

Variables Influencing the Validity of the Reading Test

Li Du

School of Humanities, Binzhou Medical University

Yantai 264003, China

346 Guanhai Street, Laishan district, Yantai

E-mail: duli19722003@yahoo.com.cn

The research is financed by The Backwash of CET on College English Teaching of Binzhou Medical University.

Abstract

The paper analyzes the variables that can influence the validity of the reading part in the standardized test of CET (College English Test). They refer to the test methods and test content. As for the test method, the widely used are multiple-choice questions and short answer questions which can influence the test takers' performance and therefore the validity of the reading part. For the test content, this paper analyzes from the perspectives of text forms and types of questions. It shows that these variables do have some influence on the validity of the reading test.

Keywords: Validity, Test method, Test content

1. Introduction

Many standardized tests in use in China, such as CET (College English Test), have got a necessary part to evaluate the language learners' comprehensive abilities and that is the reading part. This part is to test the learners' reading abilities considered to be very important among many other abilities. In order to do so, the test designers make use of several methods to test the learners that include multiple-choice questions and short answer questions. But can these ways or methods really reflect the test takers' reading abilities? Is the design of the reading part influenced by some variables?

2. Theoretical understanding of reading abilities

In order to analyze the validity of the test, it is necessary to know what validity refers to. Henning (1987, p.89) defines validity as follows: "Validity in general refers to the appropriateness of a given test or any of its competent parts as a measure of what it is purported to measure. A test is said to be valid to the extent that it measures what it is supposed to measure."

According to the definition mentioned above, if we want to prove the validity of the reading part in the CET-4, we should be clear about the population of abilities that we want to test. It is worthwhile specifying these as accurately and completely as possible.

2.1 Theory Division

Different people use the term reading abilities in different ways and much confusion can arise from consequent misunderstandings. Almost all researchers agree that reading does not only refer to the understanding of lexical words, but reading "the lines", reading "between the lines" and reading "beyond the lines". That is to say, we should not only understand the literal meaning of text, but the inferred meaning and have the critical evaluation of text.

2.1.1 Alderson's integrated theory

Alderson (2000, p.11) states that "at least part of the reading process probably involves the simultaneous and variable use of different and overlapping 'skills'. The division of 'skills' ...does not seem to be justified in practice." Although Alderson looks at the reading as the integrated process, he does not mention what are integrated and how they are integrated in the reading process.

2.1.2 Matthews' theory

Citing Eskey and Grabe's view (1988) of the importance of speed and automaticity in word recognition, Matthews suggests that, if speed and flexibility are important, then they need to be tapped in tests of reading. (Alderson, 2000, p.12) Now that it is the test, we should take the reading speed and fluency into consideration when we want to test the learners reading ability. She puts too much emphasis on the reading speed and fluency without suggesting what reading abilities are and so she somewhat deviates from the understanding of the reading abilities.

2.1.3. Carver's three-part separability theory

Carver's view of reading is that "reading should be a three-part separability of word recognition skills, reading rate or reading fluently, and problem-solving comprehension abilities." (Alderson, 2000, p.12). His "problem-solving" strategies may be useful for the resolution of many difficulties in reading, for example the deduction of the meaning of unknown words. But he seems to emphasize the importance of logical inferencing abilities in the reading process. In other words, he explains the reading abilities from the cognitive point of view without taking the social background knowledge of learners into consideration.

2.2 Reading Test of CET-4

Synthesizing all the theories mentioned above, we can conclude that reading abilities are the combination of linguistic competence including all kinds of reading skills, social background knowledge and cognitive process within the limited time. While the objectives of reading part in CET-4 are to test learners' following abilities: extracting the main idea of the text; distinguishing the facts and details of the text; understanding the literal meaning and inferring according to the text and understanding the logical relation of the context. Compared these with the understanding and evaluation of the reading theories, we can find that the designed test cannot fully reflect the learners' reading abilities.

3. Test Methods

There is some evidence in the literature that test format might affect student performance. (Weir, 1990, p.43). It is clear that the test takers' true abilities are not always reflected in the test scores and to a certain extent this is inevitable. But we should try our best to decrease the influence of external factors to achieve the better understanding of examinees' reading abilities. At the same time, we should bear in mind that there is no one "best method" in testing reading. There are many testing methods used to test reading such as multiple-choice questions, short answer questions and cloze. But theorists have different attitudes towards these methods.

