

タイトル	How to Write Useful and Effective Reading Comprehension Tests for Intermediate Level Learners
著者	MUNBY, Ian
引用	北海学園大学人文論集(75): 55-84
発行日	2023-08-31

# How to Write Useful and Effective Reading Comprehension Tests for Intermediate Level Learners

Ian MUNBY

## Abstract

If we teach reading comprehension skills to foreign language learners, we may need to test their reading ability. Although ELT publishers provide an abundance of reading comprehension tests, they may be unsuitable for our students due to issues with length, level of difficulty, or uninteresting and unfamiliar content. In addition, test items may not test the reading strategies that we wish to develop in our learners. Drawing from the research literature and personal experience, this paper describes an approach to writing a series of reading tests using needs-based specifications for text length, readability, and question type. The initial focus is on ways to modify texts from Wikipedia to suit lower intermediate to intermediate level students in an English major program in Japan. Next under the spotlight is topic selection for reading passages, especially in the light of what is known about how the level of reader interest and background knowledge may influence comprehension. With reference to a sample test in the Appendix, sometimes referred to as a practice test in this paper, the next objective is to explain methods for writing test items that foster reading strategy development. For example, to promote understanding of the rhetorical organization and main ideas of a text, advice is given on how to write paragraph headings to match with paragraphs. For testing

inferencing skills and comprehension of content and the author's purpose, some approaches to creating MC (multiple-choice) items and true or false statements are described. The paper concludes with guidelines for testing reading strategies such as understanding referents and signal words and guessing the meaning of unknown words.

*Keywords:* reading comprehension test, reading strategies

### **The teaching and learning situation**

The series of reading comprehension tests referred to in this paper form part of an assessment package for required reading courses in an EFL program for English majors at a private university on the island of Hokkaido in northern Japan. Students enrolled in these courses take mid-term and end-of-term reading proficiency tests in each of the four semesters of their first two years of study. According to Bachman (1990) one of the two main purposes of language tests is to act "as sources of information for making decisions within the context of educational programs" (p.54). In this context, the purpose of these tests is to make evaluative decisions about the students and their reading comprehension ability and changes in it over time. Each one-hour test consists of two reading passages with twenty questions each, and raw scores out of 40 are converted to percentage scores. However, Alderson (1996) cautions that reading tests such as the ones described in this paper should not be the only kind of assessment used in a reading course so each test is weighted to be worth 20% of the final grade for each of the four courses. The remainder is made up of 20% for homework reading tasks, 10% for two short story reviews, and 30% for two reading textbook chapter tests. A minimum final score of 60% is necessary to pass each course. The chapter tests are

essentially achievement tests since they are designed to test vocabulary and reading strategies that appear in the textbook and are taught in class. The mid-term and end-of-term tests are reading proficiency tests and they differ from the chapter tests because they have no separate vocabulary testing sections, and it is not possible for students to prepare for them by reviewing the textbook, for example.

### **Reading test length, level of difficulty and time allowed**

Following the advice of Bachman & Palmer (1996), the writing of any test for learners of a second language should be underpinned by philosophical considerations. For example, they believe that test writers should design tests “so as to encourage and enable test takers to perform at their highest level of ability” (p. 13). It is, therefore, most important that reading comprehension tests be carefully tailored to suit the level of the test-takers and that ample time is allowed for them to read and answer the questions and demonstrate their ability. For this reason, although the specifications for this series of reading tests are inspired by and partly modelled on the academic module of IELTS reading tests, the latter tests were considered too difficult for most students in the program, and the time allowed was too short. For example, most students in this EFL program achieve scores of TOEIC 400-600 in annual in-house tests, corresponding to IELTS level 3 or 4, or lower intermediate to intermediate.

From an empathetic or learner-friendly point of view, Anderson (1999) reminds us “you want the learners to have a successful experience practicing new reading skills” (p. 117), and excessively challenging tests do not present optimal conditions for student success. Furthermore, from a practical point of view, in his discussion of factors affecting test reliability, Bachman (1990) points out that: “If the test is too easy or too difficult for a

particular group, this will generally result in a restricted range of scores or very little variance” (p. 220). Alternatively put, a restricted range of scores will defeat the evaluative purpose of the test.

However, with reference to Table 1, the length of the reading passages in this program is gradually extended in each semester of first and second year. In addition, the readability, or level of difficulty of the text passage, is

**Table 1**

*Specifications for reading test passage length, time allowed, and level of difficulty for each course compared with a sample IELTS test.*

	Number of words	Flesch-Kincaid grade level score	Time allowed per passage
1 <sup>st</sup> . year Spring Semester	600-650	9.0-9.9	30 minutes
1 <sup>st</sup> . year Fall Semester	650-700		
2 <sup>nd</sup> . year Spring Semester	700-750	10.0-11.0	
2 <sup>nd</sup> . year Fall Semester	750-800		
Sample IELTS test	760	12.5	20 minutes

also increased. Readability is measured according to the Flesch-Kincaid formula (Kincaid et al., 1975) which assesses text difficulty by considering the total number of words in a text, the average number of words per sentence and the average number of syllables per word in a text. For the purposes of this series of tests an online calculator (Good Calculators, 2022) was used to generate a Flesch-Kincaid grade level score corresponding to the US education grade level. For example, readability parameters for the reading passages for the first-year tests were set at 9.0-10.0, corresponding to ninth grade, and 10.0-11.0 for second year tests (see Table 1). The readability score for the sample test was 9.8 (789 words). Note that Section 1 of an online sample IELTS reading test (British Council, 2022, academic module) entitled “Electroreception” was assessed at grade level 12.5 by the calculator. The twenty minutes allowed for this test was also discovered to be too short not only for the IELTS test, but also for the reading tests

described here. It was determined by trial and error that thirty minutes was ample for each section of the test.

