



PTE ACADEMIC™

Score Guide

Version 8/ October 2017



Pearson

Contents

Introduction	4
1. Reported Scores: An Overview	5
Overall score	5
Communicative skills scores	5
Enabling skills scores	6
2. Item Scoring: An Overview	7
Correct or incorrect	7
Partial credit	7
3. Item Scoring: Skills Tested and Scoring Criteria	11
Part 1 Speaking and writing	11
Read aloud	11
Repeat sentence	13
Describe image.....	15
Re-tell lecture.....	17
Answer short question.....	19
Summarize written text	20
Write essay.....	22
Scoring criteria: Pronunciation and Oral fluency.....	24
Part 2 Reading	26
Multiple-choice, choose single answer.....	26
Multiple-choice, choose multiple answers	27
Re-order paragraphs.....	28
Reading: Fill in the blanks.....	29
Reading and writing: Fill in the blanks.....	30
Part 3 Listening	31
Summarize spoken text.....	31
Multiple-choice, choose multiple answers	33
Fill in the blanks	34
Highlight correct summary	35
Multiple-choice, choose single answer.....	36
Select missing word.....	37
Highlight incorrect words	38
Write from dictation	39
4. Using PTE Academic Scores.....	40
How institutions can use PTE Academic scores	40
Overall score and communicative skills scores	40
Enabling skills scores.....	42
Alignment with CEF	43
The PTE Academic Score Scale and the CEF	43
What PTE Academic scores mean	44
PTE Academic Requirements	45
Error of measurement	48
Overall score and communicative skills scores	48

Enabling skills scores.....	49
Test reliability.....	49
5. Estimates of Concordance between PTE Academic, TOEFL and IELTS	51
Test comparisons using field test data.....	51
Information on concordances since the launch of PTE Academic.....	52
Relation to the Common European Framework	52
Validity check using BETA testing data.....	53
Concordance of PTE Academic with other measures of English	53
Estimates of concordance between PTE Academic and the descriptive scale of the CEF	54
Estimates of concordance between PTE Academic and TOEFL iBT	57
Estimates of concordance between PTE Academic and IELTS	58
6. Scored Samples	59
Automated scoring	59
Scoring written English skills.....	59
Scoring spoken English skills	60
Spoken samples	61
Example <i>Describe image</i> item	61
Test Taker responses	64
Overall performance rating.....	67
Written samples.....	68
Example <i>Write essay</i> item 'Tobacco'	68
Test Taker Responses	71
Overall performance rating.....	74
7. References.....	75
Using PTE Academic scores.....	75
Concordance to other tests.....	75

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 in order to assist education institutions and professional and government organizations that require a standard of academic English language proficiency for admission purposes.

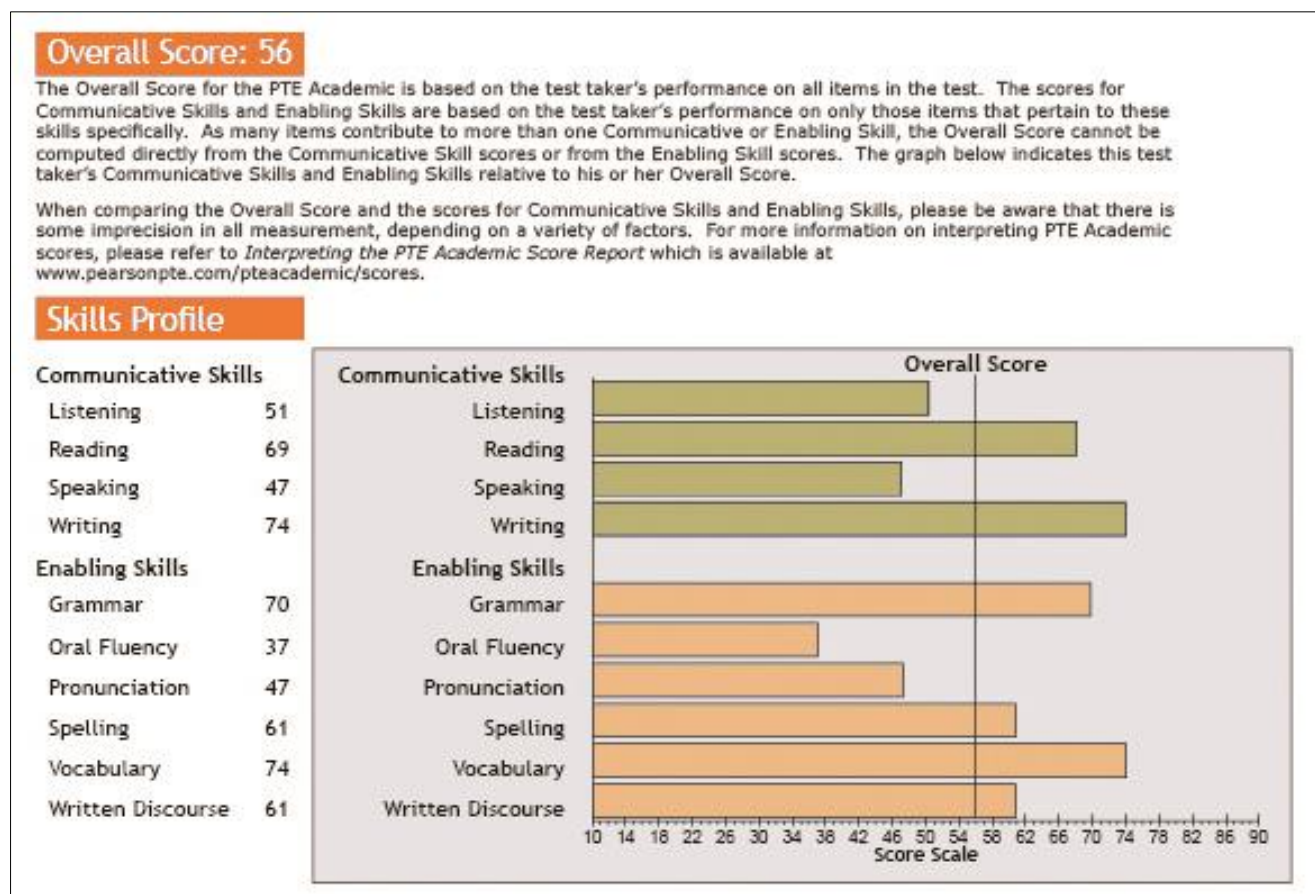
The *Score Guide* is designed for anyone who wants to learn more about how the different tasks in PTE Academic are scored. The Guide will help you to understand:

- 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.

1. Reported Scores: An Overview

PTE Academic reports an overall score, communicative skills scores and enabling skills scores.



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 does between 70 and 91 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.

Communicative skills scores

The communicative skills measured are **listening**, **reading**, **speaking** and **writing**. Items testing these communicative skills also test specific subskills. For integrated skills items (that is, 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

The enabling skills are used to rate performance in the productive skills of speaking and writing. The enabling skills measured are **grammar**, **oral fluency**, **pronunciation**, **spelling**, **vocabulary**, and **written discourse**. The scores for enabling skills are based on performance on only those items that assess these skills specifically. The score range for each skill is 10–90 points.

The enabling skills reported are described as follows:

Grammar	Correct use of language with respect to word form and word order at the sentence level.
Oral fluency	Smooth, effortless and natural-paced delivery of speech.
Pronunciation	Production of speech sounds in a way that is easily understandable to most regular speakers of the language. Regional or national varieties of English pronunciation are considered correct to the degree that they are easily understandable to most regular speakers of the language.
Spelling	Writing of words according to the spelling rules of the language. All national variations are considered correct, but one spelling convention should be used consistently in a given response.
Vocabulary	Appropriate choice of words used to express meaning, as well as lexical range.
Written discourse	Correct and communicatively efficient production of written language at the textual level. Written discourse skills are represented in the structure of a written text, its internal coherence, logical development and the range of linguistic resources used to express meaning precisely.

Scores for enabling skills are not awarded when responses are inappropriate for the items in either content or form. For example, if an essay task requires the test taker to discuss the environment, but the test taker's response is entirely devoted to the topic of fashion or sport, no score points will be given for the response, and none of the enabling skills be scored for the item.

In relation to form, if a task requires a one-sentence summary of a text and the response consists of a list of words, no score points for the response will be given.

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 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, a score of 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, but 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.

The tables that follow give an overview of how the 20 item types in the three parts of PTE Academic are scored. They also show timings, the number of items in any given test, the communicative skills, enabling skills and other elements scored.

