do not affect intelligibility. All words are easily understandable. A few consonants or consonant sequences may be distorted. Stress is placed correctly on all common words, and sentence level stress is reasonable.
3 Good	Most vowels and consonants are pronounced correctly. Some consistent errors might make a few words unclear. A few consonants in certain contexts may be regularly distorted, omitted or mispronounced. Stress-dependent vowel reduction may occur on a few words.
2 Intermediate	Some consonants and vowels are consistently mispronounced in a non-native like manner. At least 2/3 of speech is intelligible, but listeners might need to adjust to the accent. Some consonants are regularly omitted, and consonant sequences may be simplified. Stress may be placed incorrectly on some words or be unclear.
1 Intrusive	Many consonants and vowels are mispronounced, resulting in a strong intrusive foreign accent. Listeners may have difficulty understanding about 1/3 of the words. Many consonants may be distorted or omitted. Consonant sequences may be non-English. Stress is placed in a non-English manner; unstressed words may be reduced or omitted, and a few syllables added or Missed.
0 Non-English	Pronunciation seems completely characteristic of another language. Many consonants and vowels are mispronounced, mis-ordered or omitted. Listeners may find more than 1/2 of the speech unintelligible. Stressed and unstressed syllables are realized in a non-English manner. Several words may have the wrong number of syllables.
Oral fluency	
5 Native-like	Speech shows smooth rhythm and phrasing. There are no hesitations, repetitions, false starts or non-native phonological simplifications.
4 Advanced	Speech has an acceptable rhythm with appropriate phrasing and word emphasis. There is no more than one hesitation, one repetition or a false start. There are no significant non-native phonological simplifications.
3 Good	Speech is at an acceptable speed but may be uneven. There may be more than one hesitation, but most words are spoken in continuous phrases. There are few repetitions or false starts. There are no long pauses and speech does not sound staccato.
2 Intermediate	Speech may be uneven or staccato. Speech (if ≥ 6 words) has at least one smooth three-word run, and no more than two or three hesitations, repetitions or false starts. There may be one long pause, but not two or more.

1 Limited	Speech has irregular phrasing or sentence rhythm. Poor phrasing, staccato or syllabic timing, and/or multiple hesitations, repetitions, and/or false starts make spoken performance notably uneven or discontinuous. Long utterances may have one or two long pauses and inappropriate sentence-level word emphasis.
0 Disfluent	Speech is slow and labored with little discernable phrase grouping, multiple hesitations, pauses, false starts, and/or major phonological simplifications. Most words are isolated, and there may be more than one long pause.

Please note: The minimum and maximum timings indicated for the sections of each part of the test do not add up to the total timings stated. This is because different versions of the test are balanced for total length. No test taker will get the maximum or minimum times indicated.

Example of item scoring

The diagram below illustrates how different types of scores reported in the PTE Academic score report are computed for the item type **Write essay**.



The item type is rated on content; form; vocabulary; spelling; grammar; development, structure and coherence; and general linguistic range.

The item is first scored on **content**. If no response or an irrelevant response is given, the content is scored as 0.

If an acceptable response is provided (a score is received for content), the item will be scored on **form**. If the response is of the appropriate length, a score will be given and the response will then be rated on the remaining traits: **vocabulary, spelling, grammar; development, structure and coherence; and general linguistic range**.

The trait scores for content, form, vocabulary, spelling, grammar, development, structure and coherence, and general linguistic range are used to calculate the total item score.

The scores for vocabulary, spelling, and grammar contribute to those enabling skills scores. Both the development, structure and coherence trait score and general linguistic range trait score contribute to the written discourse enabling skill score.

The total item score contributes to the **communicative skills score** for writing, as well as to the **overall score** reported for performance on the entire test.

3. Using PTE Academic Scores

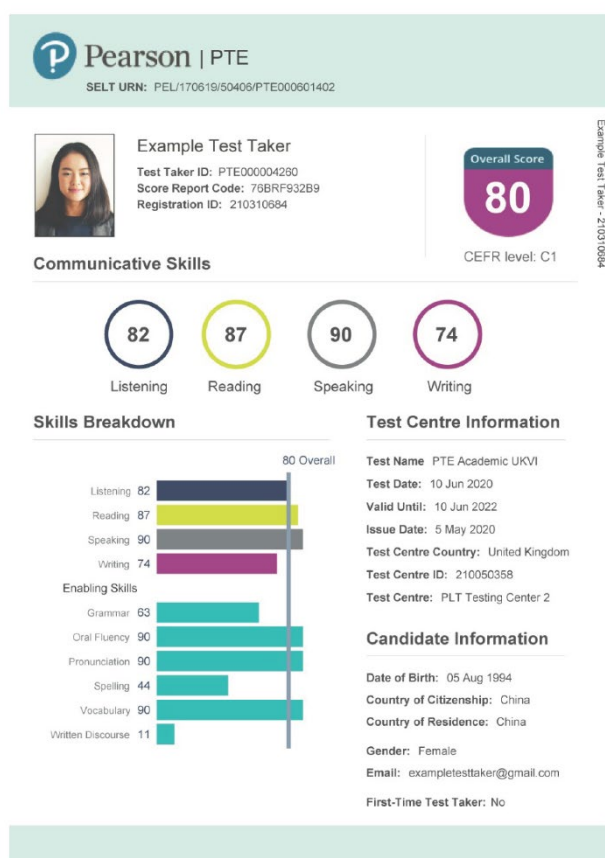
How test takers can use PTE Academic scores

Overall score and communicative skills scores

The score report provides an overall score, a score for each communicative skill and a score for each of the enabling skills.

The overall score is a general measure of a test taker’s ability to deal with English in academic settings. The score range is from 10 to 90 points.

The communicative skills scores provide information about the listening, reading, speaking and writing skills of a test taker. These skills are also scored between 10 and 90 points.



