Analyzing Characteristics of the Input and the Expected Response in TEM-4 Reading Comprehension

Enyao LI
School of Foreign Studies, Guangzhou University of Chinese Medicine
Guangzhou, Guangdong, China

Keywords: Content Validity, TEM-4, Reading Comprehension.

Abstract. Reading comprehension, as an important part of Test for English Majors-Band 4 (TEM-4), plays an important role in the English test. Drawing on the framework of Task Characteristics proposed by Bachman and Palmer (1999), this paper attempts to explore the content validity in TEM-4 (2009) reading comprehension in accordance with the testing syllabus. Through analyzing characteristics of the input and the expected response, this study shows that TEM-4 reading comprehension has relatively high content validity and the contents are internally and basically consistent with the requirements of the TEM-4 testing syllabus.

1. Introduction

TEM-4 (Test for English Majors-Band 4), as a national standardized test of English proficiency administered to English majors at the end of their foundation stage, has been increasingly gained influence in recent years. TEM has been implemented since 1991 and has already become one of the criteria for judging English teaching and learning in China. Owing to its scientific approach, consistent marking, rigorous administration and comparable scores, TEM is now well established as a large-scale standardized test that has reached international standards of educational assessment and is widely recognized in China. Reading, an essential skill in language learning, is significant in English language teaching and learning, especially for English majors. The syllabi of English teaching in secondary schools and colleges (both English majors and non-English majors) treat reading competence training as a major goal of English teaching. Meanwhile, reading comprehension, as one of the main methods to acquire a language, is regarded to be vitally important all the time.

Reliability and validity are essential terms in the language testing. The reliability of a language test is easy to be demonstrated by statistical analysis because this concept is relatively stable whereas the quality of validity is much more sophisticated than reliability [1]. A test is said to be valid if it measures accurately what it is intended to measure [2]. Alderson, Clapham and Wall divide validity into internal and external ones, of which internal validity includes face validity, content validity, response validity while external validity consists of concurrent validity and predictive validity [3]. Content validity is concerned with whether or not the content of the test is sufficiently representative and comprehensive for the test to be a valid measure of what it is supposed to measure [4]. Hughes proposes that “the greater a test’s content validity is, the more likely it is to accurately measure what it is supposed to measure” [2]. A common way to analyze the content validity of a test is to compare it with a statement of what the content ought to be. The content statement may be test specifications, a formal teaching syllabus or curriculum, or a domain specification [3].

In view of the significant role of reading comprehension ability in TEM-4, the present study conducts a case analysis of the content validity in TEM-4 (2009) reading comprehension with an aim to provide helpful implications for improving the assessment quality and developing effective reading tests.

2. Analyzing Content Validity of TEM-4 Reading Comprehension

Drawing on the framework of Task Characteristics proposed by Bachman and Palmer [1], this
paper attempts to explore the content validity in TEM-4 (2009) reading comprehension in accordance with the TEM-4 testing syllabus [5]. As Bachman and Palmer suggest, it is not their purpose to prescribe any particular types of test tasks or combination of task characteristics [1]. Taking account of the nature of reading and characteristics of reading tests, this paper focuses on the analysis of “Characteristics of the Input” and “Characteristics of the Expected Response” in TEM-4 reading comprehension.

2.1. Characteristics of the Input

Input consists of “the information contained in a given test tasks, to which the test taker is expected to respond” [6]. The characteristics of input involve length, speed, new words, readability, topics and genres [6]. In the present study, they are classified into three categories, reading speed (including length and speed), language difficulties (including new words and readability) and text types (including topics and genres).

2.1.1. Reading Speed. The question that almost all testers will consider is how long a reading test should be. Considering the fact that testees not only read the passages but also answer the questions and make choices, the test designer should make a proper proportion of reading the passages and answering the questions [7]. According to the TEM-4 testing syllabus, the total input amount of reading passages is about 1800 words, which is required to finish within 25 minutes [5]. Besides, reading speed should be around 120 words per minute [5]. It is obvious that TEM-4 reading requires both accuracy and speed. With the aid of computer counting, we summarize the number of words in each passage as well as the total input amount and reading speed in the table below.

<table>
<thead>
<tr>
<th>Total Words</th>
<th>Words for Questions</th>
<th>Words for Direction</th>
<th>Total Length</th>
<th>Time Limit</th>
<th>Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1843</td>
<td>758</td>
<td>41</td>
<td>2642</td>
<td>25 minute</td>
<td>106 words</td>
</tr>
</tbody>
</table>

As shown in Table 1, total words for four passages are 1843, which ensures the total length of about 1800 words required in testing syllabus. The reading speed (total length divided by reading time) is 106 words per minute, which almost meets the requirement of testing syllabus that the reading speed should be 120 words per minute and can also guarantee that testees have enough time to think about those questions. On the whole, reading speed (including length and speed) is in accordance with the testing syllabus.

