

Rethinking English Language Certification

New Approaches to the Assessment of English as an Academic Lingua Franca

David Newbold

3 A Critique of the Sample Material on the TOEFL, IELTS and PTE Websites

Abstract This chapter provides an in-depth look at the free sample material posted on their websites by the examining boards of their main academic exams, and highlights a number of problems. Firstly, much of the material presented is fragmented, (whereas, typically, complete practice tests are only available to be purchased). Secondly, some of the material appears to be outdated or problematic for other reasons, with a number of poor items or inadvertent language errors, which we note. This is surprising, given the desire to project a positive image of tests which we noted in chapter 2, and the financial resources available to the major boards. Overall, however, the sample material examined can give potential test takers useful insights into the scope, but also the limitations, of each certification.

3.1 The Shortcomings of Sample Materials

All boards necessarily publish information and advice about how to prepare for their certification. As well as detailed descriptions of the test structure, and information about scoring and reporting, there will usually be links to free materials, and publicity for practice tests and other materials which can be purchased. In this chapter we shall take a critical look at some of the free practice materials made available for TOEFL, IELTS, and PTE, on the assumption that would-be test takers – and teachers preparing students for the tests – will first look at these materials to get an idea of test structure, test items, and difficulty, before deciding to enrol for the test and invest in practice tests which are for sale.¹

However, it is not easy to get a complete overview of any single test. The free material tends to be fragmented, taking the form of downloadable PDF files, listening files, and partially interactive pages which focus on different parts of the exam. None of the websites administer a complete, timed, practice test in the mode (Internet-based for TOEFL, paper-based for IELTS, computer-based for PTE) used in the actual test. This is perhaps not surprising, especially if the boards want to sell practice material or complete, fully interactive (in the case of the PTE) sample tests.

¹ All the sample material analysed in this chapter was accessed in March 2017. It may of course have been removed, changed, or updated since then.

More surprising is the fact that a lot of the material is outdated, or problematic for other reasons, as we shall see. All three examining boards are guilty of posting problematic items, or texts which are plainly out of date, or which contain unintentional language errors. Of course, there is no such thing as a 'perfect test'. A 'test' is an abstraction, or rather a set of specifications, which Alderson et al. (1995, 294) refer to as "blueprints" which generate forms, the items and test tasks, which themselves, following the advice of Hughes (2003, 54) need to be sampled "widely and unpredictably" from the target language use domain. Items which do not work do not therefore necessarily indicate a 'bad test', but a glitch in the process of test production and validation.

However, it has to be of concern that all three websites offer problematic samples, some of which have been on offer for a long time. The boards do not give much information about where their sample material has been taken from, or whether it has been used in actual tests. If they are taken from previous tests, it is curious that some at least of the problematic items commented on in this chapter were not identified as such and weeded out after the test. What selection criteria were used, one might wonder, to choose sample items, if not to give an accurate idea of texts, tasks, and level of difficulty? Given the self image that boards are keen to promote of themselves (chapter 2), and the funds which they presumably have available for this purpose, it seems strange that they have not been more careful about which material to use in these key sections of their websites.

Nonetheless, for the potential user browsing these websites with a view to comparing the three tests and choosing one of them, the sample materials offer ample evidence of the ways in which the tests are different, although ostensibly testing the same skills and having the same function.

3.2 The TOEFL Sample Material

3.2.1 Critique of Sample Material for TOEFL iBT: Overview

The TOEFL website takes potential test takers to at least three different sources of free test material: a 32 page PDF document, an "interactive sampler" which can be downloaded and which reproduces the test interface, and a "Quick Prep" resource which offers further samples from all four sections of the test, in a PDF which includes embedded listening texts. Thus similar material is offered, in a paper version, in an interactive form, and in a semi-interactive paper version which connects to listening texts. This sample material is offered alongside other free resources, such as a "Test Prep Planner", and "Tips", all of which probably make it hard for potential test takers to understand where to go first to get an idea of

the test, or indeed which practice materials is likely to be the most useful, or the most recent. In addition, a separate link leads to a “TOEFL video library” with eighteen short video clips offering mini-tutorials on different parts of the test. In this critique we will consider in detail the paper and the interactive versions of the free test materials.