Part 1 Speaking and Writing (approx. 77–93 minutes)				
Item type	Time allowed	Number of items	Scoring	Communicative skills, enabling skills and other traits scored
Read aloud	30-35 minutes	6-7	Partial credit	Reading and speaking Oral fluency, pronunciation Content
Repeat sentence		10-12	Partial credit	Listening and speaking Oral fluency, pronunciation Content
Describe image		6-7	Partial credit	Speaking Oral fluency, pronunciation Content

Part 1 Speaking and Writing (approx. 77–93 minutes)				
Item type	Time allowed	Number of items	Scoring	Communicative skills, enabling skills and other traits scored
Re-tell lecture		3-4	Partial credit	Listening and speaking Oral fluency, pronunciation Content
Answer short question		10-12	Correct/ incorrect	Listening and speaking Vocabulary
Summarize written text	20-30 minutes	2-3	Partial credit	Reading and writing Grammar, vocabulary Content, form
Write essay	20-40 minutes	1-2	Partial credit	Writing Grammar, vocabulary, spelling, written discourse Content; development, structure and coherence; form, general linguistic range

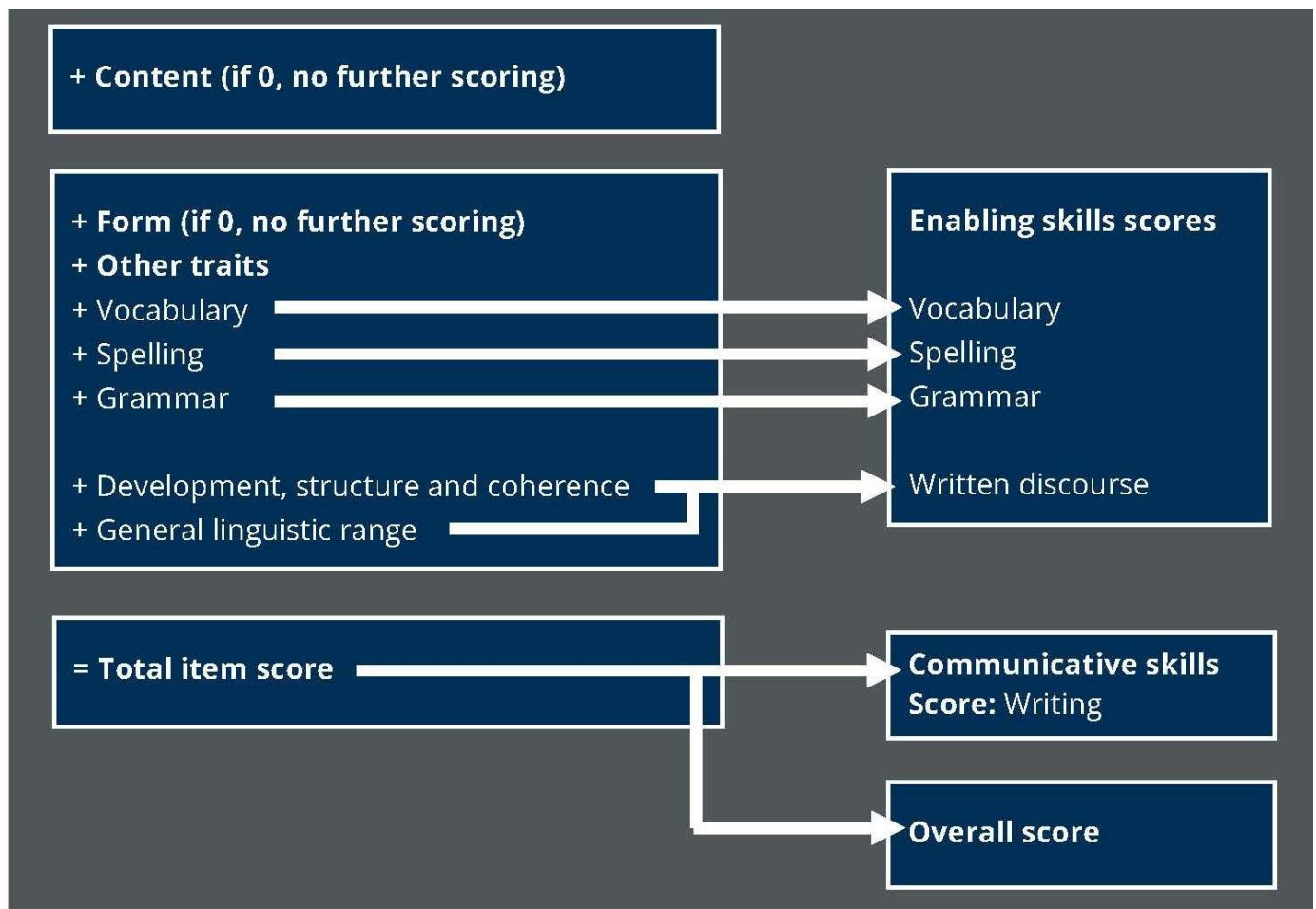
Part 2 Reading (approximately 32–41 minutes)				
Item type	Time allowed	Number of items	Scoring	Communicative skills, enabling skills and other traits scored
Multiple-choice, choose single answer	32-41 minutes	2-3	Correct/ incorrect	Reading
Multiple-choice, choose multiple answers		2-3	Partial credit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

Part 3 Listening (approx. 45–57 minutes)				
Item type	Time allowed	Number of items	Scoring	Communicative skills, enabling skills and other traits scored
Summarize spoken text	20-30 minutes	2-3	Partial credit	Listening and writing Grammar, vocabulary, spelling Content, form
Multiple-choice, choose multiple answers	23-28 minutes	2-3	Partial credit (for each correct response. Points deducted for incorrect options chosen)	Listening
Fill in the blanks		2-3	Partial credit (each correct word spelled correctly)	Listening and writing
Highlight correct summary		2-3	Correct/incorrect	Listening and reading
Multiple-choice, choose single answer		2-3	Correct/incorrect	Listening
Select missing word		2-3	Correct/Incorrect	Listening
Highlight incorrect words		2-3	Partial credit (for each word. Points deducted for incorrect options chosen)	Listening and reading
Write from dictation		3-4	Partial credit (for each word spelled correctly)	Listening and writing

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 scores for content, form and the enabling skills traits (vocabulary, spelling, grammar, development, structure and coherence, and general linguistic range) add up to the **total item score**.

The enabling skills scores awarded for the item contribute to the **enabling skills scores** reported for performance on the entire test, which for this particular item type include vocabulary, spelling, grammar and written discourse.

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. Item Scoring: Skills Tested and Scoring Criteria

Please note: The scoring criteria used by human raters for PTE Academic are given. This serves to give an understanding of what test takers need to demonstrate in their responses. The automated scoring engines are trained on scores given by human raters. The scores indicated for each trait undergo a number of complex calculations to produce the total item score.

Part 1 Speaking and writing

Read aloud

Communicative skills tested: Reading and speaking.

Subskills tested: Identifying a writer's purpose, style, tone or attitude; understanding academic vocabulary; reading a text under timed conditions.

Speaking for a purpose (to repeat, to inform, to explain); reading a text aloud; speaking at a natural rate; producing fluent speech; using correct intonation; using correct pronunciation; using correct stress; speaking under timed conditions.

Look at the text below. In 40 seconds, you must read this text aloud as naturally and clearly as possible. You have 40 seconds to read aloud.

Recorded Answer

Current Status:
Beginning in 40 seconds.

The development of easy-to-use statistical software has changed the way statistics is being taught and learned. Students can make transformations of variables, create graphs of distributions of variables, and select among statistical analyses all at the click of a button. However, even with these advancements, students sometimes still find statistics to be an arduous task.

Scoring

Communicative skills	Reading and speaking
Enabling skills and other traits scored	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 (Detailed criteria on p24)
	Oral fluency: 5 Native-like 4 Advanced 3 Good 2 Intermediate 1 Limited, 0 Disfluent (Detailed criteria on p24)

Repeat sentence

Communicative skills tested: Listening and speaking.

Subskills tested: Understanding academic vocabulary; inferring the meaning of unfamiliar words; comprehending variations in tone, speed and accent.

Speaking for a purpose (to repeat, to inform, to explain); speaking at a natural rate; producing fluent speech; using correct intonation; using correct pronunciation; using correct stress; speaking under timed conditions.

You will hear a sentence. Please repeat the sentence exactly as you hear it. You will hear the sentence only once.

Status: Beginning in 3 seconds.

Volume 

Recorded Answer

Current Status:
Beginning in 8 seconds.

Scoring

Communicative skills	Listening and speaking
Enabling skills and other traits scored	<p>Content:</p> <p>Errors = replacements, omissions and insertions only</p> <p>Hesitations, filled or unfilled pauses, leading or trailing material are ignored in the scoring of content</p> <p>3 All words in the response from the prompt in the correct sequence</p> <p>2 At least 50% of words in the response from the prompt in the correct sequence</p> <p>1 Less than 50% of words in the response from the prompt in the correct sequence</p> <p>0 Almost nothing from the prompt in the response</p>


	Pronunciation: 5 Native-like 4 Advanced 3 Good 2 Intermediate 1 Intrusive 0 Non-English (Detailed criteria on p24)
	Oral fluency: 5 Native-like 4 Advanced 3 Good 2 Intermediate 1 Limited 0 Disfluent (Detailed criteria on p24)

Describe image

Communicative skills tested: Speaking

Subskills tested: Speaking for a purpose (to repeat, inform, explain); supporting an opinion with details, examples and explanations; organizing an oral presentation in a logical way; developing complex ideas within a spoken discourse; using words and phrases appropriate to the context; using correct grammar; speaking at a natural rate; producing fluent speech; using correct intonation; using correct pronunciation; using correct pronunciation; using correct stress; speaking under timed conditions.