Example Test takers Score Report

In some university programs, the communicative skills scores may provide useful, additional information for making admissions decisions. For example, institutions may:

Set the admission requirement based on the minimum overall score alone, without taking into account communicative skills scores in admission decisions.

Set the admission requirement based on the minimum overall score in combination with a higher minimum on one of the communicative skills scores, because it is considered particularly important for the program the test taker wants to enter.

Set the admission requirement based on the minimum overall score in combination with a lower minimum on one of the communicative skills scores, because it is considered less important for the program the test taker wants to enter.

Universities may also consider other combinations of the overall score and one or more of the communicative skills scores.

Sending scores to governments or institutions

The score submission process is changing.

From April 2019, Pearson will be adding a new Score Report Code (SRC) to your PTE Academic Score Report to make it easier for institutions to access and verify your score in the future. Since we are currently in the process of upgrading the tool used by institutions to check scores, please continue to assign your score to institutions using your PTE Academic account *unless* you receive a notice from an institution that they are ready to accept the new Score Report Code.

You must send your PTE Academic score through our secure portal. Institutions do not accept paper or PDF versions of PTE Academic score reports.

Instructions

1. Login to your account.
2. Click **'View Score Reports.'**
3. Click **'Send Scores'**.
4. Type of name of your chosen institution in the field marked **'Institution/Organization/Department/School'** and click **'Search'**.
5. Tick the box next to the institution's name when it appears in the list.
6. Review your details, then scroll down the page and click **'Next'** and then **'Next'** again to confirm.

Repeat for each institution. (You can send your Score Report to an unlimited number of institutions, but you can only select up to seven recipients at any one time.)

Applying for a visa to work or study in Australia or New Zealand?

If, when booking your test, you ticked one of the boxes asking for scores to be allocated to the DHA (Australia) or INZ (New Zealand) your score will have been automatically sent to them.

4. Estimates of concordance between PTE Academic, CEFR, TOEFL iBT and IELTS

Based on research and empirical studies, Pearson has produced concordance tables showing the relationship between the PTE Academic test, the IELTS Academic test and TOEFL. The table on the next page shows Pearson's current best estimate of concordance between PTE Academic scores and the Common European Framework of Reference for Languages (CEFR). In addition, shaded score ranges indicate the PTE Academic scores that predict some degree of performance at the next CEFR level.

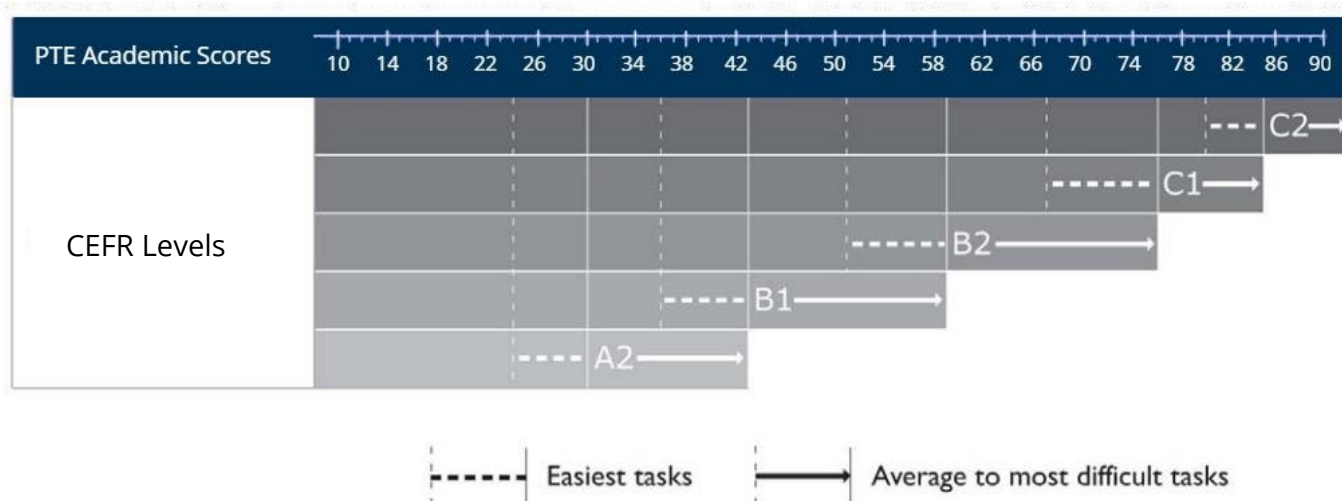
Please note that any attempt to predict a score on a particular test, based on the score observed on another test, will contain measurement error. This is caused by the inherent error in each of the tests in the comparison and in the estimate of the concordance. Furthermore, tests in the comparison do not measure exactly the same construct.

The PTE Academic Score Scale and the CEFR

The explanation of the alignment of PTE Academic to the CEFR is that to stand a reasonable chance at successfully performing any of the tasks defined at a particular CEFR level, learners must be able to demonstrate that they can do the average tasks at that level.

As students grow in ability, for example within the B1 level, they will become successful at doing even the most difficult tasks at that level and will also find they can cope with the easiest tasks at the next level. In other words, they are entering into the B2 level.

The diagram below shows PTE Academic scores aligned to the CEFR levels A2 to C2. The dotted lines on the scale show the PTE Academic score ranges that predict that test takers are likely to perform successfully on the easiest tasks at the next higher level. For example, if a candidate scores 51 on PTE Academic, this means that they are likely to be able to cope with the more difficult tasks within the CEFR B1 level. At the same time, according to their PTE Academic score, it predicts that they are likely to perform successfully on the easiest tasks at B2.



Alignment of PTE Academic scores to the CEFR

Below find descriptions of those threshold levels defined by success on the easier tasks for B2 and B1.