3.1 Multiple-choice Questions

A multiple-choice test is usually set out in such way that the examinee is required to choose the answer from a number of given options, only one of which is correct. This method is a common device for testing examinees' reading comprehension. It "allows testers to control the range of possible answers to comprehension questions, and to some extent to control the test takers' thought processes when responding." (Alderson, 2000, p.211)

Perhaps the most obvious advantage of multiple-choice questions is that scoring can be perfectly reliable, objective and rapid and economical as well. When test is carried out on a very large scale, when the scoring of tens of thousands of compositions might seem not to be a practical proposition, it is understandable that potentially greater accuracy is sacrificed for reasons of economy and convenience. (Hughes, 1989) That may be the reason that the large-scale tests such as CET-4 chooses the method of multiple-choice questions. Although the scoring can be objective, the designation of comprehension questions is not objective at all. The reliability is seldom high because there is the general disagreement on the determination of the most important part of the text on which most questions are based.

Hughes (1989) believes that another advantage is that "since in order to respond the candidate has only to make a mark in the paper, it is possible to include more items than would otherwise be possible in a given period of time." His claim is theoretically applicable, but considering the time limit of the test, we cannot include as much items as we expected. On the contrary, in the real standardized test, the number of items is usually unchangeable.

Munby used to believe that "multiple-choice questioning can be used effectively to train a person's ability to think...It is possible to set the distractors so close that the pupil has to examine each alternative very carefully indeed before he can decide on the best answer...When a person answers a comprehension question incorrectly, the reason for his error may be intellectual or linguistic or a mixture of the two. Such errors can be analyzed and then classified so that questioning can take account of these areas of difficulty." (Alderson, 2000, p.203) His positive attitude towards multiple-choice questions is problematic because answering multiple-choice items is an unreal task, as in real life one is rarely presented with four alternatives from which to make choice to signal understanding.

Alderson (1995) believes that it may be easier to control the thought process of readers with multiple-choice techniques than it is with short answer questions. Because it is easier to devise distractors to get readers to think in certain ways and this control may be desirable for the testing of inferencing in the second language. But this also implies that the method tricks the unwary into making incorrect interpretations they might not otherwise have made. It is just likely to test some abilities and not be so good at testing others.

The popular use of multiple-choice questions does not prove its validity. It has many disadvantages when the effects of this method are taken into account.

It is evident that examinees taking multiple-choice tests can learn “strategies” that inflate their scores: techniques for guessing the correct answer, for eliminating implausible distractors, for avoiding two options that are similar in meaning, for selecting an option that is notably longer than the other distractors and so on. (Alderson, 1995). So the scores gained in multiple-choice tests may be suspect. There are many test-coaching schools or classes in China to cultivate the examinees’ test wiseness. What they teach is not how to acquire the language but how to acquire the so-called test-taking techniques. Practice for the test will have a harmful effect on learning and teaching. Practice at multiple-choice items (especially when, as happens, as much attention is paid to improving one’s educated guessing as to the content of the items) will not usually be the best way for students to improve their command of the language. (Hughes, 1989)

This disadvantage is related to the designation of distractors. The most important requirement of a multiple-choice item is that the “correct” answer must be genuinely correct and there should be only one correct answer. But in reality, some dubious answers are often common in inferencing questions. Furthermore, the multiple-choice questions should be presented in context especially when testing reading abilities. The presentation of context will often reduce the possibilities of ambiguity. In fact, it is extremely difficult and time consuming to develop sufficient number of decent items on a passage. Saving in time for administration and scoring will be outweighed by the time spent on successful test preparation.

Others who are against the multiple-choice questions believe that the testers does not know why the test-takers responded the way they did. They may simply guess at their choice, or may have a totally different reason in mind from which the test constructors intended when writing the item, and they may even employ the test-taking strategies to eliminate implausible choices and be left with only choice. We can never know what part of any particular individual’s score has come about through guessing. We cannot overlook the fact that the responses on a multiple-choice test (a, b, c, d) are so simple that make them easy to communicate to other test-takers nonverbally and therefore facilitate cheating.