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



Recorded Answer

Current Status:
Beginning in 23 seconds.

Scoring

Communicative skills	Speaking
Enabling skills and other traits scored	Content: 5 Describes all elements of the image and their relationships, possible development and conclusion or implications 4 Describes all the key elements of the image and their relations, referring to their implications or conclusions 3 Deals with most key elements of the image and refers to their implications or conclusions 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 1 Describes some basic elements of the image, but does not make clear their interrelations or implications 0 Mentions some disjointed elements of the presentation
	Pronunciation: 5 Native-like

	4 Advanced 3 Good 2 Intermediate 1 Intrusive 0 Non-English (Detailed criteria on p24)
	Oral fluency: 5 Native-like 4 Advanced 3 Good 2 Intermediate 1 Limited 0 Disfluent (Detailed criteria on p24)

Re-tell lecture

Communicative skills tested: Listening and speaking.

Subskills tested: Identifying the topic, theme or main ideas; identifying supporting points or examples; identifying a speaker's purpose, style, tone or attitude; understanding academic vocabulary; inferring the meaning of unfamiliar words; comprehending explicit and implicit information; comprehending concrete and abstract information; classifying and categorizing information; following an oral sequencing of information; comprehending variations in tone, speed and accent.

Speaking for a purpose (to repeat, to inform, to explain); supporting an opinion with details, examples and explanations; organizing an oral presentation in a logical way; developing complex ideas within a spoken discourse; using words and phrases appropriate to the context; using correct grammar; speaking at a natural rate; producing fluent speech; using correct intonation; using correct pronunciation; using correct stress; speaking under timed conditions.

You will hear a lecture. After listening to the lecture, in 10 seconds, please speak into the microphone and retell what you have just heard from the lecture in your own words. You will have 40 seconds to give your response.



Status: Playing

Volume

☐

Recorded Answer

Current Status:
Beginning in 51 seconds.

Scoring

Communicative skills	Listening and speaking
Enabling skills and other traits scored	<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>
	<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>(Detailed criteria on p24)</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>(Detailed criteria on p24)</p>

Answer short question


Communicative skills tested: Listening and speaking.

Subskills tested: Identifying the topic, theme or main ideas; understanding academic vocabulary; inferring the meaning of unfamiliar words.

Speaking for a purpose (to repeat, to inform, to explain); using words and phrases appropriate to the context; speaking under timed conditions

You will hear a question. Please give a simple and short answer. Often just one or a few words is enough.

Status: Playing

Volume 

Recorded Answer

Current Status:
Beginning in 5 seconds.

Scoring

Communicative skills	Listening and speaking
	Correct/incorrect: 1 Appropriate word choice in response 0 Inappropriate word choice in response

Summarize written text

Communicative skills tested: Reading and writing.

Subskills tested: Reading a passage under timed conditions; identifying a writer's purpose, style, tone or attitude; comprehending explicit and implicit information; comprehending concrete and abstract information.

Writing a summary; writing under timed conditions; taking notes while reading a text; synthesizing information; writing to meet strict length requirements; communicating the main points of a reading passage in writing; using words and phrases appropriate to the context; using correct grammar.

Read the passage below and summarize it using one sentence. Type your response in the box at the bottom of the screen. You have 10 minutes to finish this task. Your response will be judged on the quality of your writing and on how well your response presents the key points in the passage.

'Just-in-Time' is a method of manufacturing products which aims to minimise production time, production costs, and the amount of stock held in the factory. Raw materials and supplies arrive at the factory as they are required, and consequently there is very little stock sitting idle at any one time. Each stage of the production process finishes just before the next stage is due to commence and therefore the lead-time is significantly reduced. With a 'just-in-time' production system, the level of production is related to the demand for the output (i.e. the number of orders) rather than simply producing finished goods and waiting for orders. This means that raw materials and stock only need to be ordered from suppliers as required.

Just in time...

Cut

Copy

Paste

Total Word Count: 3

Scoring

Communicative skills	Reading and writing
Enabling skills and other traits scored	Content: 2 Provides a good summary of the text. All relevant aspects mentioned 1 Provides a fair summary of the text but misses one or two aspects 0 Omits or misrepresents the main aspects of the text
	Form: 1 Is written in one, single, complete sentence 0 Not written in one, single, complete sentence or contains fewer than 5 or more than 75 words. Summary is written in capital letters
	Grammar: 2 Has correct grammatical structure

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

Write essay

Communicative skills tested: Writing

Subskills tested: Writing for a purpose (to learn, to inform, to persuade); supporting an opinion with details, examples and explanations; organizing sentences and paragraphs in a logical way; developing complex ideas within a complete essay; using words and phrases appropriate to the context; using correct grammar; using correct spelling; using correct mechanics; writing under timed conditions.

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?

Cut

Copy

Paste

Total Word Count: 0

Scoring

Communicative skills	Writing
Enabling skills and other traits scored	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

	<p>Development, structure and coherence:</p> <p>2 Shows good development and logical structure</p> <p>1 Is incidentally less well structured, and some elements or paragraphs are poorly linked</p> <p>0 Lacks coherence and mainly consists of lists or loose elements</p>
	<p>Grammar:</p> <p>2 Shows consistent grammatical control of complex language. Errors are rare and difficult to spot</p> <p>1 Shows a relatively high degree of grammatical control. No mistakes which would lead to misunderstandings</p> <p>0 Contains mainly simple structures and/or several basic mistakes</p>
Enabling skills and other traits scored	<p>General linguistic range:</p> <p>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</p> <p>1 Sufficient range of language to provide clear descriptions, express viewpoints and develop arguments</p> <p>0 Contains mainly basic language and lacks precision</p>
	<p>Vocabulary range:</p> <p>2 Good command of a broad lexical repertoire, idiomatic expressions and colloquialisms</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>0 Contains mainly basic vocabulary insufficient to deal with the topic at the required level</p>
	<p>Spelling:</p> <p>2 Correct spelling</p> <p>1 One spelling error</p> <p>0 More than one spelling error</p>

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, misordered 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

Part 2 Reading

Multiple-choice, choose single answer

Communicative skills tested: Reading

Subskills tested: Any of the following dependent on the item: Identifying the topic, theme or main ideas; identifying the relationships between sentences and paragraphs; evaluating the quality and usefulness of texts; identifying a writer's purpose, style, tone or attitude; identifying supporting points or examples; reading for overall organization and connections between pieces of information; reading for information to infer meanings or find relationships; identifying specific details, facts, opinions, definitions or sequences of events; inferring the meaning of unfamiliar words.

Parts of Australia's biggest city, Sydney, plunged into near darkness for an hour to raise awareness of global warming. The local government, environmental groups and businesses united in support of "Earth Hour" - when the city turned off the lights to save power and cut emissions that contribute to global warming. Lights on the city's iconic Harbor Bridge were switched off for the hour along with most of those on the famous Sydney Opera House. Tens of thousands of suburban homes joined in. So did hundreds of businesses.

"It's massive [climate change]. It's very difficult to grasp in all its magnitude and complexities," said Australian actress Cate Blanchett. "So I think that the potency of tonight is that it's about a very simple beginning, you know, turning off a switch." Every day millions of lights and computers are left on in deserted offices, apartments and houses. Environmental activists say that simply switching them off could cut Sydney's greenhouse gas emissions by five percent over the next year.

Per capita, Australia is one of the world's largest producers of carbon dioxide and other gases that many scientists believe are helping to warm the Earth's atmosphere, causing climate upset. Prime Minister John Howard has refused to sign the international Kyoto Protocol to halt global warming, saying it does not address key issues between developed and developing nations and hurts Australia's coal-driven economy. But he argues Australia is meeting its international emissions obligations in other ways.

A long-standing drought and serious water shortages in Australia have focused much attention on climate change in this election year. Some experts warn higher temperatures could leave this nation of 20 million people at the mercy of more severe droughts and devastating tropical cyclones.

Read the text and answer the multiple-choice question by selecting the correct response. *Only one response is correct.*

One present indicator of climate change in Australia is _____

- ☐ tropical cyclones.
- ☐ darkness.
- ☐ gas emissions.
- ☐ carbon dioxide.
- ☐ drought.