Wikipedia is recommended as an ideal source of texts for reading test passages. Nevertheless, with intermediate and lower intermediate level students, texts from this source may need to be extensively modified and simplified in order to match the specifications in Table 1. This process typically involves deleting large parts of the text, and the sample test was shortened to 789 words from an original Wikipedia entry of 2,840 words. It may also involve replacing many, but not all, of the lexical items likely to be unknown to students with known items, and shortening sentences by inserting periods, for example. Note that it is possible to reduce the Flesch-Kincaid readability grade level of a passage by dividing a small number of long sentences in half to increase the number of sentences. For instance, the readability grade level of “Electroreception” (2022), the sample IELTS reading passage referenced in Table 1, can be reduced from 12.5 to 10.9 by editing and increasing the number of sentences from 38 to 46.

### **The role of interest and prior knowledge in topic selection for reading test passages**

The first step in designing a test is the selection of an appropriate reading passage upon which to base questions. Cohen (1994) recommends that we “choose a text with a familiar topic, that is interesting, has an unambiguous intent, and is of an appropriate length” (p. 250). Interestingness, or the degree to which a text is interesting to the reader, is hard to define, but in principle, most commentators agree that if L2 readers are engaged by the topic and the content of a reading passage, their text processing will be more successful. For example, Schmitt (2002), citing Carrell and Wise (1998), claims that “if either prior knowledge or topic interest is high,

students perform better than if both prior knowledge or topic interest are low” (p. 245). These findings were partly backed up by Eidswick (2010) who investigated reading topic interest and prior knowledge and its relationship to reading comprehension performance among Japanese college-level learners. Among the 11 topics he studied, he found that the three topics with the highest learner ratings for interest and prior knowledge were biographical texts. Perhaps unsurprisingly, he found that his subjects expressed more interest in reading about Michael Jackson than about concrete and viruses. His analysis of reading comprehension test results also revealed significantly higher scores with high interest and high prior knowledge topics. A second key finding was that there was no significant difference between reading test scores for texts on high interest-low prior knowledge topics and low interest-low prior knowledge texts. Finally, level of prior knowledge corresponded significantly with reading test scores. The suggestion is that prior, or background knowledge was a more important factor contributing to reading comprehension performance than topic interest.

Before addressing the issue of prior knowledge, it is worth noting some limitations concerning evaluation of topic interest in test passage selection. To begin with, famous people such as Michael Jackson often disappear from the limelight quite quickly, perhaps rapidly becoming uninteresting and unknown. Further, it may not be wise policy to select texts solely based on perceived high interest value among the target group of test-takers for two reasons. First, with students in this reading program, for example, high interest value topics may be too narrowly focused on pop music, movies, and smart phones. Second, since learners must learn to read and understand texts on an ever-broadening range of topics as they make gains in L2 vocabulary and overall second language proficiency, it is the duty of teachers to provide a variety of reading text types covering a variety of

topics in the language classroom.

As suggested earlier, prior knowledge of reading test topic is of paramount importance when designing effective and useful reading tests. Anderson (1999) defines background knowledge as knowledge which includes “all experience that a reader brings to a text: life experiences, educational experiences, and knowledge of how texts can be organized rhetorically” (p. 11). In the literature, this is also referred to as schema, or schemata. Absence of relevant schemata may lead to comprehension failure. As Anderson points out: “readers may not have any knowledge of playing certain sports. If you have no knowledge of how the sport is played, or the vocabulary involved in it, you have no knowledge to activate prior to reading about it” (p. 12). Weir et al (1999) observe: “The content of a text should be sufficiently familiar to candidates so that candidates of a requisite level of ability have sufficient existing schemata to enable them to deploy appropriate skills and strategies to understand the text” (p. 29).

On the other hand, care should be taken to ensure that the content is not so familiar that test-takers can answer comprehension questions without having to read the test passage. For example, Saito (2023) used ChatGPT to create a reading passage with comprehension and discussion questions on the topic of Shohei Ohtani, the Japanese star baseball player. The topic is likely to be of high interest among many young adult Japanese learners compared with other reading topics to be found in ELT reading textbooks. However, although it must be stressed that the materials were designed for teaching rather than testing, many learners would probably be able to select correct MC answers by using their knowledge of the player and his abilities and achievements without having to search for them in the reading passage. With these parameters in mind, test writers need to strike a careful balance between familiar and unfamiliar content.