3.2 Short Answer Questions

In the CET-4, another common method used to test reading abilities is short answer questions. Test takers are simply asked questions which require brief and specific responses in a few words. Alderson (2000) states that the justification for this method is that it is possible to interpret students’ responses to see if they have really understood, whereas on multiple-choice questions students give no justification for the answer they have selected and may have chosen one by eliminating others. For this method, answers are not provided for the students and therefore if the students get the answer right, we are certain that this has not occurred for reasons other than comprehension of the text.

Weir (1990, p.45) mentions that “with careful formulation of the questions a candidate’s response can be brief and thus a large number of questions may be set in this format, enabling a wide coverage.” This statement may be true as far as the development of the reading abilities are concerned, but for the testing, the number of the questions is often limited and not so applicable. He also states that “activities such as inference, recognition of a sequence, comparison and establishing the main idea of a text can be done effectively through short answer questions where the answer has to be sought rather than being one of those provided.” It can test the test-takers’ reading skills mentioned above.

Weir (1990) also states that the use of long texts with short answer method is more representative of required reading in the target situation, at least in terms of length. They can also provide more reliable data about the candidates’ reading abilities. But whether the tasks designed are really related to the real life reading is doubtful.

Compared with multiple-choice questions, this method is more subjective in that it requires the test designers’ subjective attitude both in designing and scoring the test papers. The questions set in this method normally try to cover the important information in a text (overall gist, main ideas and important details) and understanding of structures and lexis that convey this. (Weir, 1993) But it is difficult for the test designers to decide which part covers the important information and which structure or lexis should be tested. The subjectivity of the decision on the formation of the questions will influence the validity of the test. Bernhardt (1991) states that “when the questioner asks the questions, the reader probably rejects his view of the text and therefore shifts his understanding. The original interaction between reader and text is interfered by the testers or the questioners’ questions and therefore may have effect on the understanding of the text.” Moreover, the questions are not easy to construct in that all possible answers should be foreseeable because individual test-takers may have different understandings of the text and therefore have different answers. Otherwise the marker will be presented with a

wide range of responses that he will have to judge as to whether they demonstrate understanding or not. (Alderson, 2000, p.227) So there is the possibility that the variability of answers might lead to marker unreliability.

According to Weir (1990), the main disadvantage of this method is that it involves the candidate in writing and there is some concern that this interferes with the measurement of the intended construct. That is to say, the method tests not only the reading abilities but writing as well. Test-takers may be able to understand the text but not be totally clear what is expected of them in the transfer phase. The guiding principle to avoid this is to keep the answers brief and to reduce writing to a minimum to avoid possible contamination from students having to write answers out in full.

Although these two methods have their own advantages and disadvantages, there are some similarities between them. They are all a series of related items based on the same reading passage. Such format makes it possible to measure examinees' understanding of the material from various perspectives and at the same time is cost-effective both for item writers and examinees.

4. Test Content

Besides the test method, there are still many other factors influencing the validity of the test and one of them is the test content. In Hughes's (1989, p.22) words, "A test is said to have content validity if its content constitutes a representative sample of the language skills, structures, etc. with which it is meant to be concerned." While Weir believes that the more a test simulates the dimensions of observable performance and accords with what is known about that performance, the more likely it is to have content validity. A comparison of test specification and test content is the basis for judgments as to content validity and the variety if the text can increase the content validity as well.

4.1 Text Forms

Insert Table 1 Here

The table quoted above is the classification of texts. From Weir's point of view, tests should include texts that mirror those as closely as possible that students have been exposed to and are likely to meet in their future target situations. But in reality, tests usually "involve short passages about unfamiliar topics that rarely approximate authentic texts and literal-level, direct content questions in multiple-choice formats." (Bernhardt, 1993, p.192)

4.2 Types of Questions

There are many types of questions used in testing reading abilities. The following is what Christine Nuttall (1982) classifies the type of questions.

Questions of literal comprehension: These are questions whose answers are directly and explicitly available in the text. Questions of this kind could often be answered in the words of the text itself.