Scoring

Communicative skills	Reading
	Correct/incorrect: 1 Correct response 0 Incorrect response

Multiple-choice, choose multiple answers

Communicative skills tested: Reading

Subskills tested: Any of the following dependent on the item: Identifying the topic, theme or main ideas; identifying the relationships between sentences and paragraphs; evaluating the quality and usefulness of texts; identifying a writer's purpose, style, tone or attitude; identifying supporting points or examples; reading for overall organization and connections between pieces of information; reading for information to infer meanings or find relationships; identifying specific details, facts, opinions, definitions or sequences of events; inferring the meaning of unfamiliar words.

The Turks and Caicos Islands are a multi-island archipelago at the southern tip of the Bahamas chain, approximately 550 miles south-east of Florida. The islands are an overseas territory of the United Kingdom although they exercise a high degree of local political autonomy. The economy of the islands rests mainly on tourism, with some contribution from offshore banking and fishing.

Primary schooling is divided into eight grades, with most pupils entering at the age of four years and leaving at twelve. After two kindergarten years, Grades 1-6 are covered by a graded curriculum in maths, language and science that increases in difficulty as pupils get older. There is little repetition and pupils are expected to progress through primary school in their age cohorts. At the end of primary schooling, pupils sit an examination that serves to stream them in the secondary school setting. Primary and secondary school enrolment is virtually universal.

There are a total of ten government primary schools on the islands. Of these, seven are large enough to organize pupils into single-grade classrooms. Pupils in these schools are generally grouped by age into mixed ability classes. The remaining three schools, because of their small pupil numbers, operate with multigrade groupings. They serve communities with small populations whose children cannot travel to a neighbouring larger primary school. Pupils in these classes span up to three grade and age groups.

As far as classroom organization is concerned, the multigrade and monograde classrooms are similar in terms of the number of pupils and the general seating arrangements, with pupils in rows facing the blackboard. There is no evidence that the multigrade teachers operate in a particularly resource-poor environment in the Turks and Caicos Islands. This is in contrast to studies conducted in other developing country contexts.

Read the text and answer the question by selecting all the correct responses. *More than one response is correct.*

According to the text, which of the following statements can be concluded about primary classes in the Turks and Caicos Islands?

- ☐ Multigrade classes are mainly found in smaller schools.
- ☐ Most primary pupils are in multigrade classes.
- ☐ Parents can choose to send their child to a multigrade school.
- ☐ Most primary pupils are in mixed ability classes.
- ☐ Multigrade classes are for the youngest three grades.

Scoring

This is the first of three item types in the test where points are deducted for incorrect responses. So if a test taker scores 2 points for two correct options, but then scores -2 for two incorrect options chosen, they will score 0 points overall for the item.

Communicative skills	Reading
	Partial credit, points deducted for incorrect options chosen: 1 Each correct response - 1 Each incorrect response 0 Minimum score

Re-order paragraphs

Communicative skills tested: Reading

Subskills tested: Identifying the topic, theme or main ideas; identifying supporting points or examples; identifying the relationships between sentences and paragraphs; understanding academic vocabulary; understanding the difference between connotation and denotation; inferring the meaning of unfamiliar words; comprehending explicit and implicit information; comprehending concrete and abstract information; classifying and categorizing information; following a logical or chronological sequence of events.

The text boxes in the left panel have been placed in a random order. Restore the original order by dragging the text boxes from the left panel to the right panel.

Unordered Options	Correct Answer
He convinced Professor Fitzgerald of the University of Hull to set up a study into this matter.	
Professor Fitzgerald and his team studied more than 47000 women.	
The women were asked to fill in a questionnaire about their diet and about their suffering from acne.	
No link was found between acne and traditionally suspect food such as chocolate and chips.	
Doctor Byron has long held that there is a link between diet and acne.	
<div>← →</div> <div><input type="checkbox"/> Key</div>	

↑
↓

Scoring

Communicative skills	Reading
	Partial credit: 1 Each pair of correct adjacent textboxes 0 Minimum score

Reading: Fill in the blanks

Communicative skills tested: Reading

Subskills tested: Identifying the topic, theme or main ideas; identifying words and phrases appropriate to the context; understanding academic vocabulary; understanding the difference between connotation and denotation; inferring the meaning of unfamiliar words; comprehending explicit and implicit information; comprehending concrete and abstract information; following a logical or chronological sequence of events.

In the text below some words are missing. Drag words from the box below to the appropriate place in the text. To undo an answer choice, drag the word back to the box below the text.

Science blogs serve a dual purpose. First, they connect scientists to each other, as modern day intellectual salons. Even scientific papers are now beginning to blogs as references. Second, they connect scientists to the general , offering a behind-the-scenes at how science progresses.

public

formal

look

view

world

cite

prescribed

serving

Scoring

Communicative skills	Reading
	Partial credit: 1 Each correctly completed blank 0 Minimum score

Reading and writing: Fill in the blanks

Communicative skills tested: Reading and writing.

Subskills tested: Identifying the topic, theme or main ideas; identifying words and phrases appropriate to the context; understanding academic vocabulary; understanding the difference between connotation and denotation; inferring the meaning of unfamiliar words; comprehending explicit and implicit information; comprehending concrete and abstract information; following a logical or chronological sequence of events.

Using words and phrases appropriate to the context; using correct grammar.

Below is a text with blanks. Click on each blank, a list of choices will appear. Select the appropriate answer choice for each blank.

The impact of a product recall can be wide-reaching, and will almost certainly extend beyond the purely financial

of taking products off shelves. David Palmer, director of product risk management for Aon Limited, points out, many years to establish a strong brand name and and it only takes one product recall to bring . Despite heightened awareness around product recalls, and significant investment in product control, some falling when it comes to implementing strategies to deal with a crisis.

observations
examinations
consultations
considerations

Scoring

Communicative skills	Reading and writing
	Partial credit: 1 Each correctly completed blank 0 Minimum score

Scoring

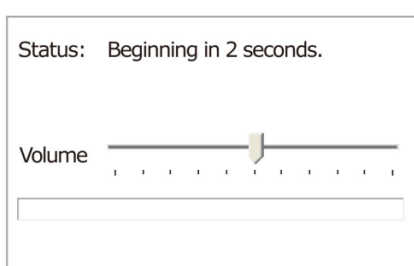
Communicative skills	Listening and writing
Enabling skills and other traits scored	Content: 2 Provides a good summary of the text. All relevant aspects are mentioned 1 Provides a fair summary of the text, but one or two aspects are missing 0 Omits or misrepresents the main aspects
Enabling skills and other traits scored	Form: 2 Contains 50-70 words 1 Contains 40-49 words or 71-100 words 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 Grammar: 2 Correct grammatical structures 1 Contains grammatical errors with no hindrance to communication 0 Defective grammatical structures which could hinder communication Vocabulary: 2 Appropriate choice of words 1 Some lexical errors but with no hindrance to communication 0 Defective word choice which could hinder communication Spelling: 2 Correct spelling 1 One spelling error 0 More than one spelling error

Multiple-choice, choose multiple answers

Communicative skills tested: Listening

Subskills tested: Any of the following dependent on the item: Identifying the topic, theme or main ideas; identifying supporting points or examples; Identifying specific details, facts, opinions, definitions or sequences of events; identifying a speaker's purpose, style, tone or attitude; identifying the overall organization of information and connections between pieces of information; inferring the context, purpose or tone; inferring the meaning of unfamiliar words; predicting how a speaker may continue.

Listen to the recording and answer the question by selecting all the correct responses. You will need to select more than one response.



What happened as a result of the arrival of colonists?

- ☐ Huge areas were deforested.
- ☐ Old farming methods were abandoned.
- ☐ Large expanses were planted with new species.
- ☐ Stone quarries were depleted.
- ☐ Sections of land were delineated with stones.

Scoring

This is the second of three item types where points are deducted for incorrect options chosen. So if a test taker scores 2 points for two correct options, but then scores -2 for two incorrect options chosen, they will score 0 points overall for the item.

Communicative skills	Listening
	Partial credit, points deducted for incorrect options chosen: 1 Each correct response - 1 Each incorrect response 0 Minimum score

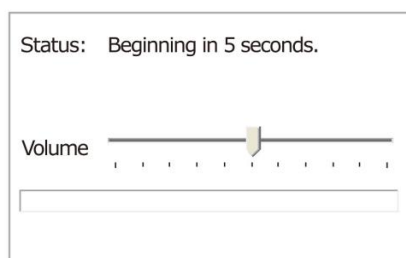
Fill in the blanks

Communicative skills tested: Listening and writing

Subskills tested: Identifying words and phrases appropriate to the context; understanding academic vocabulary; comprehending explicit and implicit information; following an oral sequencing of information.

Writing from dictation; using words and phrases appropriate to the context; using correct grammar; using correct spelling.

You will hear a recording. Type the missing words in each blank.