The text passage in the sample test describes a brown bear attack in



Hokkaido. It was chosen not only because of recent heightened interest in wild bear encounters in the local news media, but also because the test-takers in this program here would be able to draw on and deploy local knowledge to support their text comprehension. In this case, topic familiarity and associated schemata are guaranteed since residents of Japan's northern island know that brown bears are large, dangerous, and hibernate in the winter. They also know that bears are sometimes encountered in residential areas, even entering people's houses in search of food. Furthermore, local place names and the Japanese names of characters that appear in the story would also be familiar. Although the story of the Sankebetsu bear attacks is famous, test takers would almost certainly not be able to answer the questions without reading the text.

However, before spending time preparing the test reading passage and questions, several suitability criteria must be satisfied. This involves previewing the candidate reading test passage for problems that may lie ahead in designing comprehension questions, specifications for which appear in Table 2.

### **Matching paragraph headings to paragraphs to test understanding of main ideas**

Nuttall (1996) mentions three criteria for evaluating texts for reading development and these also apply to the development of effective reading tests. The first two - readability and suitability of context - have already been discussed but her third criterion of exploitability, or how easily a test writer can exploit a text, depends to a great extent on schemata, or, more specifically, the rhetorical organization of a text. To support reader comprehension, a reading passage needs to be organized or structured into paragraphs of roughly equal length, or in this series of tests, at least six

**Table 2**

*Specifications for types of reading comprehension test questions.*

Types of questions per passage	Number per passage
1. Matching paragraphs to main ideas	5
2. Multiple choice comprehension questions	
a) Guessing meaning of unknown words, or words with multiple meanings based on contextual clues in passage	2
b) Pronouns and referent words.	1
c) Signal words or expressions	2
d) Making inferences	1
e) Author's purpose	1
f) Other category comprehension questions	3
3. True or false statements	5
	20 (total)

Note: Comprehension questions should, where possible, be given in order of how information appears in passage.

Note: True-false questions should not be given in order of how information appears in passage. The goal is to force students to skim and scan the entire passage in order to locate the correct information with which to answer these questions.

paragraphs of between 100–150 words each. In addition, test writers need to ensure that they can write short headings for each paragraph if the headings in the original text are not available or suitable. Test-takers need to be able to match these headings to their paragraphs for Part 1 of the reading comprehension task, as specified in Table 2. This task involves identifying the main idea of each paragraph and is modeled on a testing format adopted in the IELTS reading tests. The matching activity is thought to promote effective reading strategies in the L2 reader. Wallace (1992), for example, makes the following claim: “Effective readers of all levels of ability will scan for specific information, skim to get an initial overview of a text, and be prepared to read and re-read with greater attention to those parts of the text that are of particular relevance to their

purpose” (p. 62). In the sample test, there are seven paragraph headings, one of which is used as an example, and one which is not used. This is to ensure that one incorrect matching does not automatically generate a second error. However, reading passages of seven or eight paragraphs can also be used with the task altered to exclude one or two initial or final paragraphs from the matching activity.

The following three points should be borne in mind to effectively exploit the paragraph matching activity and its key purpose of promoting the strategy of reading for main ideas. First, we should avoid vocabulary in the headings which is likely to be unknown to most test-takers. Second, we should try to avoid headings that are phrases that have been simply copied from the text, or headings which enable readers to perform matching without processing the text, or without activating schemata for identifying the main idea. As McNamara (2000) explains, test questions should be based on “rewordings of propositions in the text, not direct lifting of words or phrases from the text. Without paraphrase the task may require nothing more than a literal matching of similar words or phrases in the text and the question rather than an understanding of the meaning of the propositions involved” (p. 30). This may involve making alterations to some words in the text and headings to act as a disguise, by using synonyms for key words, for example. Finally, some texts have inherent, or in-built exploitability for the paragraph matching activity through their schemata. For example, biographies may exhibit readily exploitable rhetorical organization if they are structured chronologically with paragraphs focusing on, for example, early family life, education, rise to fame, height of fame, private life, and opinions of a famous person. Headings in Wikipedia entries can often be used directly or with simple modifications for the paragraph-matching activity.

Another schemata-based approach to the writing of passages for

reading tests involves structuring paragraphs on recurrent examples of themes or incidents. For instance, the sample test describes six encounters with the bear in six different paragraphs. This enables the reader to add information to existing schemata, or to adjust a recurring text pattern based on and reinforced by their predictions of text content. These predictions may include information a reader may expect to find including date, time, and place of the encounters, and details of who was involved and what happened to them and the bear, and how the encounter ended. The same strategy was used to write comprehension questions for a reading test in this series on the Loch Ness Monster, where paragraphs were organized into separate reported sightings of the monster. Similarly, this approach was used to structure paragraphs in a test describing UFO sightings in the USA. In the next section, we examine how recurrent theme-based paragraphs also add exploitability to enable the writing of comprehension questions.