Neither resource offers a complete version of the test; the description of the interactive sampler as offering “free unlimited access to past TOEFL questions from all sections of the test” is misleading; the “unlimited access” means that one can return to the same sample and redo the questions. Other, more complete resources, are offered for sale on the same webpage.

Nonetheless, the sample material gives a good overview of what to expect in the exam. Both the PDF material, and the interactive sampler, which uses different questions, we are told, comprise actual questions from previous tests. Presumably the interactive sampler will give potential test takers a closer idea of what it is like to take the test, while the paper version is offered to help students get an idea of the test even if they do not have access to a PC or the required Windows 7+ operating system.

3.2.2 The PDF Sample

The PDF document has one reading text, with questions and answers; the tapescripts for two listening tasks, with questions and answers, questions, texts and tapescripts for all six speaking tasks, and questions for both writing tasks, with two sample answers for each, and a rationale behind the scores awarded.

The single reading text offers a possible explanation for the extinction of the dinosaurs. The text (for this reader at least) is well written, and interesting. Questions follow the order in which the information in the text is presented. Five are about understanding words in context, while other functions range from extracting main ideas, to inferring information, paragraph completion, and understanding writer’s purpose.

The first 13 of the 14 items are MCQs with 4 options. To respond to the last question, number 14, which is worth two points, candidates have to select three correct pieces of information from six statements, and transfer them onto the answer sheet.

There appear to be a number of problematic items, especially those testing lexis. Q8 and Q9 can both be answered without reference to the text:

- 8 The phrase ‘tentatively identified’ on line 36 is closest in meaning to
 a identified after careful study
b identified without certainty (correct answer)
 c occasionally identified
 d easily identified

- 9 The word ‘perspective’ on line 46 is closest in meaning to
a sense of values
b point of view (correct answer)
c calculation
d complication

Both terms, “tentatively identified” and “perspective”, belong to a fairly standard academic lexis; the questions require test takers to recognize standard definitions. For a student with a European background, in higher education, or aspiring to higher education, this lexis should not be problematic; if anything, he or she might assume that the obvious answer is wrong (since MCQ options are intended to distract), and be persuaded to choose a different answer, on the assumption that the term might have acquired a different meaning in a specific context.

Q2, which asks students why the writer includes the information that dinosaurs “had flourished for tens of millions of years and then suddenly disappeared”, also appears to be problematic, since at least two of the options (c and d) seem credible.

The listening section offers the tapescripts of two texts, a dialogue (between a male basketball coach and a female member of the basketball team) and a monologue (an extract from a lecture about a novel by Wilkie Collins). The dialogue has lots of discourse markers built in, to highlight the informal tone (*yep, wow, well, oh, okay, good*), and the five questions are on understanding the main points (Qs 1-2) and recognizing communicative functions (Qs 3-5). These questions do not appear to present any particular difficulty.

The monologue (approximately 900 words) seems more demanding. It focuses on what is generally considered to be the first modern detective novel in English, *The Moonstone*. As such, at least two answers might be accessible to test takers who are familiar with the novel;

- Q7 In what way is the *Moonstone* different from earlier novels featuring a detective?
a in its unusual ending
b in its unique characters
c in its focus on a serious crime
d in its greater length

- Q8 According to the professor, what do roses in *The Moonstone* represent?
a A key clue that leads to the solving of the mystery
b A relief and comfort to the detective
c Romance between the main characters
d Brilliant ideas that occur to the detective

A major flaw, which would be easy to put right, is the key to the listening section, which gives answers for questions 15-25; but for each tapescript questions are numbered, respectively, 1-5 and 1-7.

In the speaking part of the sample, the directions remind readers that in the actual test they would be listening as well as speaking, and in some cases, reading, listening and speaking. The first two tasks are responses to invitations to “speak coherently and clearly about familiar topics”. The first of these is to remember a pleasant and memorable event at school. This may seem deceptively simple, but for many candidates (and not just those with unpleasant or indeed traumatic memories of school) it could be problematic to remember a single pleasant event. The second task, which invites a comparison between two modes of spending time with friends (at home, or in a café or restaurant) seems more accessible, and better structured.

The remaining tasks integrate reading and listening tasks with a spoken outcome, and are more university-oriented. They get students to speak about social facilities on campus, the psychological notion of ‘flow’, problems students are facing following a course in calculus, and two definitions of ‘tool’ which are presented in an extract from a biology lecture. Tapescripts, complete with phonological reductions such as *gotta* in the extract from a lecture, replace the listening files of the actual exam.