The Indonesian island of Sumatra is home to the country's fourth-largest city, Medan, a bustling [] hub that's also a jumping-off point for those who want to see one of the world's richest, yet least-known, forests, the Leuser [] in the north of the island, mostly in the province of Aceh; and the quickest way to see Leuser is by air. Mike Griffiths is a former oil company executive turned conservationist who sees Leuser as his second home, and he heads the conservation group. The Leuser International Foundation. He rattles off the names of volcanoes, rivers and other [] as our plane climbs from Medan over the [] green jungle below.

Scoring

Communicative skills	Listening and writing
	Partial credit: 1 Each correct word spelled correctly 0 Minimum score

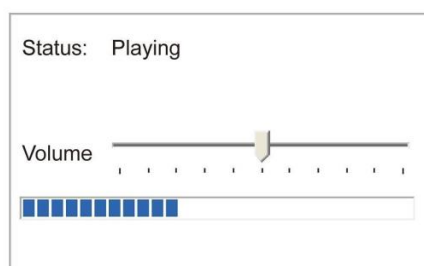
Highlight correct summary

Communicative skills tested: Listening and reading.

Subskills tested: Identifying the topic, theme or main ideas; identifying supporting points or examples; understanding academic vocabulary; inferring the meaning of unfamiliar words; comprehending explicit and implicit information; comprehending concrete and abstract information; classifying and categorizing information; following an oral sequencing of information; comprehending variations in tone, speed and accent.

Identifying supporting points or examples; identifying the most accurate summary; understanding academic vocabulary; inferring the meaning of unfamiliar words; comprehending concrete and abstract information; classifying and categorizing information; following a logical or chronological sequence of events; evaluating the quality and usefulness of texts.

You will hear a recording. Click on the paragraph that best relates to the recording.



- There have been previous lectures in this series covering aspects of abnormal behaviour. This lecture covers the models of motivation applied to normal people. Although there are numerous models of human motivation, two models will be described and compared: the homeostatic model and Maslow's model, with the main focus on the homeostatic model.
- The course has covered the motivation of abnormal people and will look at many different models of motivation and criticize them. This lecture is the first in a series of lectures on the motivation of normal people. Students need to have attended previous lectures to follow the arguments about the motivation of normal people.
- The lecture will pay equal attention to two well-known models of motivation: Maslow's model and the homeostatic model. These models are very similar both starting with basic needs like the need for food. The lecture will emphasise the importance of how the question 'Why?' is at the heart of all studies of motivation.
- The lecture will describe what drives people and scientifically describe the hunger drive. In particular, the homeostatic model of motivation will be examined in relation to the way this model is based on our perception of ourselves as individuals. Examples will be given of the lowest and highest levels of human motivation.

Scoring

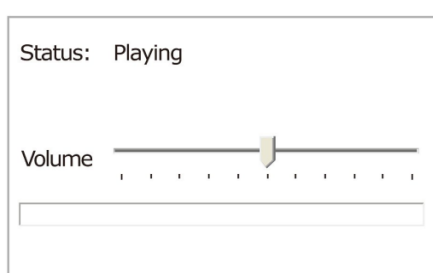
Communicative Skills	Listening and reading
	Correct/incorrect: 1 Correct response 0 Incorrect response

Multiple-choice, choose single answer

Communicative skills tested: Listening

Subskills tested: Any of the following dependent on the item: Any of the following dependent on the item: Identifying the topic, theme or main ideas; identifying supporting points or examples; Identifying specific details, facts, opinions, definitions or sequences of events; identifying a speaker's purpose, style, tone or attitude; identifying the overall organization of information and connections between pieces of information; inferring the context, purpose or tone; inferring the meaning of unfamiliar words; predicting how a speaker may continue.

Listen to the recording and answer the multiple-choice question by selecting the correct response. Only one response is correct.



What does the speaker believe is likely to happen in the future?

- ☐ The media will be owned by a small number of corporations.
- ☐ The public will eventually pay for the funding of the media.
- ☐ The media will be run primarily for the purpose of making profit.
- ☐ Most media analysts will continue to report what they see.

Scoring

Communicative Skills	Listening
	Correct/incorrect: 1 Correct response 0 Incorrect response

Select missing word

Communicative skills tested: Listening

Subskills tested: Identifying the topic, theme or main ideas; identifying words and phrases appropriate to the context; understanding academic vocabulary; inferring the meaning of unfamiliar words; comprehending explicit and implicit information; comprehending concrete and abstract information; following an oral sequencing of information; predicting how a speaker may continue; forming a conclusion from what a speaker says; comprehending variations in tone, speed and accent.

You will hear a recording about medicine. *At the end of the recording the last word or group of words has been replaced by a beep.* Select the correct option to complete the recording.

Status: Beginning in 7 seconds.

Volume



- ☐ they become dedicated
- ☐ they get exhausted
- ☐ they're rather improved
- ☐ they're quite competent
- ☐ they're quite indifferent

Scoring

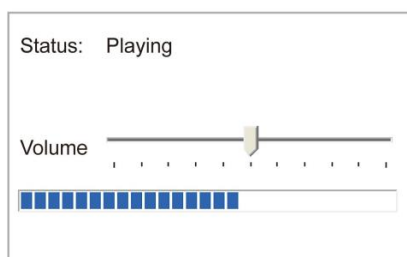
Communicative skills	Listening
	Correct/incorrect: 1 Correct response 0 Incorrect response

Highlight incorrect words

Communicative skills tested: Listening and reading

Subskills tested: Identifying errors in a transcription; understanding academic vocabulary; following an oral sequencing of information; comprehending variations in tone, speed and accent; understanding academic vocabulary; following a logical or chronological sequence of events; reading a text under timed conditions; matching written text to speech.

You will hear a recording. Below is a transcription of the recording. Some words in the transcription differ from what the speaker(s) said. Please click on the words that are different.



When explorer Roald Amundsen set out to find the Northwest Passage, his official mission was scientific - a search for the magnetic south pole. But as historian Roland Huntford describes, the real drive behind the expedition came from a deep desire to colonize the unknown. "One of the reasons that Amundsen would have been challenged by the Northwest Passage is simply that it was one of the last great geographical goals accomplished. What you have to realize is that by the 1880s, most of the earth had been discovered."

Scoring

This is the third of three item types where points are deducted for incorrect options chosen. So if a test taker scores 2 points for two correct options, but then scores -2 for two incorrect options chosen, they will score 0 points overall for the item.

Communicative Skills	Listening and reading
	Partial credit, points deducted for incorrect options chosen: 1 Each correct word - 1 Each incorrect word 0 Minimum score


Write from dictation

Communicative skills tested: Listening and writing.

Subskills tested: Understanding academic vocabulary; following an oral sequencing of information; comprehending variations in tone, speed and accent; writing from dictation; using correct spelling.

You will hear a sentence. Type the sentence in the box below exactly as you hear it. Write as much of the sentence as you can. You will hear the sentence only once.

Status: Beginning in 4 seconds.

Volume 

Total Word Count: 0

Scoring

Communicative skills	Listening and writing
	Partial credit: 1 Each correct word spelled correctly 0 Each incorrect or misspelled word

4. Using PTE Academic Scores

PTE Academic uses 20 item types, reflecting different modes of language use and requiring different response tasks and formats. All items in PTE Academic are machine scored. Scores on a number of item types are based on correctness only, while scores on other item types requiring spoken or written responses are based, in addition to correctness, on formal aspects (e.g., number of words) and the quality of the response. The quality of the responses is reflected on the PTE Academic score report in the enabling skills: grammar, oral fluency, pronunciation, spelling, vocabulary and written discourse.

How institutions 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 provides 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 discrete information about the listening, reading, speaking and writing skills of a test taker. These skills are also scored between 10 and 90 points.

Enabling skills scores

The enabling skills scores are also provided within the PTE Academic score report. They provide information about particular strengths and weaknesses of a test taker's ability to communicate in speaking or writing. This information may be useful to determine the type of further English study and coursework required to improve a test taker's English language ability. The enabling skills scores should not be used when making admissions decisions because the 'measurement error' is too large. This is discussed in the 'Error of measurement' section on page [48](#).

A definition of the enabling skills is given in the table below:

Enabling Skills	Definition
Grammar	Correct use of language with respect to word form and word order at the sentence level
Oral fluency	Smooth, effortless and natural-paced delivery of speech
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
Spelling	Writing of words according to the spelling rules of the language. All national variations in spelling are considered correct
Vocabulary	Appropriate choice of words used to express meaning precisely in written and spoken English, as well as lexical range
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

Definition of enabling skills

Alignment with CEF

To ensure comparability and interpretability of test scores, PTE Academic has been aligned to the CEF, which is recognized as a standard across Europe and in many countries outside of Europe. In the USA, the National Council of State Supervisors for Languages (NCSSFL) has introduced the use of the LinguaFolio Self-Assessment Grid (NCSSFL, 2008), which relates language levels to the scales of both the ACTFL (American Council on the Teaching of Foreign Languages) and the CEF.