### **Writing multiple choice questions and true or false statements for content comprehension**

According to Brown & Hudson (2003) “reading . . . can be efficiently tested using the multiple-choice format” (p. 69), and there are ten multiple choice questions following the paragraph matching activity. This section focuses on the testing of understanding of inferences and the author’s purpose (one question each) and “other category comprehension questions” (three questions). The latter questions aim at more general content comprehension and test writers can also exploit schemata to simplify their writing, as in the following questions from the sample test:

6. How many times was the bear shot by hunters?

- A. Two                      B. Three                      C. Four                      D. Five

7. How many people died as a result of the bear attacks?  
A. Five                      B. Six                      C. Seven                      D. Eight
8. How many people were injured by the bear?  
A. Two                      B. Three                      C. Four                      D. Five

Scanning texts for information is a process normally associated with the initial stages of reading comprehension, and the search for main ideas. However, in order to answer these questions, the test-taker is required to re-read the relevant parts of each of the six paragraphs more carefully. Only by extracting information from the descriptions of each encounter is it possible to arrive at the correct number. In order to do this efficiently, the reader probably needs to have a good grasp of the rhetorical organization of the information in each paragraph to facilitate the search for the necessary information. This said, it is not necessary to create a reading passage which is structured around recurrent themes or incidents in order to develop a useful reading test. There are many other ways that a reading passage can be organized in order to allow a reader to utilize his or her experience and knowledge to support successful text processing. Smooth navigation through a six-paragraph text can also be fostered with texts describing stages of an interesting journey, a process, or the history of a person, a business, or a campaign to name a few examples.

However, before modifying a reading passage, it is essential to check for the availability of enough “hard facts” in the text upon which to base “other category comprehension questions”. Typically, these facts should take the form of potential answers to open questions including the *wh*-questions of *who*, *what*, *why*, *when*, and *who*, and *how* questions such as *how*, *how many*, *how long*, and *how often*). In other words, test-writers are advised to ensure that there are enough numbers, dates, times, and names of people and places, or even reasons for *why* questions, in order to enable the writing of comprehension questions including a correct answer with

three distractors (sometimes spelled “distracters”), or wrong answers. From personal experience, much time can be wasted by embarking on reading passage modification and on the writing of paragraph headings before noticing that there are not enough hard facts needed for question writing. For this reason, the comprehension questions are best written before beginning text modification to ensure problems do not arise later.

In addition, test writers need to be aware of several rules of thumb regarding multiple choice question writing. Brown (2005) offers the following advice. First, all unintentional clues to answering an item should be avoided. For example, one type of unintentional clue mentioned by Bailey (2018) is an option which is longer or shorter than the other three choices in the set. Therefore, all MC options should be of roughly equal length, and if that is not possible, two should be shorter and two should be longer. Question stems and options should also be as short as possible to avoid putting too much extra text processing burden on the test-taker. Second, all the distractors should be plausible, and certainly not what Cohen (1994) describes as “absurd” (p. 231). Third, needless redundancy, or word repetition, should be avoided in the options. As a personal preference, the only exception made to this rule concerns definite and indefinite articles (*a*, *an* or *the*). Fourth, the ordering of the options must be carefully considered. Note that the number of correct A, B, C or D answers should be as equally balanced as possible with no less than 2 and no more than 3 of each letter in a set of 10 test answers. Fifth, distractors such as “none of the above”, “all of the above”, or “both A and B” should be avoided.

Next, to satisfy the specifications, we need to include an inference question. According to Nuttall (1996): “These questions oblige the students to consider what is implied but not explicitly stated”. She adds: “they may require the reader to put together pieces of information that are scattered throughout the text” (p. 188). Alternatively put, they teach students to

examine the text in its entirety (Day & Park, 2005). For example, Question 13 in the sample test focuses on the ability to make inferences:

The bear was killed . . . . . days after the attack on the Ota family home.

- A. two                      B. five                      C. nine                      D. fourteen

In order to infer that the correct answer is B, test-takers must notice in Paragraph One that the bear killed two people at the Ota family home on December 9. Although this date is not given in the text, we are told that it happened nine days after the bear returned to the Ikeda family home on November 30. Details of the bear's death on December 14 are given in Paragraph Six. Similarly, Day & Park (2005) suggest that questions requiring inferencing ability can be created from biographical information in a text where test-takers must infer and calculate a person's age at death by combining their date of birth, stated at the beginning of the text, and the date of death at the end. The date of death can be substituted for another important event in a person's life such as a marriage. While the inference questions described above are mathematical, there are other types which do not involve calculations and these will be subject of further discussion regarding both questions on the author's purpose or *true or false* items.

Alderson et al. (2015) claim that "interpreting an attitude or intention of a writer" is an important reading sub-skill (p. 74), and the test specifications require that one item should test understanding of the author's purpose. It is customarily included as the final multiple-choice item since it may either require the reader to consider the text in its entirety or at least the concluding remarks in the final paragraph. Once again test-writers are warned against spending time transforming a Wikipedia entry into a suitable reading test passage before previewing the text for material upon which to base a question on the author's purpose. Although a writer's purpose is often categorized in simple terms as falling into one of three categories: to inform, entertain, or persuade (e.g. Wallace, 1992), they can

serve as a useful starting point. For example, with persuasive reading passages dealing with controversial issues related to social justice and movements for social change, questions can be centered around identifying the author's view from amongst four commonly held opinions. With informative reading passages about past events, the author's purpose question can be based on speculation surrounding the reasons behind an event. In the sample test, for example, Question 15 asks the reader to identify which of the four reasons suggested for the bear attacks is regarded as most likely by the author. Questions which ask the reader to consider the purpose of the whole text are more challenging to write, but one useful approach can be to base a question on the main idea of a passage. This could be whether the author's main purpose is to describe a person's life, a major event in the person's life, or to discuss that person's opinions, or a social theme related to these opinions.