The short reading text on flow is potentially problematic: it is presented as an extract from a psychology textbook, and students of psychology may well find it easier than other test takers to respond to the question “Explain *flow* and how the example used by the professor illustrates the concept”. In the same way, in the extract from the biology lecture, the two definitions of tool, as used by animals in nature, will probably come more easily to students familiar with the topic.

The last section, writing, includes a reading and listening to writing activity (on the vote counting system in the US) and a free-standing essay on the topic of what makes a good teacher. The two questions are both provided with two sample answers, both of which are at the top end of the 1-5 scale, the first sample scoring four points, and the second five. This is useful feedback. However, there seems to be a large gap in level between the first two sample answers (on the vote counting system), the second one being deficient in organization, and containing a large number of formal errors not usually acceptable in an academic writing context. It begins: “The leture (*sic*) disagreed with the article’s opinions” and continues in the same vein. If this deserves a mark of 4/5, one might wonder what a score of 3 looks like.

3.2.3 The Online Sampler

One of the main functions of the on-line sampler is to get would-be candidates to have a feel of the exam, both in using the interface, and in the timings of tasks. Like the PDF material, it does not offer a complete test, but it samples rather differently: there are three reading texts (compared to only one in the PDF), but only one speaking task; the listening tasks mirror the PDF (one 'campus' type dialogue, and an extract from a lecture) and there are two writing tasks, one based on a written and spoken input, the other being the free-standing listening task.

The interface is uncluttered and user-friendly, with a time bar indicating how much time is left to listen, or to answer, but moving from one section to another can be a laborious task, as one always has to move through a pre-programmed sequence; for example, users cannot go directly from the reading to the writing section, but have to click their way through listening and speaking, and they cannot return to a previous section. A drop-down menu could have rectified this and made browsing easier. Some feedback is given; students can use a button to reveal the correct answer to the reading and listening questions, and the speaking task is provided with two sample answers.

The three reading tasks deal with science and technology (windpower and botany) and prehistoric art (cave paintings). The first, on three theories behind the cave paintings in Lascaux, is noticeable because it begins with a non-standard sentence:

"In South-West France in the 1940's playing children discovered Lascaux grotto"

in which *playing*, to indicate progressive aspect, rather than a compound structure (*playing field*, *playing card*), would normally follow the noun. The structure does not compromise understanding in any way, but it is curious since it suggests that the writer might not be a native speaker; whereas all the certifications described in this chapter are based on native speaker models of the language. We shall return to the theme of non-native input in chapters 6 and 7.

As we saw with the reading text in the PDF sample, here, too, some reading questions can be answered by candidates without actually reading the text. In the second passage, for example, (on wind farms), the first questions appeals to basic understanding of what is meant by 'wind farm':

Q1 Based on the information in paragraph 1, which of the following best explains the term wind farms?²

- a Farms using windmills to pump water
- b Research centers exploring the uses of wind
- c Types of power plant common in North Dakota
- d Collections of wind turbines producing electric power**

Logical inference might help the candidate to choose the correct answer in other cases, again without reading the text, as in the following question on cave paintings:

Q8 According to paragraph 4, why do some scholars believe that the paintings were related to hunting?

- a Because some tools used for painting were also used for hunting
- b Because cave inhabitants were known to prefer animal food rather than plant food
- c Because some of the animals are shown wounded by weapons**
- d Because many hunters were also typically painters

in which options a, b and d take for granted that a lot was known about the prehistoric cave dwellers, while only c reasons from the evidence which emerges from the paintings themselves.

As with the PDF reading, here, too, there are a number of vocabulary items the meanings of which test takers are expected to infer from context. Some of them, however, are likely to be recognizable to European candidates because they are cognate with words in their own languages (e.g.: *methods*, *emit*, *accompanied*, *massive*) and, if so, not ideal items in a test of reading.