B2 Threshold Level

Has sufficient command of the language to deal with most familiar situations but will often require repetition and make many mistakes. Can deal with standard spoken language but will have problems in noisy circumstances. Can exchange factual information on familiar routine and non-routine matters within his/her field with some confidence. Can pass on a detailed piece of information reliably. Can understand the information content of the majority of recorded or broadcast material on topics of personal Interest delivered in clear standard speech.

B1 Threshold Level

Has limited command of language, but it is sufficient in most familiar situations provided language is simple and clear. May be able to deal with less routine situations on public transport e.g., asking another passenger where to get off for an unfamiliar destination. Can re-tell short written passages in a simple fashion using the wording and ordering of the original text. Can use simple techniques to start, maintain or end a short conversation. Can tell a story or describe something in a simple list of points.

What PTE Academic scores mean

PTE Academic alignment with the CEFR can only be fully understood if it is supported with information showing what it really means to be 'at a level'. In other words, are test takers likely to be successful with tasks at the lower boundary of a level; do they stand a fair chance of doing well on any task, or will they be able to do almost all the tasks, even the most difficult ones, at a particular level? The table below shows for each of the CEFR levels A2 to C2 which PTE Academic scores predict the likelihood of a test taker performing successfully on the easiest, average and most difficult tasks within each of the CEFR levels.

PTE Academic scores predicting the likelihood of successful performance on CEFR level tasks			
CEFR Level	Easiest	Average	Most Difficult
C2	80	85	N/A
C1	67	76	84
B2	51	59	75
B1	36	43	58
A2	24	30	42

For example, if a test taker's PTE Academic score is 36, this predicts that they will perform successfully on the easiest tasks at B1. From 36 to 43, the likelihood of successfully performing the easiest tasks develops into doing well on the average tasks at B1. Finally, reaching 58 predicts that a candidate will perform well at the most difficult B1 level tasks.

Estimate of concordance between PTE Academic scores and TOEFL iBT

TOEFL iBT Score	PTE Academic Score	TOEFL iBT Score	PTE Academic Score
No data	85 - 90	90	61
120	84	89	60
119	83	87-88	59
118	82	86	58
117	81	85	57
115-116	80	83-84	56
114	79	82	55
113	78	81	54
112	77	79-80	53
110-111	76	78	52
109	75	76-77	51
107-108	74	74-75	50
106	73	72-73	49
105	72	70-71	48
103-104	71	67-69	47
102	70	65-66	46
101	69	63-64	45
99-100	68	60-62	44
98	67	57-59	43
97	66	54-56	42
95-96	65	52-53	41
94	64	48-51	40
93	63	45-47	39
91-92	62	40-44	38
		No data	10 - 37

Estimate of concordance between PTE Academic and IELTS

IELTS Score	PTE Academic Score
9.0	N/A
8.5	89 - 90
8.0	84 - 88
7.5	76 - 83
7.0	66 - 75
6.5	56 - 65
6.0	46 - 55
5.5	36 - 45
5.0	29 - 35
4.5	23 - 28
No data	10 - 22

5. Scored Samples

Automated scoring

As the worldwide leader in publishing and assessment for education, Pearson uses several proprietary, patented technologies to automatically score test takers' performance on PTE Academic. Academic institutions, corporations and government agencies around the world have selected Pearson's automated scoring technologies to measure the abilities of students, staff or applicants. Pearson customers using automated spoken and written assessments include eight of the 2008 Fortune Top 20 companies; 11 of the 2008 Top 15 Indian BPO companies; the U.S., German and Dutch governments; world sports organizations, such as FIFA (organizers of the World Cup) and the Asian Games; major airlines and aviation schools; and leading universities and language schools.

An extensive field test program was conducted to test PTE Academic's test items and evaluate their effectiveness as well as to obtain the data necessary to train the automated scoring engines to evaluate PTE Academic items. Data was collected from more than 10,000 test takers from 38 cities in 21 countries who participated in PTE Academic's field test. These test takers came from 158 different countries and spoke 126 different native languages, including (but not limited to) Cantonese, French, Gujarati, Hebrew, Hindi, Indonesian, Japanese, Korean, Mandarin, Marathi, Polish, Spanish, Urdu, Vietnamese, Tamil, Telugu, Thai and Turkish. The data from the field test were used to train the automated scoring engines for both the written and spoken PTE Academic items.

By combining the power of a comprehensive field test, in-depth research and Pearson's proven, proprietary automated scoring technologies, PTE Academic fills a critical gap by providing a state-of-the-art test that accurately measures the English language speaking, listening, reading and writing abilities of non-native speakers.

Scoring written English skills

The written portion of PTE Academic is scored using the Intelligent Essay Assessor™ (IEA), an automated scoring tool that is powered by Pearson's state-of-the-art Knowledge Analysis Technologies™ (KAT™) engine. Based on more than 20 years of research and development, the KAT engine automatically evaluates the meaning of text by examining whole passages. The KAT engine evaluates writing as accurately as skilled human raters using a proprietary application of the mathematical approach known as Latent Semantic Analysis (LSA). Using LSA (an approach that generates semantic similarity of words and passages by analyzing large bodies of relevant text) the KAT engine "understands" the meaning of text much the same as a human does.

IEA can be tuned to understand and evaluate text in any subject area and includes built-in detectors for off-topic responses or other situations that may need to be referred to human readers. Research conducted by independent researchers as well as Pearson supports IEA's reliability for assessing knowledge and knowledge-based reasoning. IEA was developed more than a decade ago and has been used to evaluate millions of essays, from scoring student writing at elementary, secondary and university level, to assessing military leadership skills.

Scoring spoken English skills

The spoken portion of PTE Academic is automatically scored using Pearson's Ordinate technology. Ordinate technology is the result of years of research in speech recognition, statistical modeling, linguistics and testing theory. The technology uses a proprietary speech processing system that is specifically designed to analyze and automatically score speech from native and non-native speakers of English. In addition to recognizing words, the system locates and evaluates relevant segments, syllables and phrases in speech and then uses statistical modeling technologies to assess spoken performance.