The CEF includes a set of consecutive language levels defined by descriptors of language competencies. The six-level framework was developed by the Council of Europe (2001) to enable language learners, teachers, universities or potential employers to compare and relate language qualifications by level.

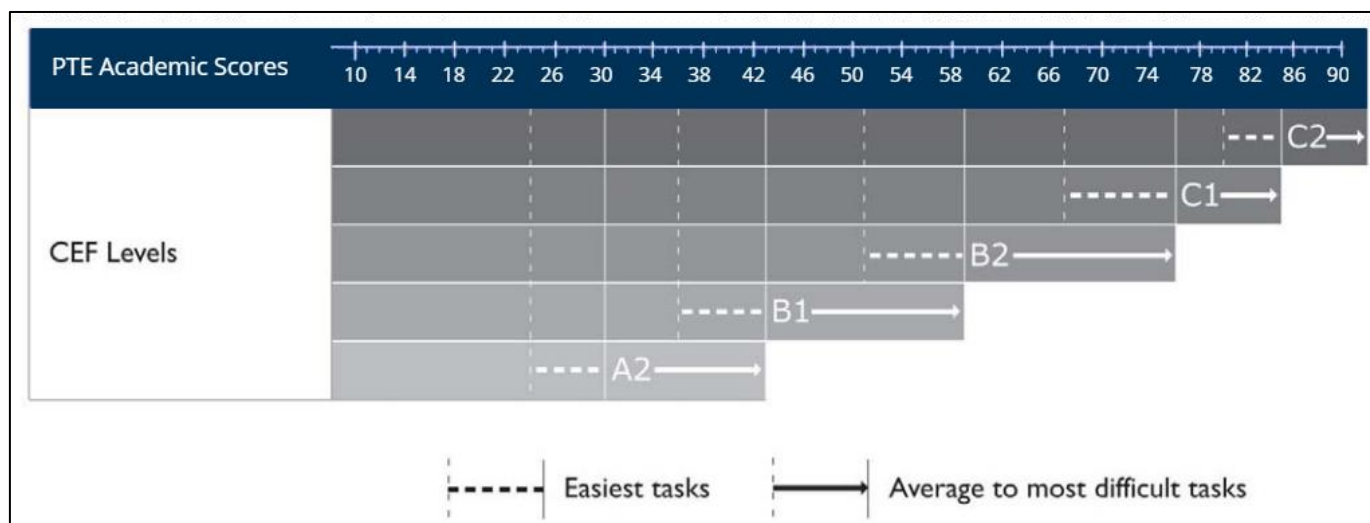
Alignment of PTE Academic to the CEF levels provides a means to interpret PTE Academic scores in terms of the level descriptors of the CEF. As these descriptors focus on what an English language learner can do, scores that are properly aligned to the CEF give educators and institutions more relevant information about a test taker's ability.

The PTE Academic Score Scale and the CEF

The explanation of the alignment of PTE Academic to the CEF is that to stand a reasonable chance at successfully performing any of the tasks defined at a particular CEF 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 CEF 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 CEF 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 CEF

What PTE Academic scores mean

PTE Academic alignment with the CEF 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 CEF 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 CEF levels.

PTE Academic scores predicting the likelihood of successful performance on CEF level tasks			
CEF Level	Easiest	Average	Most Difficult
C2	80	85	NA
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.

The table under [PTE Academic Requirements](#) (below) shows what PTE Academic scores in the range from A1 to C1 mean. The table includes shaded score ranges that predict some degree of performance at the next higher level, and it describes what a test taker is likely to be able to do within those score ranges.

PTE Academic Requirements

A score of **at least 36** is required for UKBA tier 4 student visas for students wanting to study on a course below degree level.

A score of **at least 51** is required for UKBA tier 4 student visas for students wanting to study on a course at or above degree level at an institution that is not a UK Higher Education Institution.

If students wish to study at degree level or above at a UK Higher Education Institution, then it is the university that decides on the score required. Our experience suggests that most universities require:

- for **undergraduate studies** a minimum score **between 51 and 61**
- for **postgraduate studies** a minimum score **between 57 and 67**
- for **MBA studies** a minimum score **between 59 and 69**

PTE Academic Score	Common European Framework Level	Level Descriptor ¹	What does this mean for a score user?
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.
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	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

¹ © 2001 The copyright of the level descriptors reproduced in this document belongs to the Council of Europe.

PTE Academic Score	Common European Framework Level	Level Descriptor ¹	What does this mean for a score user?
		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.	education, including both coursework and student life.
51 - 58	Scores in this range predict success on the easiest tasks at B2	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.	
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?
36 - 42	Scores in this range predict success on the easiest tasks at B1	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.	
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.	A1 is an insufficient level for academic level participation.

PTE Academic Score	Common European Framework Level	Level Descriptor ¹	What does this mean for a score user?
		Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.	

PTE A scores, CEF level descriptors and what scores mean

Error of measurement

Tests aim to provide a measure of ability. PTE Academic measures the ability 'to use English in academic settings'. Obviously, measures of a test taker's English language abilities will vary; some candidates will have higher scores than others. The degree to which scores among test takers vary is the 'score variance'. The purpose of testing is to measure 'true variance' in ability among students, but all measurement contains some error.

The degree to which the score variance is due to error is called the 'error of measurement'. The remainder of the variance is due to 'true variance' in ability among test takers. The error of measurement is related to the reliability of the test: a smaller measurement error means higher reliability of test scores.

The error of measurement can be interpreted as follows: the true score of a test taker is within a range of scores around the reported score. The size of that range is defined by the error of measurement. For example, if the reported score is 60 and the error of measurement is 3, then the true score, with 68% certainty, is within one measurement error from the reported score; that is within the range of 57 (60-3) and 63 (60+3). The true score, with 95% certainty, is within twice the measurement error; that is within the range of 54 (60-2x3) to 66 (60+2x3).

Overall score and communicative skills scores

There are two main approaches to estimating the error of measurement. In Classical Test Theory (CTT) the reliability estimate is assumed to apply to any score on a test, irrespective of whether the score is low, medium or high. Therefore, the error of measurement is assumed to be the same size anywhere on the test's score scale. That is why in CTT we speak of the Standard Error of Measurement (SEM). Many test providers report the SEM and for PTE Academic this is 2.32. This figure is based on test data from 30,000 test takers.

An alternative approach to estimating the error of measurement is used in modern test theory, commonly referred to as Item Response Theory (IRT). IRT recognizes that the reliability of a test is not uniform across an entire score scale. Tests tend to be less reliable towards the extreme low and high score ranges. Consequently, the size of the error of

measurement tends to be larger towards these extreme scores. The size of the error is therefore conditional on the score and so in IRT we speak of Conditional Errors of Measurement (CEM).

The table below shows the average size of the CEM at five levels (A2 to C2) on the CEF for the overall score and for the communicative skills scores that are provided on the PTE Academic score report. The size of the error at each score point is estimated by averaging scores across a random sample of 100 test forms from the PTE Academic item bank.

PTE Academic Scores		Average Measurement Error				
		A2	B1	B2	C1	C2
Overall		2.5	2.4	2.7	3.2	3.5
Communicative skills	Listening	3.7	3.4	3.8	4.4	4.9
	Reading	3.9	4.0	4.4	5.2	5.8
	Speaking	3.6	3.9	4.4	5.1	5.6
	Writing	4.3	3.7	4.1	4.8	5.3

Measurement error for overall score and communicative skills scores at levels A2 to C2

Enabling skills scores

The error on the enabling skills scores is too large to justify use in high-stakes decision-making. The table on the next page shows the average error in score points for the enabling skills.

PTE Academic Scores	Average Measurement Error				
Enabling skills	A2	B1	B2	C1	C2
Grammar	20.7	21.6	20.5	18.7	17.8
Oral fluency	6.5	6.1	6.0	6.1	6.3
Pronunciation	6.4	6.5	6.3	6.3	6.4
Spelling	18.2	18.7	14.9	14.5	15.7
Vocabulary	10.9	10.7	10.8	11.4	12.3
Written discourse	28.5	29.6	28.1	26.6	26.6

Measurement error for enabling skills scores at levels A2 to C2

Test reliability

Directly related to measurement error is test reliability, which is another way of expressing the likelihood that test results will be the same when a test is taken again under the same

conditions, and therefore how accurately a reported test score reflects the true ability of the test taker.