One question (n. 6) related to the third passage (opportunist vs competing plants) asks “Which of the sentences below best expresses the essential information in the highlighted sentence in the passage?”; but no sentence appears to be highlighted. As with the incorrect answer key in the PDF, this could easily be corrected.³

There are two listening tasks, preceded by rather lengthy instructions, which include a rationale about what the listening section does. The first task involves listening to a conversation between a professor and a student who has missed a class. The dialogue is clear, and the five questions straightforward. At one point the professor offers to lend the student a video tape, asking her if she has a VCR at home – dating the passage,

2 ‘wind farms’ is not highlighted in any way, through italics or quote marks.

3 The sentence is however highlighted in a PDF version of the sample material. URL http://toefl.uobabylon.edu.iq/papers/ibt_2015_1821899.pdf (2017-11-01).

probably, to the pre-DVD 1990s, and impacting negatively on the image of the test.

The second listening text is, following the pattern we have already seen, much more academic in style. It is an extract from a lecture, it is long, and describes a quite complex process of crystallization. There are six questions.

The third section has two speaking tasks. The first one asks candidates to talk about on-campus accommodation for students. We are provided with a single, mid-level, sample response. The candidate speaks carefully and slowly, in clear accurate English. He answers the question well, makes virtually no formal errors, produces language such as:

“In my opinion it would be in the better interest of a first year student to live in a dormitory on campus but I wouldn’t make it a requirement but make it a personal choice.”

and yet is called ‘not fluent’ because he speaks slowly.

He seems to have been knocked back only on speed of response. He has (in the opinion of the author) addressed the question appropriately, and thoughtfully, and the judgement we read “He does provide some relevant information but in general the topic is not sufficiently developed to score at the highest level” thus seems unfair. If this is a bad performance, or ‘mid-level’, it would have been more useful to have a sample of a good performance. It would also been useful to have a numerical score for ‘mid-level’: does the performance score 3/4 or 2/4?

The problem of the sample responses is compounded with the second speaking activity, in which test takers read an introductory sentence (in this case about taming herd animals), listen to an extract from a lecture on the same topic, and then have to explain how the behaviour of horses and antelopes as herd animals relates to their suitability for domestication. The sample performance is (rightly) flagged as low level, while poor pronunciation and a background hiss make it extremely difficult to follow. But this is of little help to the would-be test taker; far more useful would have been to provide an example (or examples) of a good response, at a mid-to-high level.

The final section contains two writing tasks. The first of these is a carefully structured test of writing from a reading and listening input. Candidates are given three minutes to read a short text (approximately 250 words, on the altruism of meerkats), after which the text is removed and they listen to a lecturer refuting some of the information given in the text. This second part lasts for about two minutes. The writing task is to summarize the lecture, and show how it sheds doubt on the reading passage. However, instead of being given the opportunity to write the text, candidates are then shown three responses, at different levels (high, mid, and low). There is no comment on these sample responses, but there does appear to be a clear gap in level between them, in terms of content, accuracy, and range.

The free-standing task, in contrast, offers test takers the chance to write using the test interface. They have 30 minutes to complete the task (a 300 word essay on whether or not telling the truth should always be the most important consideration in a human relationship). One useful interactive feature is the word count, which charts students' progress as they write. However, there are no sample answers to compare with one's own.

3.2.4 TOEFL Sample Material: Concluding Remarks

Although there is plenty of sample material, in a range of formats, it is badly organized, and it is not possible to do a complete practice test. The website shows signs of age, as does the material (some of which has been left untouched for years), while newer pages and texts have been added. The result is a lack of guidance for potential test-takers coming to the site hoping to have a clear idea of how the test is structured and what they will have to do.

3.3 The IELTS Sample Material

3.3.1 Critique of Sample Material for IELTS: Overview

The sample material is easily accessed through a user-friendly website. A drop-down menu "About the test" on the uncluttered homepage takes students to pages headed "two types of test" (which distinguishes between the academic version of the test and the general training version), "test format", where students can click on one of the four section headings (listening, reading, writing, speaking) to find out more about the test structure, and a third page, "test format in detail" which uses the same interactive section headings, but this time by clicking on them students find out more about the tasks, the question types and the scoring. The fourth page has links to sample material for each section of both versions of the test. There are other pages, too, about scoring, test development, fairness and security (amongst other things) but the first four pages give enough information for potential candidates to have a very good idea of what to expect in the test itself.

3.3.2 The Sample Material in Detail

There are nine sample listening texts, which take the form of mp3 files. This compares with the four texts of the actual exam. Along with the

questions and the answers, tapescripts of the texts are also supplied, but, inexplicably, only for the first seven tasks.

