

On the De-Common-Sense Validity of the Multiple Choice Questions in the Reading Comprehension of College English Test Band 4

Jun Liu

College of Foreign Languages, North China Electric Power University, Beijing, P. R. China

To Link this Article: <http://dx.doi.org/10.6007/IJARPED/v3-i1/559>

DOI:10.6007/IJARPED/v3-i1/559

Published Online: 21 March 2014

Abstract

The researcher selected 10 multiple choice questions from the Reading Comprehension of College English Test Band 4, and asked 102 freshmen and 126 postgraduates to answer the questions without reading the source passages. The result shows that the percentage of the correct choice of each of the 7 multiple choice questions out of 10 exceeds the average percentage of the 4 choices for each question. This suggests that the students' common sense does play a role when they answer the multiple choice questions of Reading Comprehension. The researcher, therefore, proposes a new concept "De-Common-Sense Validity" and discusses two techniques for the improvement of this validity.

Keywords: De-Common-Sense Validity, Reading Comprehension, Multiple Choice Question, College English Test Band 4

Introduction

In scientific research, the term validity refers to the extent that a study is capable of addressing the questions that it is intended to answer. Validation is of great importance, but extremely difficult, since the judgment of validation entails the judgment of the intended answers as well as that of the designed questions. In theories of testing, validity is often described as "Pandora's Box", "the fantastic trove of complex interconnected problems (Markus & Borsboom, 2013, p.1)." In testing theories of second language education, it is now universally acknowledged that the validity of a test for second language learners is the most important criterion for its quality. A test with high validity is such that its items are adequately the intended focus of the test and in proportion to the importance and value of each individual item that it is claimed to deserve. If we can justify that a certain test does not measure what it claims to measure, then we say it has a validity deficiency in the design of the test. Deficiencies may lie at different levels. This paper examines the multiple choice questions in the test item of reading comprehension, which is popular in almost all large-scale tests for second language learners. Occasionally, test takers have no time to finish the reading passage, but they have sufficient time to read the questions and guess the answers. An ideal question is such that test takers cannot choose the correct answer out of pure guess. But more often than not the questions of the reading passage are about the common sense of test takers, and they can easily guess the correct answers.

Testing researchers (e.g. Bachman, 1990; Bachman & Palmer, 1996; Castello, 2008) have realized that the problem of validity involves individual differences of the test takers, especially the differences of their knowledge and experience. These are the external factors of test validity, and it is hard for test designers to consider these external factors into account when they design foreign language tests. In order to address this issue, this research proposes a new concept --- de-common-sense validity. The author studied some test papers of College English Test Band 4 (CET-4), which is administered nationwide twice a year, and analyzed the multiple choice questions of the item of Reading Comprehension so as to justify the concept and improve the validity of tests for learners of English as a foreign language.

De-Common-Sense Validity

De-common-sense validity of a foreign language test refers to the degree of the avoidance of the common sense that test takers attempts to resort to when they answer the multiple choice questions for reading comprehension. Whether an item measures what it is intended to measure depends upon whether, and if so, how much, it avoids the common sense of test takers. "Common sense" in this context means a test taker's general knowledge and experience about the world. For instance, in a passage that discusses the independence of the United States and mentions the Independence Day. If one of the questions designed is to ask test takers to tell the date of the celebration of the Independence Day, then this question is invalid since it can be answered by resorting to the general knowledge, or common sense, of test takers, without reading the passage.

Teachers and students often come across this kind of reading comprehension exercise: a math problem for the learner to solve. If the learner can understand the problem and the questions, he can provide correct answers; if he cannot understand, he will be at loss and leave the questions unanswered. It seems that the learner who cannot understand the problem and questions is poor in English, but if the learner's prior knowledge and experience are taken into account, we can see that he cannot understand the problem and the questions because he lacks mathematical ability. Or even if he can understand everything about the math problem and the questions, it is most likely that he still cannot solve the problem. If the English proficiency of the learner is to be measured by the degree of the correct solution to the problem, then the target may well be missed. It is one thing to understand the problem and the questions; and it is quite another to solve the math problem. Therefore, we say that this test is not scientific and its de-common-sense validity is low. Extreme as the examples are, they serve perfectly well to show what the concept of de-common-sense validity is. A test of high de-common-sense validity is such that a test taker will get a wrong answer to the questions in reading comprehension if he only resorts to his common sense.