2.1.2. Language Difficulties. The TEM-4 testing syllabus states that the language difficulty for reading should be at a moderate level and key words should be within the stipulated scope [5]. When it comes to language difficulties, new words and readability are involved. Readability is known as the combination of structural and lexical difficulty of a passage and is mainly affected by sentence length, new words, and other language forms. Since the new word is a factor affecting readability, the focus in this section is to find out the readability of each passage to illustrate the level of difficulty in reading comprehension. In order to make the readability of passage measurable, many a formula has been worked out, such as the SMOG index, the Fry readability estimate, and the Flesch formula. The common purposes of all these methods are to provide an easy and operable way to assess the difficulty level of the input passage. In the present study, the Flesch formula is adopted for the calculation of the readability [8]. The formula is \( RE = 206.835 - 0.846 \times \text{NSYLL} - 1.015 \times \frac{\text{W/S}}{\text{W}} \), where NSYLL is the average number of syllables per 100 words and W/S is the average number of words per sentence. The readability of four passages in TEM (2009) is demonstrated in Table 2.

<table>
<thead>
<tr>
<th>Text A</th>
<th>Text B</th>
<th>Text C</th>
<th>Text D</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>65.06</td>
<td>62.2</td>
<td>74.85</td>
<td>53.75</td>
<td>63.97</td>
</tr>
</tbody>
</table>
Table 3. Reference Table of Readability.

<table>
<thead>
<tr>
<th>Score</th>
<th>Description of style</th>
<th>Typical magazine</th>
<th>Potential audience</th>
<th>School grade</th>
<th>% of US Adults</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-30</td>
<td>Very difficult</td>
<td>Scientific</td>
<td>College</td>
<td>4.5</td>
<td>4.5</td>
</tr>
<tr>
<td>30-50</td>
<td>Difficult</td>
<td>Academic</td>
<td>High school</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>50-60</td>
<td>Fairly difficult</td>
<td>Quality</td>
<td>Some H.S.</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>60-70</td>
<td>Standard</td>
<td>Digest</td>
<td>7th or 8th grade</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td>70-80</td>
<td>Fairly easy</td>
<td>Slick-fiction</td>
<td>6th grade</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>80-90</td>
<td>Easy</td>
<td>Puff-fiction</td>
<td>5th grade</td>
<td>86</td>
<td></td>
</tr>
<tr>
<td>90-100</td>
<td>Very easy</td>
<td>Comics</td>
<td>4th grade</td>
<td>90</td>
<td></td>
</tr>
</tbody>
</table>

By comparing Table 2 with Table 3, it can be found that the readability score for these four passages is from 50 to 80, which indicates that Text A and Text B are standard, Text C is fairly easy, and Text D is fairly difficult. Although the readability formula is designed for English native learners instead of foreigners, it could be regarded as a reference for test designers and provides a basis to build a uniform comparison criterion [7]. It can be drawn from the above analysis that most passages get a good control of difficulty from the perspective of readability, which also meets the requirement of testing syllabus.

2.1.3. Text Types. As required in the TEM-4 testing syllabus, the test should cover a variety of topics and a wide range of subjects from society, science, technology, culture, economy, day-to-day knowledge to biographies [5]. Genre is a means for organizing formal aspects of a text to reflect specific functional intentions of a group, discipline or culture, including five types, description, narration, exposition, argumentation and practical writing. Good tests of reading and good assessment procedures in general will ensure that readers have been assessed for their ability to understand texts in a range of different topics [9]. Ideally we need to select the topic from a suitable genre, at an appropriate level of specificity, and the topic should not be culturally biased or favor one section of the test population [10].

Table 4. Topics and Genres in TEM-4 (2009) Reading Comprehension.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Text A</th>
<th>Text B</th>
<th>Text C</th>
<th>Text D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social life and social science</td>
<td>Science and technology</td>
<td>Day-to-day life (personal experience)</td>
<td>Day-to-day life (personal experience)</td>
<td></td>
</tr>
<tr>
<td>Genre</td>
<td>Exposition</td>
<td>Exposition</td>
<td>Narration</td>
<td>Narration</td>
</tr>
</tbody>
</table>

Table 4 shows that four texts in TEM-4 (2009) are focused on matters of social issues or cultural issues, covering topics from social life and social science, science and technology to day-to-day life. It also shows that exposition and narration are the most commonly employed text types in the TEM-4 (2009) reading comprehension. To some extent, topics and genres meet the requirement of the TEM-4 testing syllabus. However, there is a lack of economic and cultural topics. In addition, passages of argumentation and other practical writings are inadequate. Therefore, more balanced distribution of topics and genres should be taken into consideration in the future tests to promote English teaching and learning.