To understand the way that the Ordinate technology is "taught" to score spoken language, think about a person being trained by an expert rater to score speech samples during interviews. First, the expert rater gives the trainee rater a list of things to listen for in the test taker's speech during the interview. Then the trainee observes the expert testing numerous test takers, and, after each interview, the expert shares with the trainee the score he or she gave the test taker and the characteristics of the performance that led to that score. Over several dozen interviews, the trainee's scores begin to look very similar to the expert rater's scores. Ultimately, one could predict the score the trainee would give a particular test taker based on the score that the expert gave.

This, in effect, is how the machine is trained to score, only instead of one expert teaching the trainee, there are many expert scorers feeding scores into the system for each response, and instead of a few dozen test takers, the system is trained on thousands of responses from hundreds of test takers. Furthermore, the machine does not need to be told what features of the speech are important; the relevant features and their relative contributions are statistically extracted from the massive set of data when the system is optimized to predict human scores.

Ordinate technology powers the Versant™ line of language assessments, which are used by organizations such as the U.S. Department of Homeland Security, schools of aviation around the world, the Immigration and Naturalization Service in the Netherlands, and the U.S. Department of Education. Independent studies have demonstrated that Ordinate's automated scoring system can be more objective and more reliable than many of today's best human-rated tests, including one-on-one oral proficiency interviews.

More information about automated scoring is available on our website:
www.pearsonpte.com/organisations/teachers-teaching-resources/scoring/

Enabling skill scores and trait scores examples

The PTE Academic test taker's score report includes:

1. Overall test scores,
2. Communication scores of the four skills of Reading, Writing, Listening and Speaking,
3. Enabling skills in Writing and Speaking – Grammar, Oral fluency, Pronunciation, Spelling, Vocabulary, and Written Discourse.

The sections below show examples of how scoring operates for Speaking and Writing items. The scoring mechanisms in these items are based on collecting data on multiple relevant traits in each item, giving them each a score, and then converting them all to an overall score in either speaking or writing. The automated system is trained on the trait scores of hundreds of items scored by human expert markers. Once trained, our automated systems can then quickly score all new speaking and writing items quickly and accurately. The traits measured in PTE Academic include:

- Content
- Oral Fluency
- Pronunciation
- Form
- Development, structure and coherence
- Grammar
- General linguistic range
- Vocabulary

Some of these traits are reported also as discrete enabling skills, while other traits are combined to contribute to more generic enabling skills such as written discourse.

Spoken samples

The PTE Academic automated scoring system correlates highly with human ratings. Studies have been carried out to compare human and machine scores for the speaking item type **Describe image** using tasks such as the example below.

Example Describe image item

Look at the graph below. In 25 seconds, please speak into the microphone and describe in detail what the graph is showing. You will have 40 seconds to give your response.

Year	Funding council grants	Research	Overseas fees	Services	Endowments
1994-5	520	180	50	70	20
1997-8	550	200	50	100	20
2001-2	700	300	80	100	20

Recorded Answer

Current Status:
Recording

Samples of test taker responses at B1, B2 and C1 were collected as well as comments from the Language Testing division of Pearson. The ratings on each response include a machine score and scores from at least two human raters. In cases where the two human rater scores differed, an adjudicator was used to provide a third human rating.

Scoring

The *Describe image* item is scored on 3 different traits:

Traits	Maximum raw score	Human rating	Machine score
Content	5	5	5
Oral fluency	5	5	5
Pronunciation	5	5	5
Maximum item score	15	15	15

These traits are scored as shown in the chart that follows.

Content	Pronunciation	Oral fluency
<p>5: Describes all elements of the image and their relationships, possible development and conclusion or implications</p>	<p>5 Native-like: All vowels and consonants are produced in a manner that is easily understood by regular speakers of the language. The speaker uses assimilation and deletions appropriate to continuous speech. Stress is placed correctly in all words and sentence-level stress is fully appropriate.</p>	<p>5 Native-like: Speech shows smooth, rhythm and phrasing. There are no hesitations, repetitions, false starts or non-native phonological simplifications.</p>
<p>4: Describes all the key elements of the image and their relations, referring to their implications or conclusions</p>	<p>4 Advanced: Vowels and consonants are pronounced clearly and unambiguously. A few minor consonant, vowel or stress distortions do not affect intelligibility. All words are easily understandable. A few consonants or consonant sequences may be distorted. Stress is placed correctly on all common words, and sentence level stress is reasonable.</p>	<p>4 Advanced: Speech has an acceptable rhythm with appropriate phrasing and word emphasis. There is no more than one hesitation, one repetition or a false start. There are no significant non-native phonological simplifications.</p>
<p>3: Deals with most key elements of the image and refers to their implications or conclusions</p>	<p>3 Good: Most vowels and consonants are pronounced correctly. Some consistent errors might make a few words unclear. A few consonants in certain contexts may be regularly distorted, omitted or mispronounced. Stress-dependent vowel reduction may occur on a few words.</p>	<p>3 Good: Speech is at an acceptable speed but may be uneven. There may be more than one hesitation, but most words are spoken in continuous phrases. There are few repetitions or false starts. There are no long pauses and speech does not sound staccato.</p>
<p>2: Deals with only one key element in the image and refers to an implication or conclusion. Shows basic understanding of several core elements of the image</p>	<p>2 Intermediate: Some consonants and vowels are consistently mispronounced in a non-native like manner. At least 2/3 of speech is intelligible, but listeners might need to adjust to the accent. Some consonants are regularly omitted, and consonant sequences may be simplified. Stress may be placed incorrectly on some words or be unclear.</p>	<p>2 Intermediate: Speech may be uneven or staccato. Speech (if >= 6 words) has at least one smooth three-word run, and no more than two or three hesitations, repetitions or false starts. There may be one long pause, but not two or more.</p>
<p>1: Describes some basic elements of the image, but does not make clear their interrelations or implications</p>	<p>1 Intrusive: Many consonants and vowels are mispronounced, resulting in a strong intrusive foreign accent. Listeners may have difficulty understanding about 1/3 of the words. Many consonants may be distorted or omitted. Consonant sequences may be non-English. Stress is placed in a non-</p>	<p>1 Limited: Speech has irregular phrasing or sentence rhythm. Poor phrasing, staccato or syllabic timing, and/or multiple hesitations, repetitions, and/or false starts make spoken performance notably uneven or discontinuous. Long utterances may have one or two long pauses</p>