Questions involving reorganization or reinterpretation: These are questions that require students to obtain literal information from various parts of text and put it together, or to interpret information. Such questions are valuable in making students consider the text as a whole rather than thinking of each sentence on its own.

Questions of inference: These are questions obliging students to read between the lines, to consider what is implied but not explicitly stated. These questions are more difficult because they require students to understand the text well enough to work out its implications. The difficulty is intellectual rather than linguistic in most cases.

Questions of evaluation: Evaluative questions involve the reader in making a considered judgment about the text in terms of what the writer is trying to do and how far he has achieved it.

Questions of personal response: Of all the types of questions, the answer to this type depends most on the reader and least on the writer. The reader is not asked to assess the techniques by means of which the writer influences him, but simply to record his reaction to the content of the text.

Understanding the classification of the type of questions can help us make clear what are tested by these questions. But the test difficulty should not lie in the questions but in the text, and the language of the questions should be easier than the text. Conversely, the questions should not be so simple or answerable from world knowledge that makes it possible for examinees to answer them without looking at the texts. Furthermore, the interdependence of items should be avoided and answering one question should not be dependent on ability to answer another.

5. Conclusion

From the analysis above, it is clear that these variables or factors do have some influence on the validity of the

reading test. The reading part of CET is no exception. In fact, there is no perfect method to apply to test the examinees' reading ability and the possible solution to this problem is to combine different methods and reduce the subjective influence as much as possible.

References

- Alderson, J. Charles. (1991) *Language Testing in the 1990s*. London: Modern English Publication in association with the British Council.
- Alderson, J. Charles. (1995) *Language Test Construction and Evaluation*. Cambridge: Cambridge University Press.
- Alderson, J. Charles. (2000). *Assessing Reading*. Cambridge: Cambridge University Press.
- Bachman, Lyle. F. (1991) *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bachman, Lyle. F. (1995). *An Investigation into the Comparability of Two Tests of English as a Foreign Language*. Cambridge: Cambridge University Press.
- Bachman, Lyle. F and Andrew, D. Cohen. (1998). *Interfaces Between Second Language Acquisition and Language Testing Research*. Cambridge: Cambridge University Press.
- Bernhardt, Elizabeth, B. (1991). *Reading Development in a Second Language: Theoretical, Empirical, and Classroom Perspectives*. New Jersey: Ablex Publishing Corporation.
- Clapham, Caroline. (1996). *The Development of IELTS: A Study of the Effect of Background Knowledge on Reading Comprehension*. Cambridge: Cambridge University Press.
- Green, Alison. (1998). *Verbal Protocol Analysis in Language Testing Research: A Handbook*. Cambridge: Cambridge University Press.
- Henning, Grant. (1987). *A Guide to Language Testing: Development, Evaluation, Research*. London: Newbury House.
- Hughes, Arthur. (1989). *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- Kunnan, Anthony, J. (1995). *Test Taker Characteristics and Test Performance: A Structural Modeling Approach*. Cambridge: Cambridge University Press.
- Nuttall, Christine. (1982). *Teaching Reading Skills in a Foreign Language*. London: Heinemann International.
- Weir, Cyril, J. (1990). *Communicative Language Testing*. London: Prentice Hall.
- Weir, Cyril, J. (1993). *Understanding and Developing Language Tests*. London: Prentice Hall.

Table 1. Text forms. DIALANG Assessment Specifications for Reading Comprehension, Version 6 (Alderson, 2000, p.127)

Text forms	Examples (text types)	
Descriptive	impressionistic descriptions	e.g. travel accounts
	technical descriptions	e.g. reference books
Narrative	stories, jokes	
	reports: biographical notes, news, historical accounts	
Expository	definitions	brief, one-line dictionary definitions
	explications	broader accounts of esp. phenomena e.g. newspaper articles, educational materials
	outlines	e.g. initial abstract, introductory paragraph
	summaries	...of phenomena, e.g. in an encyclopedia
	text interpretations	e.g. book review
Argumentative	comments	e.g. newspaper leader, letter-to-the-editor, column, book/film review
	formal argumentation	e.g. scientific articles
Instructive	personal instructions	e.g. signs, notes
	practical instructions	e.g. signs, recipes, technical instructions
	statutory instructions	e.g. directions, rules, regulations, law text