The test specifications also call for the inclusion of five true or false items (as featured in the IELTS reading module), and these are useful when it is not possible to conceive enough distractors for a MC item. They also allow the test-writer further opportunities for testing inferencing skills and author purpose alongside regular content comprehension. As mentioned earlier, inferencing can be based on problems that a reader can solve by combining numerical information from two or more places in the text. However, an alternative approach to inferencing involves exploiting the subtext. For example, in the sample test passage, we can infer that the group of thirty armed guards made a serious mistake. In Paragraph 3 of the reading passage in the sample test we learn that they "headed down a river on what they thought was Kesagake's (the bear's) trail". The implication is not only that they were incompetent, but also that they could have prevented the attack on the Miyoke family home. Second, we can conclude from the information in the text that the bear Kesagake himself was quite



smart, especially in deciding not to enter a house where hunters lay in wait for him (Paragraph 5). During final editing of the test, true or false items based on these implications were removed.

We can also use true or false statements to exploit potential misunderstandings of the text that may arise from imprecise comprehension of a text. For example, Q20 in the sample test (The bear attacked a horse at the Ikeda family home) is designed to test degree of comprehension precision of the following lines from the reading passage: “the surprise encounter (with the bear) threw the family horse into a panic at the Ikeda family home (Paragraph One)”. The word “threw” is potentially suggestive of a physical attack on the horse when none took place. Finally, as a rule of thumb when writing true or false items, Cohen (1994) also warns that negatives and double-negatives should also be avoided because they can create ambiguity and confusion. Also, as mentioned earlier in notes related to Table 2, the true or false items do not appear in the order in which clues to answers are found in the text. The aim is to allow test-takers to activate scanning skills to locate the required information using his or her understanding of the structure of the text, especially the way it is organized into paragraphs.

### **Writing multiple choice questions for testing reading strategies: referent words, cohesive devices, and lexical inferencing**

The remaining five items in the MC section of the test specifications (see Table 2) are reserved for questions focusing on the following three categories, or reading strategies: understanding referent words, cohesive devices, and lexical inferencing, with one or two of each category required for inclusion in the set of ten MC items. In the sample test, there is only one question on referent words, with two each for the other two categories. In

my experience, it is not necessary to preview a reading passage for the availability of these items in the way that is recommended for the other five MC items. This is because it should always be possible to design questions for these categories in any reading text that you select.

Although referent words are sometimes described as cohesive devices in the literature (e.g. Nuttall, 1996), for the sake of clarity, and to avoid ambiguity in this paper, I do not class them as such. Identifying and understanding referent words involves what Alderson et al. (2015) refer to as the reading subskill of “understanding grammatical relationships of words or phrases across text” (p. 74). The referent words used most commonly in these tests are pronouns such as *it*, *he*, *she*, *they* or *them* or possessive pronouns such as *its*, *his*, *her*, or *their* which refer to nouns or noun phrases in the text. Understanding of reference can also be tested using demonstrative pronouns such as *this*, *that*, *these*, or *those*, which can also be exploited for probing reader understanding of both anaphoric and cataphoric reference. For example, *this* in a sentence, or part of a sentence, such as “this is what he did” can either refer to a verb phrase in a previous part of a text (anaphoric reference), or a part which follows (cataphoric reference). Referent words are underlined and numbered in subscript within the reading passage. For example, in the practice test, the item is presented as follows: “She heard a rumbling noise in the garden, but before she could determine what <sub>(10)</sub> it was, the bear broke in through a window and entered the house”.

Q10. In Paragraph Four, it refers to ...

- A. the noise.      B. the bear.      C. the garden.      D. the meal.

One design consideration to be borne in mind is that the referent word should not always be closest to the noun it refers to. However, the distractors should all be words or phrases that are present in the vicinity of the referent to be reasonable, or plausible options.

The second category tests the reader’s understanding of cohesive devices, also known as transition signals, or discourse markers. These are removed from the reading passage and replaced with a numbered space. For example, in the sample test, the item is presented as follows: “On December 13, a team of six hunters waited inside the house. [12] , the bear reappeared, but stopped and seemed to check the inside of the house by looking through the windows before returning to the forest”.

Q12. Complete Paragraph Five with the best word or phrase:

- A. Furthermore      B. For example      C. As expected
- D. In other words

Most of the cohesive devices used in this series of tests are taken from a list (see Table 3) adapted from Nordquist (2020). Some of the devices on

**Table 3**

*List of cohesive devices from Nordquist (2020)*

Type of transitions	Examples
Addition	also, furthermore, in addition, moreover, to begin with, next, finally, first, second, third, to make matters worse.
Cause-effect	as a result, for this reason, therefore, as expected
Comparison	in the same way, similarly
Contrast	but, fortunately, however, in contrast, nevertheless, on the other hand, unfortunately
Summary	at last, finally, in conclusion
Example	for example, for instance
Insistence	in fact, indeed
Place	above, beneath, behind, in front, on top of, under, to the left, to the right
Restatement	in other words, in short
Time	at the same time, currently, in the future, in the meantime, in the past, then, until now.