Reliability is expressed as a number between 0 and 1, where 0 means no reliability at all and 1 means perfectly reliable. For tests that are used to make important decisions, high reliability (0.90 or higher) is required. The table below provides the reliability estimates of the overall score and the communicative skills scores within the PTE Academic score range of 53 to 79, which is the most relevant range for admission decisions. For further information on the reliability of PTE Academic, refer to the paper *Establishing Construct and Concurrent Validity of Pearson Test of English Academic*, available at pearsonpte.com/organisations/researchers/research-notes/

Score	Overall	Listening	Reading	Speaking	Writing
Reliability	0.97	0.92	0.91	0.91	0.91

Reliability estimates for scores in the range 53–79

5. Estimates of Concordance between PTE Academic, TOEFL and IELTS

Test comparisons using field test data

PTE Academic has been field tested using over 10,400 test takers. Field testing took place in 2007 and 2008. Test takers were representative of the global population of students seeking admission to universities and other tertiary education institutions where English is the language of instruction. Test takers were born in 158 different countries and spoke 126 different languages.

During the field tests several sets of secondary data were collected. Among these were ratings for all test takers on descriptive scales published by the Council of Europe (2001). In addition, a number of test takers reported their scores on other tests of English, including TOEIC, TOEFL PBT, TOEFL CBT, TOEFL iBT and IELTS.

A limited number of the self-reported data were invalid as the reported scores were outside the possible score range for the particular test. A small number of the test takers also submitted copies of their official score reports on the tests, for which they had provided self-reported data. The table below shows the following for each test: the numbers of self-reported data, how many of these were valid, the mean self-reported scores, the number of official score reports sent in, the mean official scores and the correlations with the PTE Academic field test scores. All correlations are significant at $p < .01^2$.

Test	Self-Reported Data				Official Score Report		
	N Total	N Valid	Mean	Correlation	n	Mean	Correlation
TOEIC	328	327	831.5	0.76	No data	-	-
TOEFL PBT	96	92	572.3	0.64	No data	-	-
TOEFL CBT	110	107	240.5	0.46	No data	-	-
TOEFL iBT	144	140	92.9	0.75	19	92.1	0.95
IELTS	2436	2432	6.49	0.76	169	6.61	0.73

PTE Academic field tests: test takers on other tests of English

From the table, it can be concluded that the self-reported scores are, in general, quite accurate. Indeed, the correlation between the self-reported results and the official score reports was .82 for TOEFL iBT and .89 for IELTS. This finding is in agreement with earlier research on self-reported data. For example, Cassady (2001) found students' self-reported Grade Point Average (GPA) scores to be 'remarkably similar' to official records. The data are also consistent. According to ETS (2005, p.7) the score range 75–95 on TOEFL iBT is

²Significant at $p < .01$ means there is less than 1% chance to observe this correlation if the measures are not related.

comparable to the score range 213–240 on TOEFL CBT and to the score range 550–587 on TOEFL PBT. The mean self-reported scores in the table for these three tests are therefore comparable.

In addition, according to ETS (2001, p.3) a score range of 800–850 on TOEIC corresponds to a score range of 569–588 on TOEFL PBT, which makes the self-reported TOEIC mean score of the test takers on the PTE Academic field test also fall in line with data published by ETS. Based on the data presented in the table, concordance between PTE Academic and other tests of English can be estimated, taking into account a less than optimal effort of test takers during field testing where test results have no direct relevance to the test takers.

Information on concordances since the launch of PTE Academic

At the time of the launch of PTE Academic in November 2009 we presented concordance of PTE Academic with other measures of English as ‘preliminary’. Since then additional information has become available supporting our preliminary estimates. This new information comes from:

- the tens of thousands of test takers who have taken PTE Academic annually since launch
- the use of test scores by thousands of tertiary education institutions
- additional concordance data gathered via surveys
- publications by third parties

Relation to the Common European Framework

The relation of the PTE Academic score scale with the descriptive scale of the Common European Framework of Reference for Languages (CEF) is based on both an item-centered and a test taker-centered method. For the item-centered method, the CEF level of all items was estimated by item writers, reviewed and, if necessary, adapted in the item-reviewing process. For the test taker-centered method, three extended responses (one written and two spoken) per test taker were each rated by two independent, trained raters. If there was a disagreement between the two independent raters, a third rating was gathered and the two closest ratings were retained. A dataset of over 26,000 ratings (by test takers self-reporting, by items and by raters) on up to 100 different items was analyzed using the computer program FACETS (Linacre, 1988; 2005). Estimates of the lower boundaries of the CEF levels, based on the item-centered method, correlated at .996 with those based on the test taker-centered method, which effectively means that the two methods yielded the same results except for less than 1% of error variance.

Validity check using BETA testing data

In addition to the initial field testing of 10,400 students during 2007–08, a further 364 test takers participated in the 2009 BETA testing of PTE Academic. The concordance between the score scale of PTE Academic and the score scales of TOEFL iBT and IELTS (each estimated from the field test data) were used as predictors of TOEFL iBT and IELTS scores of test takers participating in BETA testing. Test takers provided self-reported scores and a smaller, partially overlapping, number of test takers sent in copies of their official score reports.

The table below shows the mean scores as self-reported and from the official score reports; the mean scores for the same test takers as predicted from their PTE Academic score and the correlations between the reported scores and the predictions from PTE Academic. All correlations are significant at $p < .01^3$. It can be concluded that this concordance produces fairly accurate and coherent predictions.

Test	Self-Reported Data				Official Score Report			
	n	Mean	Predicted	Correlation	n	Mean	Predicted	Correlation
TOEFL iBT	42	98.9	97.3	0.75	13	92.2	98.2	0.77
IELTS	57	6.80	6.75	0.73	15	6.60	6.51	0.83

PTE Academic BETA: test takers on other tests of English

Concordance of PTE Academic with other measures of English

Based on the research described, Pearson has produced concordance tables. The table on page [54](#) shows Pearson's current best estimate of concordance between PTE Academic scores and the CEF. In addition, shaded score ranges indicate the PTE Academic scores that predict some degree of performance at the next CEF level.

The table on page [57](#) shows the relation between scores on TOEFL iBT and PTE Academic. The table on page [58](#) shows the relation between scores on IELTS and PTE Academic.

It must be noted 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.

³ Significant at $p < .01$ means there is less than 1% chance to observe this correlation if the measures are not related.

Estimates of concordance between PTE Academic and the descriptive scale of the CEF

PTE Academic Score	Common European Framework Level	Level Descriptor ⁴	What does this mean for a score user?
>85	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.
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	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.

⁴ © The copyright of the level descriptors reproduced in this document belongs to the Council of Europe.

PTE Academic Score	Common European Framework Level	Level Descriptor ⁴	What does this mean for a score user?
		explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.	
51 – 58	Scores in this range predict success on the easiest tasks at B2	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.	
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.
36 – 42	Scores in this range predict success on the easiest tasks at B1	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	

PTE Academic Score	Common European Framework Level	Level Descriptor ⁴	What does this mean for a score user?
		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.	
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.

Estimates of concordance between PTE Academic and TOEFL iBT

TOEFL iBT Score	PTE A Score
No data	85 - 90
120	84
119	83
118	82
117	81
115-116	80
114	79
113	78
112	77
110-111	76
109	75
107-108	74
106	73
105	72
103-104	71
102	70
101	69
99-100	68
98	67
97	66
95-96	65
94	64
93	63
91-92	62

TOEFL iBT Score	PTE A Score
90	61
89	60
87-88	59
86	58
85	57
83-84	56
82	55
81	54
79-80	53
78	52
76-77	51
74-75	50
72-73	49
70-71	48
67-69	47
65-66	46
63-64	45
60-62	44
57-59	43
54-56	42
52-53	41
48-51	40
45-47	39
40-44	38
No data	10 - 37

Estimates of concordance between PTE Academic and IELTS

IELTS Score	PTE A Score
9.0	86 - 90
8.5	83 - 85
8.0	79 - 82
7.5	73 - 78
7.0	65 - 72
6.5	58 - 64
6.0	50 - 57
5.5	42 - 49
5.0	36 - 41
4.5	29 - 35
No data	10 - 28

6. Scored Samples

Automated scoring

As the worldwide leader in publishing and assessment for education, Pearson is using several of its 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. Test 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.