Content	Pronunciation	Oral fluency
	English manner; unstressed words may be reduced or omitted and a few syllables added or missed.	and inappropriate sentence-level word emphasis.
<p>0: Mentions some disjointed elements of the presentation</p>	<p>0 Non-English: Pronunciation seems completely characteristic of another language. Many consonants and vowels are mispronounced, mis-ordered or omitted. Listeners may find more than 1/2 of the speech unintelligible. Stressed and unstressed syllables are realized in a non-English manner. Several words may have the wrong number of syllables</p>	<p>0 Disfluent: Speech is slow and labored with little discernable phrase grouping, multiple hesitations, pauses, false starts, and/or major phonological simplifications. Most words are isolated, and there may be more than one long pause</p>

Test Taker responses

Test-taker A: mid B1 Level

 [Listen to audio sample 'Test taker A'](#)

Comment on response

The response lacks some of the main contents. Only some obvious information from the graph is addressed. Numerous hesitations, non-native-like pronunciation, poor language use and limited control of grammar structures at times make the response difficult to understand.

How the response was scored

The table below and subsequent tables under 'How the response was scored' show the machine scores and the human ratings that have been assigned to this response. When the cells in the adjudicator column are empty, the adjudicator score does not deviate from the scores given by the first and second human rater.

Trait name	Maximum raw score	Machine score	Human rater 1	Human rater 2	Adjudicator
Content	5	1.69	2	2	
Oral fluency	5	1.62	4	2	2
Pronunciation	5	1.41	2	2	
Total item score	15	4.72	8	6	6

Test taker B: mid B2 Level

 [Listen to audio sample 'Test taker B'](#)

Comment on response

The test taker discusses some aspects of the graph and the relationship between elements, though some key points have not been addressed. The rate of speech is acceptable. Language use and vocabulary range are quite weak. There are some obvious grammar errors and inappropriate stress and pronunciation.

How the response was scored

Trait name	Maximum raw score	Machine score	Human rater 1	Human rater 2	Adjudicator
Content	5	2.50	2	3	2
Oral fluency	5	3.71	4	5	3
Pronunciation	5	3.28	3	4	2
Total item score	15	9.49	9	12	7

Test taker C: mid C1 Level

 [Listen to audio sample 'Test taker C'](#)

Comment on response

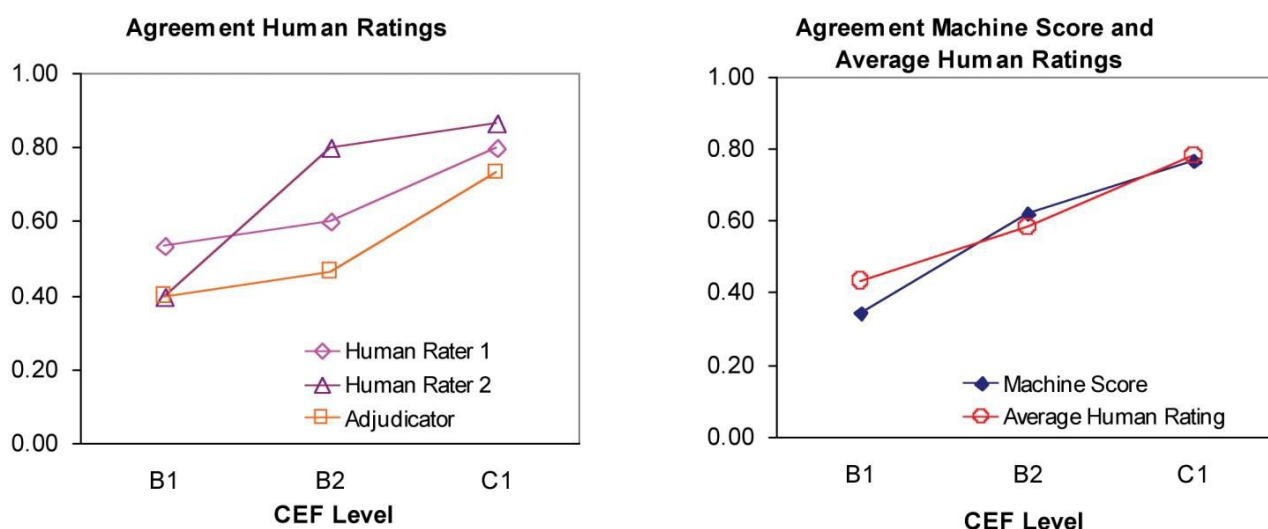
The test taker discusses the major aspects of the graph and the relationship between elements. The response is spoken at a fluent rate and language use is appropriate. There are few grammatical errors in the response. The candidate demonstrates a wide range of vocabulary. Stress is appropriately placed.