The first seven are short texts, some very short, and are representative of the first three listening tasks, which include two conversations and a short monologue, such as an announcement. Texts eight and nine are much longer (more than five minutes each), as is the final monologue in the actual exam.

Why does IELTS offer so many listening texts? The obvious answer is to offer potential test takers as much practice as possible. But since one of the most publicized features of the test is the range of (native speaker) accents it uses, this could also account for the wide sample. We hear UK, US and Australian accents in the samples, although some of them (such as the male 'American' in sample 4 and the female 'American' in sample 5) appear to be British English speakers attempting to put on American accents. This lack of authenticity is even more obvious in the first sample, in which a British English speaker plays the role of a Kenyan man (who is presumably not a mother tongue speaker of English) wanting to ship goods back to Africa.

The listening samples are also organized so that each one gives practice with different question types, such as matching, sentence completion, multiple choice, and labelling. The seven reading samples are similarly organized. Topics range from dung beetles to rocket science and pollution from cars, the risks of cigarette smoke, and agriculture and the environment. Like the TOEFL sample, much of the material seems dated: the reference to the 1986 Round Table, for example, on multilateral trade agreements, suggests that this is a recent event the results of which have not yet been felt.

There are other ways in which the samples are likely to be less than satisfactory, or even confusing, for would-be test takers. One of the samples (on the dung beetle) is used twice (samples one and seven) to be exploited through different question types. Another sample, on agriculture and the environment (sample six) is recycled in part in the next text (sample seven), which begins: "All these activities may have damaging environmental impacts". It thus begins with a reference to part of text which is not shown, although the introductory rubric reads: "The text preceding this extract explained how subsidies can lead to activities which cause uneconomical and irreversible changes to the environment". A student coming to this sample might wonder if it is standard practice for texts to begin like this one *in medias res*; there is no answer to this question.

A further curiosity is the choice of (very) different fonts for the different reading texts. There seems to be no reason for this, unless it is to indicate that they come from different sources, and thereby hint at 'authenticity'. But they are not facsimiles, and although we read in the introductory material that "texts are taken from books, journals, magazines and newspapers, and have been written for a non-specialist audience", no credits are given to indicate the actual sources - if indeed they have not been specially written for the exam.

The overall effect then, of the sample listening and reading material, is to give candidates an idea of the type of questions they will have to answer, but not the feel of a complete reading paper, which would have only three long texts, but to each of which would be appended two or more question types.

The third part of the test, writing, is adequately covered in the sample. In the actual exam candidates have to write a 150 word report synthesizing or summarizing material from a visual input such as a graph or chart. There are two example questions. This is followed by a 250 word argumentative essay, in which the writer is invited to agree or disagree with an opinion, and provide arguments in support of their choice. Here, too, there are two sample questions. However, (unlike the TOEFL pages) there are no sample answers.

What is most striking about this material is that it is decades out of date. Both of the graphic input questions present data from the 1990s, while the first essay question reads:

The first car appeared on British roads in 1888. By the year 2000 there may be as many as 29 million vehicles on British roads. Alternative forms of transport should be encouraged and international laws introduced to control car ownership and use. To what extent do you agree or disagree?

To refer to the year 2000 as if it belonged to the distant future, rather than a rapidly receding past age, is likely to lead to a moment of disbelief for the would-be candidate looking at this material. Admittedly it could be justified as providing an example of the kind of writing task a candidate could be faced with. But to use it as an actual test item today would be unthinkable – how would the candidate begin to answer it? – and one can only wonder why a more updated sample has not been provided.

The speaking part of IELTS is a free standing test which can be taken on a different day from the rest of the exam. The sample material covers all three phases, the introduction, the ‘long turn’, in which the candidate talks about a topic which is provided by the examiner, and the discussion. Separate PDF files are provided with the questions, and tapescripts and mp3 listening files with the partial answers of one candidate. It is not clear why we are not given the candidate’s complete performance; after all, the complete speaking test lasts only for around twelve minutes. A further failing is that we are not provided with any indication of how the candidate is rated. Although he speaks clearly, makes few formal errors, and provides thoughtful, intelligent answers to questions, there are long pauses; any would-be candidate listening to this sample would probably want to know if they are penalized for the pauses.