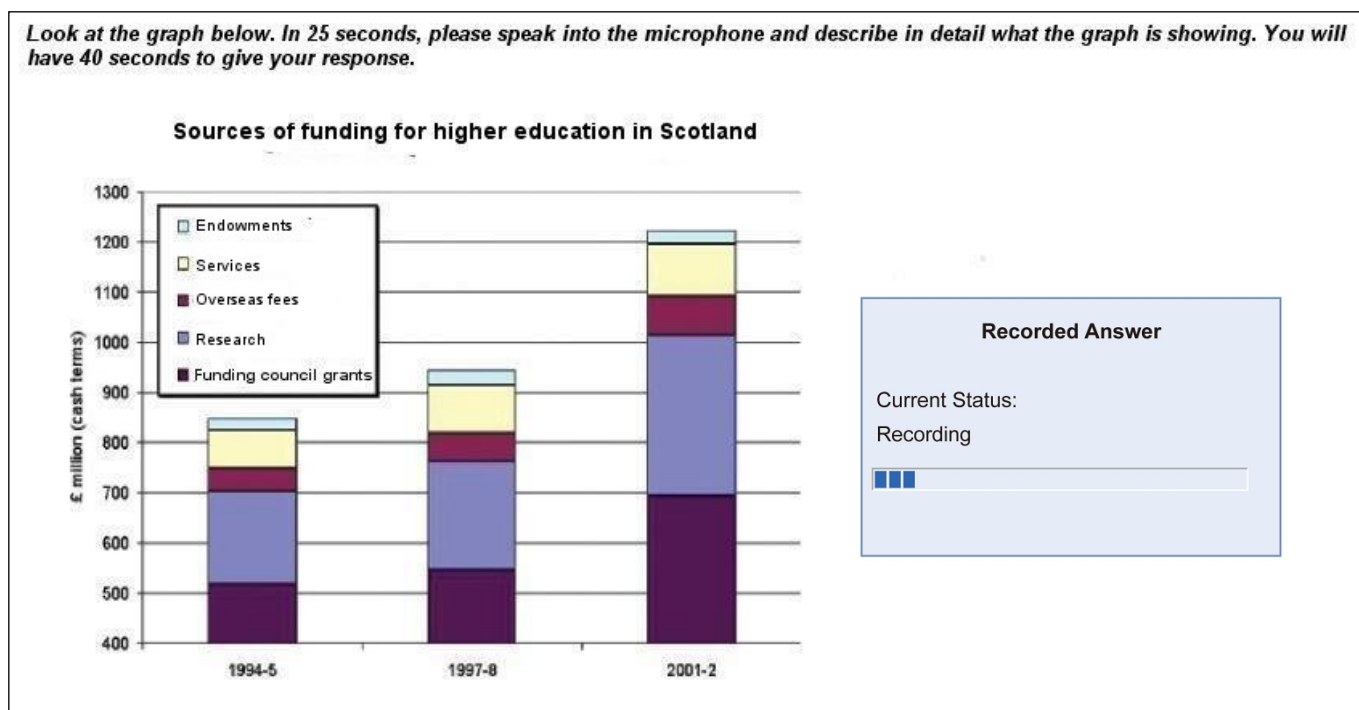
This, in effect, is how the machine is trained to score, only instead of one expert teaching the trainee, there are many expert scorers feeding scores into the system for each response, and instead of a few dozen test takers, the system is trained on thousands of responses from hundreds of test takers. 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. Further information about automated scoring is available on our website www.pearsonpte.com/organisations/teachers-teaching-resources/scoring/

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



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	+	+
Oral fluency	5	+	+
Pronunciation	5	+	+
Maximum item score	15	15	15

These traits are scored as follows:

Content	Pronunciation	Oral fluency
5: Describes all elements of the image and their relationships, possible development and conclusion or implications	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	5 Native-like: Speech shows smooth, rhythm and phrasing. There are no hesitations, repetitions, false starts or non-native phonological simplifications
4: Describes all the key elements of the image and their relations, referring to their implications or conclusions	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	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: Deals with most key elements of the image and refers to their implications or conclusions	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	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: 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	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	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: Describes some basic elements of the image, but does not make clear their interrelations or implications	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-	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

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
0: Mentions some disjointed elements of the presentation	0 Non-English: Pronunciation seems completely characteristic of another language. Many consonants and vowels are mispronounced, misordered 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	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

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

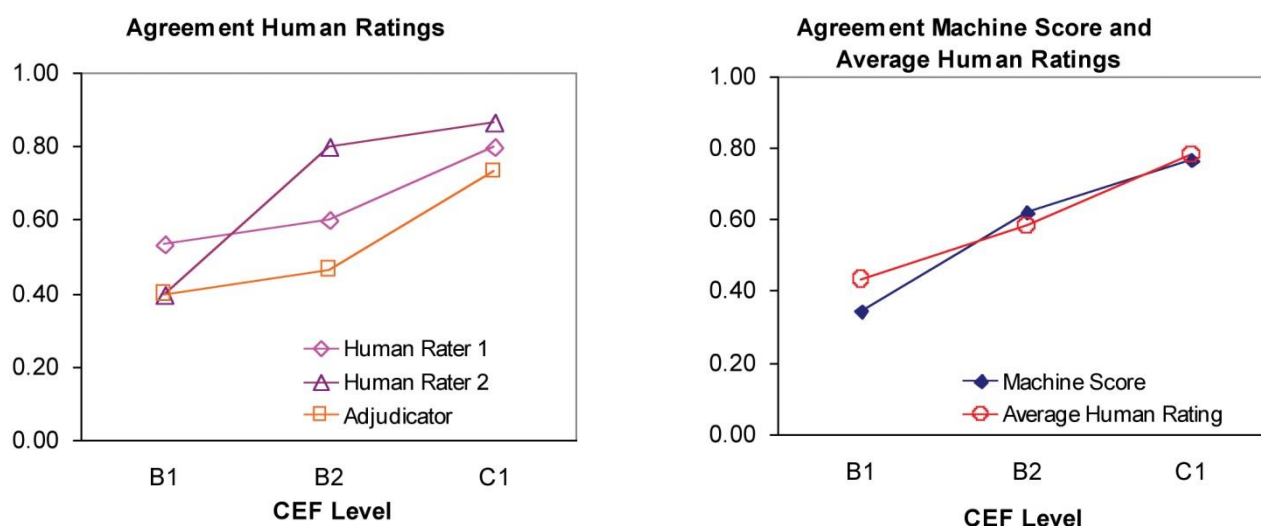
The test taker discusses the major aspects of the graph and the relationship between elements. The response is spoken at a fluent rate and language use is appropriate. There are few grammatical errors in the response. The candidate demonstrates a wide range of vocabulary. Stress is appropriately placed.

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	+	+
Form	2		+
Development, structure and coherence	2	+	+
Grammar	2	+	+
General linguistic range	2	+	+
Vocabulary range	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
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	2: Good command of a broad lexical repertoire, idiomatic expressions and colloquialisms	2: Correct spelling
1: Sufficient range of language to provide clear descriptions, express viewpoints and develop arguments	1: Shows a good range of vocabulary for matters connected to general academic topics. Lexical shortcomings lead to circumlocution or some imprecision	1: One spelling error
0: Contains mainly basic language and lacks precision	0: Contains mainly basic vocabulary insufficient to deal with the topic at the required level	0: More than one spelling error

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 is appropriate. There are very few grammar errors. Spelling is excellent.

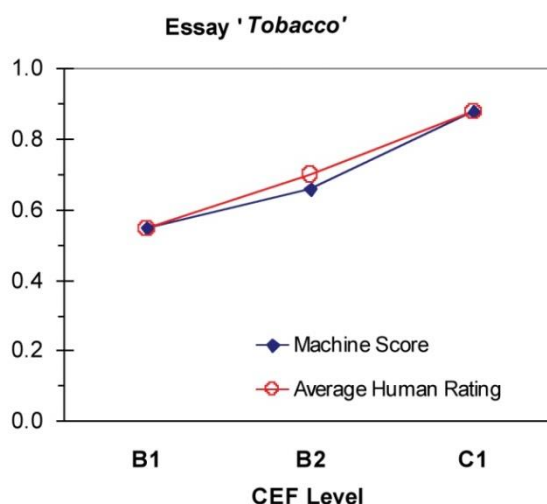
How the response was scored

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

7. References

Using PTE Academic scores

American Council for the Teaching of Foreign Languages (1986) ACTFL Proficiency Guidelines. Hastings-on-Hudson, NY

American Council for the Teaching of Foreign Languages (1999) ACTFL Proficiency Guidelines Speaking, (Revised), actfl.org/files/public/Guidelinespeak.pdf (retrieved 2009-08-08)

American Council for the Teaching of Foreign Languages (2001) ACTFL Proficiency Guidelines Writing, Revised actfl.org/files/public/writingguidelines.pdf (retrieved 2009-08-08)

Council of Europe (2001) Common European Framework of Reference for Languages: Learning, Teaching Assessment Cambridge: CUP

National Council of State Supervisors for Languages (2008) Linguafolio Self-Assessment Grid, ncssfl.org/links/LFGrid.pdf (retrieved 2009-08-08)

Concordance to other tests

Cassady, Jerrell C. (2001) Self-Reported GPA and SAT Scores. ERIC Digest. ERIC Identifier: ED458216

Council of Europe (2001) Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Cambridge: CUP

ETS (2001) TOEFL Institutional Testing Program (ITP) and TOEIC Institutional Program (IP): Two On-Site Testing Tools from ETS at a Glance. Handout Berlin Conference 2001. Princeton: Educational Testing Service

ETS (2005) TOEFL ® Internet-based test: Score comparison tables. Princeton: Educational Testing Service

Linacre, J.M (1988; 2005) A Computer Program for the Analysis of Multi-Faceted Data. Chicago, IL: Mesa Press