How the response was scored

Trait name	Maximum raw score	Machine score	Human rater 1	Human rater 2	Adjudicator
Content	5	2.70	3	4	3
Oral fluency	5	4.03	4	5	4
Pronunciation	5	4.02	5	4	4
Total item score	15	10.75	12	13	11

Overall performance rating

As shown from the scoring tables on the responses presented, the human ratings at trait level differed up to two score points out of six possible scoring categories (0 - 5). The two graphs below show the level of agreement of the total item score (sum of traits) of the human raters (graph on the left) and the agreement of the machine score with the average of the human ratings (graph on the right). The total item scores are rendered as a proportion of the total maximum item score (15) for the item. The human ratings vary substantially, especially for the B2 candidate, from a score that is only slightly higher than the score given to the B1 test taker, to a score that is close to the one given to the C1 test taker.



Note that these ratings were given by trained raters who had all recently passed a rater's exam. This example is therefore not typical for the human rating in general, but it shows that in some instances, especially for spoken responses, human raters have a hard time deciding on the most fitting score.

The automatic scoring system that has been trained on more than 100 human raters agrees quite well with the average human rating as shown in the graph on the right.

The machine-human comparison is part of the validation studies based on the field test responses for speaking, where 450,000 spoken responses were collected and scored, generating more than 1 million human ratings. The correlation between the human raw scores and the machine-generated scores for the overall measure of speaking was 0.89. In order to neutralize the effect of differences in severity amongst human raters, the human scores were scaled using Item Response Theory (IRT). The correlation with the machine scores then increases to 0.96. The reliability of the measure of speaking in PTE Academic is 0.91.

Score type	Human-human	Machine-human
Raw scores	0.87	0.89
IRT scaled	0.90	0.96

Written samples

The PTE Academic automated scoring system correlates highly with average human ratings. Studies were carried out to compare human and machine scores for the writing item type **Write essay**, using tasks such as the example below.

Example *Write essay* item 'Tobacco'

You will have 20 minutes to plan, write and revise an essay about the topic below. Your response will be judged on how well you develop a position, organize your ideas, present supporting details, and control the elements of standard written English. You should write 200-300 words.

Tobacco, mainly in the form of cigarettes, is one of the most widely-used drugs in the world. Over a billion adults legally smoke tobacco every day. The long term health costs are high - for smokers themselves, and for the wider community in terms of health care costs and lost productivity.

Do governments have a legitimate role to legislate to protect citizens from the harmful effects of their own decisions to smoke, or are such decisions up to the individual?

From the studies using these items, samples of test taker responses at B1, B2 and C1 are given as well as a comment from the Language Testing division of Pearson. Ratings on each response are provided including a machine score and scores from at least two human raters. In cases where the two human rater scores differed, an adjudicator was used to provide a third human rating.

Scoring

The item type *Write essay* is scored on 7 different traits:

Traits	Maximum raw score	Human rating	Machine score
Content	3	3	3
Form	2		
Development, structure and coherence	2	2	2
Grammar	2	2	2
General linguistic range	2	2	2
Vocabulary range	2	2	2
Spelling	2		
Maximum item score	15	11	15

The form and spelling traits do not require human ratings for training the automatic scoring systems as they can be objectively scored. It can be assumed (if the human raters work error-free) that the human rating on these two traits would have been identical to the machine score.

To make the total score from human rating comparable to the machine score, we need to take the score as a proportion of the maximum obtainable score by dividing the observed total score by the maximum possible score.

An item is not scored if the test taker’s response does not meet the minimum requirements for the traits content and form (i.e., when a test taker scores 0 for content and/or form).

The traits are scored as follows:

Content	Form	Development, structure and coherence	Grammar
3: Adequately deals with the prompt			
2: Deals with the prompt but does not deal with one minor aspect	2: Length is between 200 and 300 words	2: Shows good development and logical structure	2: Shows consistent grammatical control of complex language. Errors are rare and difficult to spot
1: Deals with the prompt but omits one major aspect or more than one minor aspect	1: Length is between 120 and 199 or between 301 and 380 words	1: Is incidentally less well structured, and some elements or paragraphs are poorly linked	1: Shows a relatively high degree of grammatical control. No mistakes which would lead to misunderstandings
0: Does not deal properly with the prompt	0: Length is less than 120 or more than 380 words. Essay is written in capital letters, contains no punctuation or only consists of bullet points or very short Sentences.	0: Lacks coherence and mainly consists of lists or loose elements	0: Contains mainly simple structures and/or several basic mistakes

General linguistic range	Vocabulary range	Spelling
<p>2: Exhibits mastery of a wide range of language to formulate thoughts precisely, give emphasis, differentiate and eliminate ambiguity. No sign that the test taker is restricted in what they want to communicate</p>	<p>2: Good command of a broad lexical repertoire, idiomatic expressions and colloquialisms</p>	<p>2: Correct spelling</p>
<p>1: Sufficient range of language to provide clear descriptions, express viewpoints and develop arguments</p>	<p>1: Shows a good range of vocabulary for matters connected to general academic topics. Lexical shortcomings lead to circumlocution or some Imprecision.</p>	<p>1: One spelling error</p>
<p>0: Contains mainly basic language and lacks precision</p>	<p>0: Contains mainly basic vocabulary insufficient to deal with the topic at the required level</p>	<p>0: More than one spelling error</p>

Test Taker Responses

Test taker A: mid B1 Level

Tobacco, mainly in the form of cigarettes, is one of the most widely-used drugs in the world. Over a billion adults legally smoke tobacco everyday. Recently, it is not only the adult. Even the high school students or college students smoke just because they want to know how it feels. It is also not limited by gender. Lots of women are smokers. Even the old people still smoke, as if they do not care about their healthy. Become a smoker is like make someone just care about the good feeling of smoking and makes them to forget the risks they will face in the future. The long term health costs are high - for smokers themselves, and for the wider community in temrs of health care costs and lost productivity. The worst risk that the smokers will face is lung cancer, which can cause death. The governments have a legitimate role to legislate to protect citizens from the harmful effects of their own decisions to smoke. For example they make rule about no smoking area, in the street, and public place. But it also the decisions of each individual wheter they want to continue their life as a smoker and take all the risk, or stop and learn to life healthier. (211 words)

Comment on response

The response is a simple essay which gives a minimal answer to the question. The argument contains insufficient supporting ideas. The structure is lacking in logic and coherence. There is frequent misuse of grammar and vocabulary. Vocabulary range is limited and inappropriate at times.