At this point, we can refer to the band descriptors for speaking (public version) which are available on the IELTS website. Fluency, we note, is one of the assessment criteria, but ‘content-related hesitation’ does not

prevent a high mark (band 8 or 9) from being awarded. Interestingly, this seems to be in contrast with the approach to fluency in the TOEFL listening sample which we referred to earlier in this chapter, and in which performance was penalized because of the slow delivery.

The question is, whether or not the long (eight seconds!) pause in the sample material is ‘content related’; the candidate seems to be thinking of something more to say in the ‘long turn’, (so hesitating as he searches for content), rather than experiencing any particular language problem, until the silence is broken by the examiner with a prompt. There are other long hesitations in the same file, and as such it seems like a strange choice to offer as a sample candidate performance.

3.3.3 IELTS Sample Material: Concluding Remarks

The sample material does not do justice to the overall well-designed website. As with the TOEFL website, we find material that is incomplete, or dated; and for the subjective parts (writing and speaking) there are no examples of candidate performance (writing) or evaluation of performance (speaking). It is difficult to understand why this should be, since these could have been provided fairly easily.

Like TOEFL the IELTS website does not offer a complete version of an exam for practice, but, again like the TOEFL website, it offers a range of practice materials, including complete tests, for sale.

3.4 PTE Academic Sample Material

3.4.1 Critique of Sample Material for PTE Academic: Overview

The PTE is the most recent academic certification on the market (2009). It has an uncluttered home page making it easy to find and access the sample material through the drop-down menu. The “Test taker” menu leads to a “Preparation” page, where one option for “free materials” is given alongside a range of materials which are for sale: “scored practice test”, “sample questions”, and course books.

The “free materials” link offers four features: two PDF documents “Test tutorial” and “Top Tips”, a “Skills video” which is a collection of short YouTube clips showing the range of tasks the candidate has to perform in the test, and an “Offline practice test”. This latter brings together sample questions and answers, with comments, for the productive skills, on sample performances of test takers. Since the same sample items (or at least, some of them) turn up in all four blocks of materials, we shall focus

primarily on the Offline practice test, which offers the most materials.

The “Test tutorial” and “Top Tips” documents are similar in form and size, the former being a 38 page document, the latter having 40 pages. The tutorial focuses mostly on test procedure, while the “Top Tips” pages pepper the sample tasks with brief exam-taking strategies such as “Use punctuation to help you decide when to pause when you read” or “Skim the text before the reading begins”; the tips frequently border on the trite, such as “Make good use of the 40 second speaking time”. Some of the tips are also to be found in the tutorial document, and some of them are self evident, such as “Use correct punctuation for writing tasks” or “Don’t click NEXT before you have completed the task and are ready to move on”. Given the overlap of function and content, these two documents could probably have been more usefully combined as a single document giving procedural information and exam strategies.

The sample videos have been uploaded onto YouTube. Rather misleadingly called “Skills” videos, they are brief, approximately half minute clips which familiarize potential test takers with the procedures for the many different tasks. However, the clips fade out after the instructions have been completed, or during sample student responses; they are not intended as practice material.

3.4.2 The Offline Practice Test

The “Offline practice test” is more than a test, since it offers multiple items for some of the shorter task types, such as “repeat sentence”, “describe image” and “answer short question”. Readers see screenshots of the tasks, and on a later page are given the answers (for objective type questions) or can read or listen to sample candidate answers.

The sample responses are a strong feature of the practice test, since for each task three responses are given, illustrating three key levels B1, B2 and C1 of the CEFR. It is interesting that no reference is made to the Pearson “Global Scale of English” (see chapter 2, 2.3), so we do not know the exact score for each response. However, the Framework levels will probably be far more meaningful for prospective universities, which are likely to discard B1 candidates as below level, to view B2 candidates as potential students, and C1 candidates as fulfilling all language requirements.

The samples are either written, or, for speaking activities, available on audio files. Besides the attribution of a CEFR level, each response is described in four or five lines. This is likely to give useful feedback about the test to would-be candidates, who can identify those features which are clearly below level. One comment, for a B1 level response to a ‘describe image’ task, can suffice:

Two basic elements of the graph are described, but the main idea is not discussed. While there are a few phrases spoken at a natural rate, fluency is negatively affected by multiple hesitations and long pauses. Incorrect pronunciation of consonants might require listeners to adjust to the accent of the speaker. There is limited control over simple grammatical and lexical structures. This response lasts for 31 seconds.