the original list, *notwithstanding*, for example, were removed because they were judged to be infrequent lexical items and likely to be unknown to intermediate or lower intermediate level learners. Even here, it is worth

teaching the phrases to test-takers before the test to limit the possibility that items designed to test learner understanding of text cohesion are simply testing vocabulary knowledge. One rule to follow in writing such items is to ensure that the correct answer and the three distractors each belong to a different type of transition. For example, in Question 12 above, the distractors and correct item represent the following four transition types: addition, example, cause-effect, and restatement. Typically, learners find MC questions in the cohesive device category more challenging than items testing referents. This is probably because they must process a larger quantity of text, or a larger proportion of the reading passage to answer the question. Alternatively put, these items may wield greater discriminatory power than questions on referents that require the reader to process more localized information.

The third and final category aims to test the reader's ability to guess the meaning of unknown words from context in a process also known as lexical inferencing or lexical guessing. Anderson (1999) claims: "L2 readers should be taught to use context to effectively guess the meanings of less frequent vocabulary" (p. 25). The process he recommends for doing this is based on a 5-step process described by Nation (2008) and begins with deciding its part of speech (p. 75-76) and following this with close examination of the context in the search for useful clues to meaning. These clues can exist in the same sentence as the unknown word, or further afield. As with the referents, the vocabulary guessing items are underlined and numbered in the reading passage as in the following example from the sample test. "In the panic that followed the cooking pot next to the <sup>(11)</sup> hearth was overturned, extinguishing the flames, and the oil lamp was put out as well, plunging the house into darkness".

Q11. In Paragraph Four, the word hearth probably means . . .

- A. a fireplace.      B. a table.      C. a shelf.      D. a window.

Needless to say the three distractors and the correct answer should all be in the same part of speech as the underlined word, which should also be highly unlikely to be known to the test-takers. In evaluating candidate vocabulary items for test-takers in Japan, it is important to ensure that the selected items do not exist as loan words in Japanese. Indeed, the author had selected the word “aggressive” for Question 14 of the sample test before noticing that the word had recently come to be used in soccer commentary. It was therefore replaced with the word “hostile”.

## **Conclusion**

A final step in the creation of a reading test would normally involve trialing the test with a group of learners who are similar in level to the target test-takers. Following trialing, test answers should be inputted into a spreadsheet and submitted for item analysis using tools available on SPSS, for example. The main aim of the analysis is to determine the degree of item effectiveness, and test validity and reliability, but this was considered beyond the scope of this paper at the time of writing. With the practice test in the Appendix, two trials were conducted and one item (identifying referents) was replaced after the first trial because it was answered successfully by all the learners in the group taking the test. In this case, it was therefore possible to identify a failed item without the use of software. Further trials are required to confirm validity because a problem with one item mentioned earlier (guessing the meaning of unknown words), causing it to be replaced, was not discovered until after the second trial. Although the second trial produced a promising and satisfactory range of scores, it needs to be said that the validation process of the sample test has barely begun. Nevertheless, the test serves its intended purpose of illustrating the test writing process described in this paper, and should also be useful for

teaching students about the format of the test.

In overview of the pursuit of effective test-writing Cohen (1994) concludes: “It is very difficult to write successful items” (p. 231). This is especially true if, as pointed out on more than one occasion in previous sections, the test writer does not preview the reading passage for availability of suitable information with which to craft questions, particularly MC items. The key question is whether it is possible to simplify the process of writing useful reading tests through AI. In a previous section, it was reported that Saito (2023) had used ChatGPT to create a reading passage, reading comprehension and discussion questions (see Figure 1 for the specifications used).

### **Figure 1**

*Specifications for ChatGPT creation of a reading passage, reading comprehension and discussion questions for lower intermediate level learners of English*

<p>Step 1: Write a 300-word article about Shohei Ohtani and set the level to be for A2 English learners.</p> <p>Step 2: Make a vocabulary list of ten words from the article that A2 English students should learn with a format of a table including English, Japanese, and other example sentences., Write the first letter of English words as a capital letter. Write examples which are not related to the article. Make “fill in” questions about the vocabulary.</p> <p>Step 3: Provide 5 comprehension questions with four answer choices each about the article.</p> <p>Step 4: Provide correct answers for each question and reasons for the answers.</p> <p>Step 5: Please explain the correct answers and the reasons in Japanese.</p> <p>Step 6: Provide 5 discussion questions about the article.</p>
---

Creating a useful reading test like the one included in the Appendix, and with the same specifications, would almost certainly require a much longer list of steps than in Figure 1. However, the possibility that effective tests can be created with AI is one that needs to be explored in further research.