The purpose and aim of CET-4 is to objectively and accurately measure the English proficiency of the English learners in colleges and universities in China. We all know that the learner's linguistic knowledge and competence can only be measured by his linguistic performance (listening, speaking, reading, writing or translation). The learner's linguistic performance cannot be conducted without the application of his prior general knowledge and experience, and consequently it becomes difficult to test the learner's English proficiency without employing his prior knowledge and experience. However, it is one thing to demonstrate the linguistic proficiency in passage reading while using the prior knowledge and experience; it is quite another to give a correct answer to a given question by using the common sense. Special attention should be paid to the de-common-sense validity of the questions. Test designers should avoid questions that can be answered by the common sense of test takers.

There are problems of de-common-sense validity in almost all the items in CET-4. For instance, Question 20 in the Listening Comprehension of the CET-4 of June, 2005: What is the chief advantage of having the engagement period? [A] The two people can learn about each other's likes and dislikes. [B] The two people can have time to decide if they are a good match. [C] The two people can have time to shop for their new home. [D] The two people can earn enough money for their wedding. The correct answer is [B]. If a text taker (who was brought up in the Chinese cultural circumstances) could not understand the listening passage, his prior knowledge and experience would enable him to use his common sense for the correct answer. To most people in China, the engagement period before the formal marriage is a time for the two people to formally and publicly claim their relationship and to stay together in order to see if they can be a good match. Therefore the degree of the overlap between the test takers' prior knowledge needed to answer the question and their common sense is high, and hence the de-common-sense validity is low.

Analysis and Discussion of Some Questions in the Reading Comprehension of CET-4

In order to examine the problem of de-common-sense validity in CET-4, I went over the Reading Comprehension questions of the CET-4 papers across a number of years. The reason why I chose the questions in Reading Comprehension is that I thought the problem there might be prominently displayed, though problems also exist in Listening Comprehension, Cloze and Composition. When I studied the tests, I attempted to compare the correct answers to the questions with my own knowledge and experience. I finally picked out 10 questions and formed a questionnaire, which was administered to 102 first-year freshmen and 126 postgraduate students in North China Electric Power University. In the questionnaire, the students were not given the reading passages but only the questions with multiple choice answers. They were required to choose what they thought might be the most appropriate answer to each question. The results of the collected data are as follows:

Table 1

Data from the Freshmen

Questions	1	2	3	4	5	6	7	8	9	10
Question No. in the Test	26 Jan. 2002	16 Jan. 1998	27 Jan. 1999	31 Jan. 1999	36 July 2000	28 July 2002	32 Jan. 2002	38 Jan. 2002	39 Jan. 2002	30 July 2003
Answers	A	B	D	D	D	C	B	B	A	D
Percentage	33%	47%	14%	65%	63%	51%	29%	39%	8%	22%

Table 2

Data from the Postgraduates

Questions	1	2	3	4	5	6	7	8	9	10
Question No. in the Test	26 Jan. 2002	16 Jan. 1998	27 Jan. 1999	31 Jan. 1999	36 July 2000	28 July 2002	32 Jan. 2002	38 Jan. 2002	39 Jan. 2002	30 July 2003
Answers	A	B	D	D	D	C	B	B	A	D
Percentage	70%	59%	28%	65%	84%	68%	40%	40%	20%	13%

Table 3

Data Comparison between the Freshmen and the Postgraduates

Questions	1	2	3	4	5	6	7	8	9	10
Question No. in the Test	26 Jan. 2002	16 Jan. 1998	27 Jan. 1999	31 Jan. 1999	36 July 2000	28 July 2002	32 Jan. 2002	38 Jan. 2002	39 Jan. 2002	30 July 2003
Answers	A	B	D	D	D	C	B	B	A	D
Freshmen	33%	47%	14%	65%	63%	51%	29%	39%	8%	22%
Postgraduates	70%	59%	28%	65%	84%	68%	40%	40%	20%	13%
Compared	+37%	+12%	+14%	0	+21%	+17%	+11%	+1%	+12%	— 9%

The above tables of the collected data show:

(1) Problems of de-common-sense validity do exist in the multiple choice questions in the Reading Comprehension of CET-4. The average percentage for each choice ought to be 25%, but the percentage of correct answers to 7 questions out of 10 (except questions 3, 9 and 10) is above 25%. The result is consistent with our expectations. Different percentage indicates difference in the degree of the correlation of the respondents' common sense with the meaning of the questions. Take Questions 4 and 5. The percentage of correct answers of both the freshmen and the postgraduates is above 60%, which indicates high correlation between the common sense and the meaning of the questions. Question 4 is: "Advertising can persuade the consumer to buy worthless products by _____." The four choices are: [A] stressing their high quality, [B] convincing him of their low price, [C] maintaining a balance between quality and price, and [D] appealing to his buying motives. The common sense of the students enabled them to choose [D]. Question 5 is: "The chief function of a uniform is to _____." The four choices are: [A] provide practical benefits to the wearer, [B] make the wearer catch the public eye, [C] inspire the wearer's confidence in himself, and [D] provide the wearer with a professional identity. A uniform is a particular type of clothing worn by all the members of a group or organization. Its function is to characterize the wearer as a member belonging to a certain organization, that is, the professional identity.