How the response was scored

The table below and subsequent ones under 'How the response was scored' show the machine scores and the human ratings that have been assigned to this response. When the cells in the adjudicator column are empty, the adjudicator score does not deviate from the scores given by the first and second human rater.

Trait name	Maximum raw score	Machine score	Human rater 1	Human rater 2	Adjudicator
Content	2	1.80	2	2	
Development, structure and coherence	2	1.35	0	1	1
Form	2	2.00	n/a	n/a	
General linguistic Range	2	1.03	1	1	
Grammar	2	1.07	1	1	
Spelling	2	0.00	n/a	n/a	
Vocabulary range	2	0.93	1	2	1
Total item score	14	8.18	5	7	6

Test taker B: mid B2 Level

In my opinion it should be a combined effort of both government and an individual. In some countries specially in UK, government is trying to impose laws and regulations which discourage smoking, for example the law which prohibits smoking in pubs, bars and public areas. Also there are TV commercials and banners which explain the long term effects of smoking. As a result there has been some reduction in the number of people smoking before the law and now. But this effort is not enough. Until and unless an individual doesn't make an effort himself the problem cannot be solved. One has to have control of his own body and will power to overcome this habit turned necessity of the body. There has been a significant increase in amount of people who are approaching medical practitioners and NHS to help them to overcome this problem. There are also some NGO's who are working in this field. I think if we can spread awareness about the ill effects of smoking to teenagers, there will be less number of people who start smoking at the first place. It is a collective responsibility of government and parents as well. To conclude I can say that youngsters are the people who get fascinated by the whole idea of smoking, thus this concept should be changed by the efforts of government, media and by us as an individual. (234 words)

Comment on response

A systematic argument with appropriate highlighting of significant points and relevant supporting detail has been developed. Ability to evaluate different ideas or solutions to a problem has been demonstrated. However, some obvious grammar errors and inappropriate use of vocabulary can be found. There are also quite a number of spelling errors.

How the response was scored

Trait name	Maximum raw score	Machine score	Human rater 1	Human rater 2	Adjudicator
Content	3	2.25	3	1	2
Development, structure and coherence	3	1.17	2	1	2
Form	3	2.00	n/a	n/a	
General linguistic range	3	1.42	1	1	
Grammar	3	1.68	1	2	3
Spelling	3	0.00	n/a	n/a	
Vocabulary range	3	1.32	1	1	
Total item score	14	9.84	8	6	9

Test taker C: mid C1 Level

Outlawing tobacco use would create unprecedented controversy. Billions of people worldwide smoke; whether they are chain smokers or recreational smokers. Also, there are several multi-million dollar cigarette companies that will also suffer many consequences if tobacco use is made illegal. We must also consider the thousands of employees who will be left unemployed if such a legislation is made. Unfortunately, it is an industry that makes ridiculous amounts of money for many people, so the likelihood of banning it is minimal.

Nonetheless, it is a change that would benefit society on many levels in the long run. Smoking causes so many health care issues, so if smoking is made illegal, morbidity and mortality rates would be reduced significantly. Quality of life will be improved dramatically, and it will allow more people to enjoy their lives significantly longer.

Legislators must also consider the rights of the individual. Shouldn't every individual have the right to choose how they treat their body? The government can argue that these individuals may do as they wish, but then they must also suffer the consequences without government funding. They must take full responsibility for any health issues developed as a result of tobacco use, and not expect Medicare or health insurance to cover costs caused by their own irresponsible negligent decisions.

In essence, if individuals wish to make their own decisions to smoke, they must consider all the possible outcomes, and be willing to deal with these outcomes accordingly. (243 words)

Comment on response

Clear, well-structured exposition on the topic which touches upon the relevant issues. Points of view are given at some length with subsidiary points. Reasons and relevant examples are demonstrated. General linguistic range and vocabulary range are excellent. Phrasing and word choice are appropriate. There are very few grammar errors. Spelling is excellent.

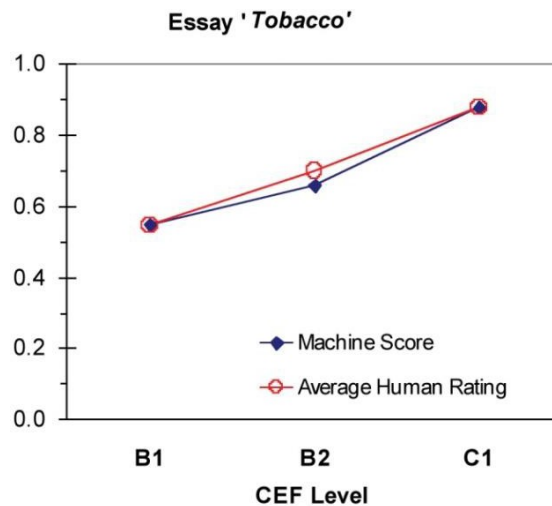
How the response was scored

Trait name	Maximum raw score	Machine score	Human rater 1	Human rater 2	Adjudicator
Content	3	2.74	1	2	3
Development, structure and coherence	3	1.97	2	2	
Form	3	2.00	n/a	n/a	
General linguistic range	3	2.00	2	2	
Grammar	3	1.70	2	2	
Spelling	3	1.00	n/a	n/a	
Vocabulary range	3	1.82	1	2	2
Total item score	14	13.23	8	10	11

Overall performance rating

As can be seen from the scoring tables on the essay responses, the machine scores correspond closely to the average human score. Although there is some variation at the trait level, the total item scores agree to a high degree. To illustrate this agreement the graph below shows the machine scores and the average human scores.