### **References**

Alderson, J. C. (1996). The testing of reading. In C. Nuttall (Ed.), *Teaching*

- Reading Skills in a foreign language* (pp. 212-228), Macmillan.
- Alderson, J. C., Haapakangas, E.-L., Huhtis, A., Nieminen, L., & Vllakonaja, R. (2015). *The diagnosis of reading in a second or foreign language*. Routledge, Taylor & Francis Group.
- Anderson, N. J. (1999). *Exploring second language reading. Issues and strategies*. New York: Newbury House Teacher Development. Heinle & Heinle. (International Thomson Publishing Company).
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F. & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Bailey, K. M. (2018). Multiple-choice item format. *The TESOL Encyclopedia of English Language Teaching*, 1-8. <https://doi.org/10.1002/9781118784235.eelt0369>
- British Council. (2022, October 15). Electroreception. Retrieved from: <https://takeielts.britishcouncil.org/take-ielts/prepare/free-ielts-practice-tests/reading-academic/section-1>
- Brown, J. D. (2005). *Testing in language programs*. New York: McGraw-Hill.
- Brown, J. D., & Hudson, T. (2003). *Criterion-referenced language testing*. Cambridge University Press.
- Carrell, P. L., Wise, T. E. (1998) The relationship between prior knowledge and topic interest in second language reading. *Studies in Second Language Acquisition* 20: 285-309
- Cohen, A. D. (1994). *Assessing language ability in the classroom*. Heinle & Heinle.
- Day, R., & Park, J. (2005). Developing reading comprehension questions. *Reading in a Foreign Language*, 17(1), 60-73.
- Eidswick, J. (2010). Interest and Prior Knowledge in Second Language Reading Comprehension. *JALT Journal*, 32(2), 149-168.
- Kincaid, J. P., Fishburne, J., Robert P., R., Richard L., C., & Brad S. (1975). Derivation of new readability formulas (automated readability index, fog count and Flesch reading ease formula) for Navy enlisted personnel. <https://doi.org/10.21236/ada006655>
- Good Calculator. (2022, October 14). Retrieved from <https://goodcalculators.com/flesch-kincaid-calculator/>

- McNamara, T. F. (2000). *Language testing*. Oxford University Press.
- Nation, I.S.P. (2008). *Teaching vocabulary: Strategies and techniques*. Heinle.
- Nordquist, R. (2020). *Cohesion strategies: A list of transitional words and phrases*. ThoughtCo. Retrieved November 17, 2022, from <https://www.thoughtco.com/transitional-words-and-phrases-1690557>
- Nuttall, C. (1996). *Teaching reading skills in a foreign language*. Macmillan Heineman.
- Saito, Y. (2023, May 14). *Exploring the Potential of Metaverse and ChatGPT in ELT* [Paper presentation] JALT PanSIG Conference 2023, Kyoto Sangyo University, Kyoto, Japan.
- Sankebetsu brown bear incident. (2022, October 6). In *Wikipedia*. [https://en.wikipedia.org/wiki/Sankebetsu\\_brown\\_bear\\_incident](https://en.wikipedia.org/wiki/Sankebetsu_brown_bear_incident)
- Schmitt, N. (2002). *An introduction to applied linguistics*. Arnold.
- Wallace, C. (1992). *Reading*. Oxford University Press.
- Weir, C., Porter, D., & Green, R. (1999). *Language Testing: Materials*. Reading, UK: The University of Reading.



**Appendix: Practice Reading Exam (Sample Test)**

Instructions: The reading passage is followed by 20 questions. Read the passage and write your answers on the answer sheet. You have 30 minutes to complete the test.

① The Sankebetsu brown bear incident of December 1915 was the worst bear attack in Japanese history. Events began at dawn in mid-November 1915 when the bear first appeared at the home of the Ikeda family in Sankebetsu, near the west coast of Hokkaido. The surprise encounter threw the family horse into a panic, but the bear ran away after helping itself to harvested corn. When the bear reappeared on November 30, a local hunter shot and hit it, but failed to kill it. Nine days later, the giant brown bear turned up at the home of the Ota family attacking Abe Mayu and killing a baby she was taking care of. Mayu fought back, apparently by throwing firewood, but she was overtaken, knocked down, and dragged into the forest when she tried to escape.

② The next day a group of thirty men entered the forest and had advanced no more than 150 meters when they came across the bear, later named Kesagake. Five men shot at it, but only one managed to land a hit, whereupon the angry animal retreated, and the men escaped injury. Later, the hunters searched the area and discovered dried blood on the snow next to a tree. Beneath the snow were the half-eaten remains of Mayu. The bear had buried the corpse there to preserve it, as well as to keep it safe from other animals.

③ Believing the bear would come back, several villagers gathered at the Ota family's home with guns. Around eight o'clock on the night of

December 12, the bear reappeared and one man managed to shoot at it but missed his target. Hearing gunfire, a group of armed guards staying 300 meters away at the Miyoke house soon arrived, but the bear had vanished into the woods. The group decided to follow it and headed down a river on what they thought was Kesagake's trail. Meanwhile, when women and children from the neighborhood first received news of the attack at the Ota family home, they sought safety at the Miyoke family home. (9) , that was where the bear decided to head to next.

4 In the Miyoke family home, Yayo was preparing a late meal while carrying her youngest son, Umekichi, on her back. She heard a rumbling noise in the garden, but before she could determine what (10) it was, the bear broke in through a window and entered the house. In the panic that followed the cooking pot next to the (11) hearth was overturned, extinguishing the flames, and the oil lamp was put out as well, plunging the house into darkness. Yayo tried to escape, but Kesagake attacked her and bit Umekichi. Although Yayo survived, Umekichi died later from his wounds. Odo, the only guard who remained at the house, attempted to take cover behind furniture but was clawed in the back. The bear then attacked Kinzo, the third son of the Miyoke family, followed by his friend Haruyoshi, killing them both. It then bit Haruyoshi's brother Iwao, who survived with serious injuries. Next, Take, who was pregnant, was attacked and killed. Later her unborn child was retrieved alive from her body but died shortly after. When the group of guardsmen finally arrived back from the riverside, Kesagake escaped into the night.