2.2. Characteristics of the expected response

Expected response is the test of “language use” or “physical response” that attempts to be elicited by the instruction, the task, and the input provided. The characteristics of the expected response are analyzed mainly from response types and reading skills.

2.2.1. Response Types. Response types include “selected response”, “limited production response” and “extended production response”. The multiple-choice questions (MCQ) format, as a selected response type, is widely used in reading testing, and the TEM-4 testing syllabus stipulates the adoption
of this response type [5]. Due to its objectivity of answers, MCQ can be scored by machine, which ensures that marking is rapid, entirely objective and reliable. In addition, a series of statistical tools are available for item analysis, test analysis, score equating, item banking of selected response items, and this type of tests can be pre-tested fairly easily [11]. However, there are demerits of MCQ format measuring reading ability. As Weir argues, “answering multiple-choice items is an unrealistic task, as in the real life one is rarely presented with several alternatives from which to make a choice to signal understanding” [11]. Moreover, MCQ format allows guessing. Statistically a reader can get 25% right ratio by guessing each item. In order to ensure the validity of examination, it is suggested that the tester take a variety of response types when designing test questions in reading comprehension.

2.2.2. Reading Skills. Reading consists of the deployment of a range of separate skills, abilities or strategies [9]. There are six reading skills specified in the TEM-4 testing syllabus as follows [5].

01 Understanding explicitly stated information
02 Making simple judgment and inference
03 Understanding the purpose and gist
04 Understanding the author’s intention and attitude
05 Deducing unfamiliar words, phrases and sentences through context clues
06 Understanding the basic structure of passages

According to the six reading skills listed above, the distribution of the skills in TEM-4 (2009) reading part is presented in the following table.

Table 5. Reading Skills in TEM-4 (2009) Reading Comprehension.

<table>
<thead>
<tr>
<th>Reading Skills</th>
<th>01</th>
<th>02</th>
<th>03</th>
<th>04</th>
<th>05</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>6</td>
<td>9</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

It can be shown in Table 5 that five reading skills are covered in the test, which is consistent with the testing syllabus, but they vary in weight according to different proficiency levels they each stand for. Skills 01 and 02 account for 75%, the biggest proportion of total items, which suggests that understanding of the literal input and inference can exert great impact on the total score. By contrast, skills 03, 04 and 05 hold a relatively small proportion in the test. However, it is an indispensable part for English majors to understand the purpose or gist of the passage and the author’s intention or attitude. Therefore, the tester should take full account of various reading skills in the future test.

3. Conclusion

This paper conducts an analysis of content validity in TEM-4 reading comprehension within the framework of Task Characteristics proposed by Bachman and Palmer (1999) with the help of TEM-4 testing syllabus. Results show that reading comprehension in TEM-4 (2009) has relatively high content validity. The contents are greatly consistent with the requirements of the testing syllabus. However, there are still deficiencies affecting the content validity, and thus some suggestions are proposed for further improvement. Firstly, the passage length should be increased to improve the reading speed. New testing syllabus has set a higher requirement for testees. Meanwhile, the tester should follow the requirements of testing syllabus to design questions to effectively test the testees’ actual reading ability. Secondly, topics and genres should be enlarged. The analysis in Section 2.1.3 shows that text types are limited in TEM-4 reading comprehension, and a wider range of topics and genres should be selected in the future tests, especially those that have not been touched upon over the years. Thirdly, test formats should be enriched. The present test methods are quite limited and only one test method the Multiple-Choice Questions (MCQ) is employed. Although the scoring of MCQ is highly reliable, there are still shortcomings (as stated in 2.2.1) affecting the validity. Therefore, other methods could be adopted, such as short answer questions, information transfer. To sum up, although it is insufficient to depict the full picture of content validity in TEM-4 reading comprehension with a case analysis, the present study could provide useful information for the development of English teaching and learning in the future.
Acknowledgement

This research was financially supported by the 2017 Higher Education Teaching and Reform Project of Guangzhou University of Chinese Medicine and the Municipal Social Science Fund of Guangzhou (Grant NO. 2017GZGJ60).

References