This response is below level because the task has not been fulfilled, it is not sufficiently fluent, consonant production is problematic, and there are grammatical and lexical inaccuracies. The reference to “incorrect pronunciation” is interesting, since it suggests that the need for the listener to “adjust to the accent” is a negative feature of the response, whereas listeners always have to adjust to accents, whether native or non-native, and the sample texts which students have to listen to in this offline test contain a range of British, American and Australian accents.

The range of questions on offer is helpful, but also points to problems. For example, one of the speaking activities, “Answer short questions” is more of a listening comprehension and vocabulary check than a speaking task. Candidates listen to definitions and identify the word. This in itself is not necessarily an inappropriate task in an academic test, but it could be relabelled, or re-presented, as a listening task. The problems arise in the variety of the nature of the task. In some sample questions, students have to identify one of three given words, such as:

Which is the longest - a decade, a millennium, or a century?

making it a three-option multiple choice question, whereas in other questions, candidates are not given the target word, but have to work it out from themselves, as in:

What key mineral makes seawater different from freshwater?

or

If a figure is hexagonal, how many sides does it have?

This last question could easily have been turned into “If a figure is hexagonal, does it have five, six or seven sides?” making it into a qualitatively different type of question.

There also seems to be a labelling problem with the “Retell lecture” task. The first of three samples comes across as an extract from a lecture. The pauses, added emphases, and overall intonation patterns all give it an authentic feel. The other two samples, however, are extracts from inter-

views, both involving two speakers, an interviewer, who takes two turns, and an interviewee, making them dialogues rather than monologues, and the title “lecture” a misnomer.

A further problem with items arises in the “Highlight incorrect words” in the listening section, in a sample task which features in the “Tutorial” but is not repeated in the offline test. This concerns possible test bias (which we also noted in a TOEFL reading passage). In this task, test takers have to listen and highlight the words which are different from the text they have in front of them on the screen. In the following introduction, however, the story of Amundsen’s quest for the north-west passage may be familiar to many European students, enabling them to identify the incorrect words for the wrong reasons:

When explorer Roald Amundsen set out to find the Northwest *Pasture*, his official mission was scientific – a search for the magnetic *south* pole. (italics added)

A further limitation of the sample material is to be seen in the quality of some of the written texts. Here are two extracts from the short texts used as examples for the “fill in the blanks” task in the reading section:

Up until our research the predominate wisdom in the scientific community was that umami was not a separate sense.

Peering into the future seldom produces a clear picture. But this is not the circumstances with bio-energy.

“Predominate” as an adjective? All learner dictionaries give this word as a verb, and only by going to a big dictionary, such as the Oxford Shorter, do we find the entry “Now rare. [App. a mistaken form for predominant]”. “Circumstances” a singular noun with a plural marker? Or should it have been “These are not the circumstances”?

These two errors – or slips – seem all the more problematic in that both “predominate” and “circumstances” are the target words in the activity, which test takers have to select to complete the texts. Here, the danger is that they discard the correct answers for the right reasons: their superior (to the item writer’s) knowledge of the language.

3.4.3 PTE Academic Sample Material: Concluding Remarks

Of the three certifications, the PTE website probably offers the most complete information to potential candidates. As we noted, the provision of sample responses, with comments, at three different levels, for all the

productive tasks, is a strong feature. In addition, in keeping with the user-friendly home page, clarity of design is also a feature of the downloadable booklets. These are written in an appropriate style for learners of English, addressing them directly, and highlighting important information in bold.

The sample material is packaged in a zipped file (which may take some time to download). Apart from the problematic items, which we have commented on, one noteworthy feature is the disparity in the number of samples per task type. For most task types there are two examples, but there are six “read aloud” and “describe image” questions, and no fewer than ten examples of “repeat sentence” and “answer short question”. Why, one wonders, does Pearson want us to have ten examples of such a straightforward activity as “Please repeat the sentence exactly as you hear it”? The unintended message that may be inferred by potential candidates is that this activity has to be repeated several times in the actual test. Perhaps their interests would be better served if this file were to be labelled “Sample questions and answers” rather than “Practice Test”.