5 It was decided that the bear would probably try to retrieve the bodies of those it had killed. A new approach was therefore proposed to lure the bear back to the Miyoke house using the body of a victim. At first,

the idea was widely condemned, especially by the families of the victims, but they finally agreed that for the future of the village it was the best course of action. On December 13, a team of six hunters waited inside the house. [12] , the bear reappeared, but stopped and seemed to check the inside of the house by looking through the windows before returning to the forest.

[6] On the morning of December 14, a hunter named Yamamoto spotted the bear resting by a tree and approached within 20 meters. His first shot hit the bear's heart and the second shot hit his head, killing the animal which weighed 380 kilograms and was 2.7 meters tall. While many people believed that the attacks occurred due to the bear being woken by hunters during hibernation, others thought that it had woken too early in the winter because it was hungry. Another theory is that brown bears had to search for food close to humans after the land around Sankebetsu had been deforested. However, the bear did not attack anyone until it was shot. It therefore seems more likely that the bear had become <sup>(14)</sup> hostile due to its injuries.

**Questions 1 – 5**

*[1 point each]*

**Paragraph Headings**

*Match each paragraph with a heading. Do not use answer G; it is used in the example. There is one other extra heading that you won't use. Write the answers on the answer sheet.*

Example: Paragraph One Answer: G

List of headings

- A. Attack at dinnertime
- B. The death of Kesagake
- C. Return to the Ota farm
- D. A body is discovered in the woods
- E. The bear is seen by a river
- F. A new plan ends in failure
- G. The first attack (used in example)**

- 1. Paragraph Two
- 2. Paragraph Three
- 3. Paragraph Four
- 4. Paragraph Five
- 5. Paragraph Six

**Questions 6 – 15**

*[1 point each]*

*Choose the correct answer, A, B, C, or D. Fill in the appropriate letter on the marksheet.*

6. How many times was the bear shot by hunters?

- A. Two                      B. Three                      C. Four                      D. Five

7. How many people died as a result of the bear attacks?

- A. Five                      B. Six                      C. Seven                      D. Eight

8. How many people were injured by the bear?

- A. Two                      B. Three                      C. Four                      D. Five

9. Complete Paragraph Three with the best word or phrase:  
A. Similarly            B. As a result            C. In addition  
D. Unfortunately
10. In Paragraph Four, it refers to . . .  
A. the noise.            B. the bear.            C. the garden.            D. the meal.
11. In Paragraph Four, the word hearth probably means . . .  
A. a fireplace.            B. a table.            C. a shelf.            D. a window.
12. Complete Paragraph Five with the best word or phrase:  
A. Furthermore            B. For example            C. As expected  
D. In other words
13. The bear was killed . . . . . days after the attack on the Ota family home.  
A. two            B. five            C. nine            D. fourteen
14. In Paragraph Six, the word hostile probably means being . . .  
A. weak and wanting to sleep.  
B. angry and wanting to attack.  
C. afraid and wanting to hide.  
D. hungry and wanting to eat.
15. The author believes the bear attacked humans because . . .  
A. it was woken by hunters.            C. the area was deforested.  
B. it was very hungry.            D. it had been injured.

**Questions 16 – 20**

*[1 point each]*

*Mark A for TRUE or B for FALSE.*

16. The bear stole some food from the Ikeda family home.
17. When the bear attacked, Odo tried to escape from the Miyoke family home.
18. Mayu tried to defend herself from the bear.
19. After the bear left the Ota family home on December 12, it went down to a river.
20. The bear attacked a horse at the Ikeda family home.

**Answer Guide.**

1. D Paragraph 2
2. C Paragraph 3
3. A Paragraph 4
4. F Paragraph 5
5. B Paragraph 6
6. C Other category questions. The bear was shot once each in paras 1 & 2, and twice in para 6.
7. C Other category questions. Mayu & the baby (para 1), Umekichi, Kinzo, Haruyoshi, Take, & an unborn child (para 4).
8. B Other category questions. Yayo, Odo & Iwao (para 4)
9. D Signal words or expressions.
10. A Pronouns and referent words.
11. A Guessing the meaning of unknown words.
12. C Signal words or expressions.
13. B Inference. The attack on the Ota family home occurred on December 9 (nine days after it returned to the Ikeda family home on November 30). The bear was killed on December 14.
14. B Guessing the meaning of unknown words.
15. D Author's purpose.
16. A (True) Para 1. The bear stole some corn.
17. B (False) Para 4. Odo tried to hide behind furniture.
18. A (True) Para 1. Mayu fought back/she threw firewood at the bear
19. B (False) Para 3. The bear went to the Miyoke family home/The guardsmen headed down a river on what they thought was Kesagake's trail
20. B (False) Para 1. The bear only caused the horse to panic.