The graph illustrates the total (proportional) item score from the machine and from the human ratings for the essay responses. The results show that the machine generated total item scores are closely aligned with the average over the human ratings.



The machine-human comparison is part of the validation studies based on the field test responses for writing, where 50,000 written responses were collected and scored, generating about 0.6 million human ratings.

The correlation between the human raw scores and the machine-generated scores for the overall measure of writing was 0.88. In order to neutralize the effect of differences in severity amongst human raters, the human scores were scaled using IRT. The correlation with the machine scores then increases to 0.93. The reliability of the measure of writing in PTE Academic is 0.89.

Score Type	Human-Human	Machine-Human
Raw scores	0.87	0.88
IRT scaled	0.90	0.93

6. Glossary

CEFR (also known as CEF) – The Common European Framework of Reference for Languages put together by the Council of Europe to standardize the levels of language exams in different regions. Other exams are mapped to the CEFR.

Communicative skills – Listening, reading, speaking and writing. These are the scores you will be judged on for visa/HEI entry.

Concordance studies – The relationship between two or more scales of measurement.

Enabling skills – Skills that are used to rate performance in the communicative skills of speaking and writing. The enabling skills measured in the PTE Academic are grammar, oral fluency, pronunciation, spelling, vocabulary, and written discourse. Test takers receive reported scores for enabling skills (unlike traits).

Enabling skills, grammar – Correct use of language with respect to word form and word order at the sentence level.

Enabling skills, oral fluency – Smooth, effortless and natural-paced delivery of speech.

Enabling skills, pronunciation – Ability to produce speech sounds in a way that is easily understandable to most regular speakers of the language. Regional or national pronunciation variants are considered correct to the degree that they are understandable to most regular speakers of the language.

Enabling skills, spelling – Writing of words according to the spelling rules of the language. All national variations in spelling are considered correct.

Enabling skills, vocabulary – Appropriate choice of words used to express meaning precisely in written and spoken English, as well as lexical range.

Enabling skills, written discourse – Correct and communicatively efficient production of written language at the textual level. Written discourse skills are manifested in the structure of a written text, its internal coherence, logical development, and the range of linguistic resources used to express meaning precisely.

Integrated skills items – Items on the test that require the use of more than one skill such as assessing reading and speaking, listening and speaking, reading and writing, listening and writing, or listening and reading.

Item Response Theory (IRT) – A testing theory. IRT is based on the relationship between an individual's performance on a test item and that individual's levels of performance on an overall measure of the ability that item was designed to measure.

Formal aspects – The form of a response: for example, whether it is over or under the word limit for a particular item type.

IELTS – International English Language Testing System. This test measures the language proficiency of people who want to study or work where English is used as a language of communication.

Lexical – Relating to the vocabulary of a language: For example, *lexical knowledge*.

Ordinate technology – A proprietary speech processing system that is specifically designed to analyze and automatically score speech from native and non-native speakers of English.

Overall score – Score based on test taker’s performance on all test items.

PTE Academic – Pearson Test of English Academic. PTE Academic is a 3-hour long, computer-based assessment of a person’s English language ability in an academic context. The test assesses an individual’s communicative skills of Reading, Writing, Listening and Speaking skills through questions using authentically-sourced material. In addition, the test provides feedback on enabling skills in the form of Oral Fluency, Grammar, Vocabulary, Written Discourse, Pronunciation and Spelling.

TOEFL iBT – A test that measures your ability to use and understand English at the university level, and evaluates how well you combine your reading, listening, speaking, and writing skills to perform academic tasks.

Traits – Items measured in PTE Academic that contribute to overall scores. These include content; oral fluency; pronunciation; form; development, structure and coherence; grammar; general linguistic range; and vocabulary.

Appendix

The following table shows the PTE Academic score aligned to the CEFR discussed in section 2 and describes performance indicators at these levels.

PTE Academic Score	Common European Framework Level	Level Descriptor ¹	What does this mean for a score user?
85 - 90	C2	Can understand with ease virtually everything heard or read. Can summarize information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.	C2 is a highly proficient level and a student at this level would be extremely comfortable engaging in academic activities at all levels.
76 - 84	C1	Can understand a wide range of demanding, longer texts and recognize implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organizational patterns, connectors and cohesive devices.	C1 is a level at which a student can comfortably participate in all post-graduate activities including teaching. It is not required for students entering university at undergraduate level. Most international students who enter university at a B2 level would acquire a level close to or at C1 after living in the country for several years, and actively participating in all language activities encountered at university.

¹ © The copyright of the level descriptors reproduced in this document belongs to the Council of Europe.
Version 15 - April 2021

PTE Academic Score	Common European Framework Level	Level Descriptor ¹	What does this mean for a score user?
59 - 75	B2	Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialization. Can interact with a degree of oral fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.	B2 was designed as the level required to participate independently in higher level language interaction. It is typically the level required to be able to follow academic level instruction and to participate in academic education, including both coursework and student life.
43 - 58	B1	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst in an area where the language is spoken. Can produce simple connected text on topics, which are familiar or of personal interest. Can describe experiences and events, dreams, hopes and ambitions and briefly give reasons and explanations for opinions and plans.	B1 is insufficient for full academic level participation in language activities. A student at this level could 'get by' in everyday situations independently. To be successful in communication in university settings, additional English language courses are required.

PTE Academic Score	Common European Framework Level	Level Descriptor ¹	What does this mean for a score user?
30 - 42	A2	Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g., very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.	A2 is an insufficient level for academic level participation.
10 - 29	A1 or below	Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.	A1 is an insufficient level for academic level participation.

© Copyright Pearson Education Ltd 2021. All rights reserved; no part of this publication may be reproduced without the prior written permission of Pearson Education Ltd.