The percentage of the correct choice of Question 10 is lower than the average percentage (25%). Question 10 is: "A company's efforts to keep expenses low and profits high may result in _____." The four choices are: [A] reduction in the number of employees, [B] improvement of working conditions, [C] fewer disputes between labor and management, and [D] a rise in workers' wages. A company's efforts to keep expenses low and profits high may result in any of the four choices, but to a Chinese student, the probability of [A] is higher than the corrected answer [D], and therefore, 59% of the postgraduates chose [A].

(2) The degree of de-common-sense validity is related to the amount of knowledge of the individuals. Test takers have a certain amount of common knowledge, but individual differences do exist. The percentage indicates the number of test takers who chose the correct answers, and also indicates that these test takers share the same common sense when confronted with the question. More people chose correct answers to 7 questions out of 10, which shows that they share the same prior or background knowledge. If a question evokes the prior knowledge of the test takers and they can judge the answer by their common sense, then the question is bound to have a low validity of de-common-sense. Although the

percentage of the correct answers for 7 questions is higher than the average, it is far from 100%. This indicates that test-takers differ in their knowledge and judgment based on their common sense.

(3) In most cases, the percentage of the correct answers of the postgraduates is generally higher than that of the freshmen (except No. 4 and No. 10). This indicates that more postgraduates used their common sense. This agrees with our expectation before the research: students' knowledge and common sense judgment improve with their experience and education.

Measures to Improve the De-Common-Sense Validity

An effective way to increase the de-common-sense validity is to reduce test takers' reliance on their common sense when answering questions. The avoidance of test takers' reliance on their common sense when answering multiple choice reading comprehension questions does not mean that test designers should avoid test takers employing their prior knowledge in their reading process, which is inevitable. Rather, test takers' prior knowledge should be used to lead to the wrong choice if they answer the questions merely based on their common sense. The concept of de-common-sense validity is of great importance, both theoretically and practically. It is one of the criteria for the measurement of the quality of a test of any kind. With this criterion, the validity of the reading comprehension questions in language test will improve. With the above analysis and discussion, we suggest two techniques to raise the de-common-sense validity.

(1) Choosing reading materials that is new, or differ from or contrast to the common sense. This means that the reading materials should be beyond test takers' common sense knowledge.

(2) Designing questions that avoid test takers' common sense knowledge. In order to improve the de-common-sense validity, test designers will have to avoid those questions that test takers can answer merely by resorting to their common sense. For instance, the 31st question of the June 2005 CET-4 is: "According to the passage, the NCC found it outrageous that ____." The four choices are: [A] all the products surveyed claim to meet ISO standards; [B] the claims made by products are often unclear or deceiving; [C] consumers would believe many of the manufactures' claim; [D] few products actually prove to be environment friendly. Test takers have no idea of the organization NCC, and they have no access to the correct choice just from the four given choices. All the information comes from the reading passage. Without reading the passage, test takers will be at loss as to what the correct answer is. Another example is the 28th question of the June 2002 CET-4. The question is: "A note in the pocket can hardly serve as a reminder because ____." The four choices are: [A] it will easily get lost; [B] it's not clear enough for you to read; [C] it's out of your sight; [D] it might get mixed up with other things. This is a typical example of inference from the common sense. There is a high coincidence between the correct answer and the common sense of the test takers, and hence low de-common-sense validity. The common sense tells us that a note in the pocket often fails to serve as a reminder because it's out of sight. This is why we often write an importance message on the palm of a hand or place the note where we can see.

Corresponding Author

Jun Liu: College of Foreign Languages, North China Electric Power University, No 2, Beinong Road, Huilongguan, Changping District, Beijing 102206, China. E-mail: liuj59@126.com

References

- Castello, E. (2008). *Text Complexity and Reading Comprehension Tests*. Bern: Peter Lang AG, International Academic Publishers.
- Markus, K. A., & Borsboom, D. (2013). *Frontiers of Test Validity Theory: Measurement, Causation, and Meaning*. New York, NY: Routledge.
- Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bachman, L. F., and Palmer, A. S. (1996). *Language Testing in Practice*. Oxford: Oxford University Press.