

## On the Content Validity of the Reading Part of 2019 TEM4

Yueyang Qi

Dongguan Campus, Xinhua College, Sun Yat-sen University, No. 7, Yanjiang West First Road, Machong Town, Dongguan City, Guangdong Province, China

qiyueyang220@163.com

**Keywords:** TEM4; Reading Comprehension; Content Validity

**Abstract:** As one of the national standard examinations for College English majors, Test for English Majors (TEM4) is well recognized in China. It is not only the main instrument to measure the language proficiency of English majors, but also a mirror to reflect the situation of English teaching in China. Whether or not the TEM4 is valid and efficient is important to both English learners and teachers, and forth is worthy thinking of. Since reading comprehension, which tests the reading ability of English learners, weighs much in the TEM4, this paper tries to examine the content validity of the reading part of 2019 TEM4, in accordance to the TEM4 test syllabus and test specifications.

### 1. Introduction

Since 1990, TEM4 (Test for English Majors-Band 4) has been implemented nationwide for almost twenty years. As the main proficiency test for college English majors, it has received increasing attention by English learners, teachers and researchers in China in recent years. Since TEM4 is well recognized in China and over the world, it means much to the test-takers. For all the English Majors, TEM4 is a bend mark for their efforts and achievements in English learning. And for English teachers, TEM4 is a common way to measure their teaching result. Virtually, TEM4 is expected to reflect the real situation of English Learning and teaching in China. How and to what extent does the TEM4 measure the real language proficiency of English learners is not simply a question, but has developed into a complex science, which deserves careful research. Meanwhile, as is known, reading is considered as an essential skill in language learning and reading ability is of great importance to English learners, especially to English majors. Moreover, reading competence is a major goal of English teaching according to college English teaching syllabus. Therefore, reading comprehension part always accounts much in any type of English tests, including such national standard examination<sup>0</sup> as CET4, CET6, TEM4 and TEM8, etc. Because of the significant status of TEM4 and the crucial role of reading comprehension, it is necessary to develop an efficient and reasonable reading comprehension test to guarantee the validity and reliability of TEM4. Due to limited time and space, this paper tries to conduct a non-empirical investigation on the content validity of the reading comprehension part of 2019 TEM4.

### 2. Theoretical Background for the Current Evaluation of Language Testing

#### 2.1 Rationality for Validity.

“There could be no science as we know it without measurement” (Henning, 2001:1). As one form of measurement, language testing plays an important role in determining to what degree language learners has mastered the target language. However, there would be potential for error in language testing just as we measure any material object. To measure the real proficiency of English learners, the test makers must reduce errors to the greatest extent and provide a relatively reliable and valid standard for measuring. Before the test initially designed is put into service, it is necessary to make a comprehensive evaluation of the test so as to ensure its efficiency. To do so, the test makers must take many factors into consideration, including the purpose of the test, the

characteristics of the examinees, the accuracy of measurement, the suitability of format and features of the test, the developmental sample, the availability of the equivalent or equated forms, the nature of the scoring and reporting of scores, the cost, the procurement, and the political acceptability of the test (Henning, 2001:9). Among all these factors, the first and foremost consideration in selecting and developing a test is to meet the purpose of the test. An efficient test must measure what is supposed to measure. That is to say, it should be ascertained whether or not the test is valid for its intended use.

## **2.2 Definition of Validity.**

As a significant indicator in evaluating a test, validity in general refers to the appropriateness of a given test or any of its component parts as a measure of what it is purported to measure (Henning, 2001: 89). According to Henning, a test is said to be valid to the extent that it measures what it is supposed to measure. It follows that the term valid, when used to describe a test, should usually be accompanied by the preposition *for*. Here, Henning emphasizes the matching of the initial purpose of a particular test. Messick (1989) describes validity as “an integrated evaluative judgment of the degree adequacy and appropriateness of inferences and actions based on test scores”. Similarly, according to the standards of educational and psychological testing, validity always refers to the degree to which that evidence supports the inferences that are made from the scores (Bachman, 1990:236). Different as the various definitions are, the essence of validity is the quality to make a test efficient enough to reflect the authentic information of the test-takers, with minimum error.

Talking of validity, people may find it rather abstract and difficult to grasp. Detailed classification will make it easier to understand. Alderson & Clapham (2000) classified validity into internal and external validity. The former relates to the studies of perceived content of the test and its perceived effect; while the latter relates to studies comparing students’ test scores with measures of their ability gleaned from outside of the test (Alderson & Clapham, 2000:171). Usually, the external validity includes face validity, content validity and response validity and the internal validity consists of concurrent validity, predictive validity and construct validity.

## **2.3 Face Validity & Content Validity.**

Many testing specialists believe that there are no distinction between face validity and content validity. So face validity is usually mentioned together with content validity. However, some researchers do differentiate face validity and content validity by noting that face validity, unlike content validity, is often determined impressionistically (Henning, 2001:94). Face validity refers to the appearance of validity or the extent to which the test appeals or is acceptable to the test-takers and test users (Bachman, 1996:42). According to the definition, face validity seems to be a kind of intuitive inspection on a test and lacks scientific evidence for the evaluation. Despite its non-scientific nature, face validity is important to a test. Imagine that if a test even does not appear to be valid, how would it serve seriously and reliably? Face validity is the prerequisite condition for the presentation of other kinds of validity. Content validity, as the name implies, 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. (Henning, 2001:94). Content validity is also deficient in empirical data and is dependent on testing experts’ judgment. However, it is logic and reasonable because content validity of a test is often inspected according to certain standardized norms.

## **2.4 Response Validity.**

Gathering information on how individual test takers respond to test items has become increasingly common aspect of test validation (Alderson & Clapham, 2000:176), Henning (2001) points out that the response validity is intended to describe the extent to which examinees responded in the manner expected by the test developers. It concerns the methods which the test-takers exploit to fulfill a testing task. Theoretically, the test-takers are expected to complete a test in such appropriate manner that they can perform to their full capacities. For example, the choice questions are expected to be answered according to the test-takers’ potential knowledge and

skills rather than guessing. Unclear instructions and unfamiliar test format will also affect the test-takers' response to the test. So it requires much to guarantee a good test with response validity.

## **2.5 Concurrent Validity.**

Generally speaking, concurrent validity is criterion-related. According to Hughes (2000), criterion-related validity is to see how far results of the test agreed by those provided by some independent and highly dependable assessment of candidates' ability. Here, this dependable assessment is the 0criterion measure against which the test is validated. Concurrent validity is established when the test and the criterion administrated at the same time" (Hughes, 2000:23). This criterion could be an existing test which is authoritative and administrated at the same time; it could also be the teachers' ratings or other independent assessment of the students' performance in a similar task at the same0 time. To evaluate the concurrent validity, data must be collected and formulas be applied to produce the validity coefficient.

## **2.6 Predictive Validity.**

Predictive validity of a test concerns "the degree to which the test can predict candidates' future performance" (Hughes, 2000:25). It is also a kind of criterion-related validity. Predictive validity is usually reported in the form of a correlation coefficient with some measure of success in the field or subject of interest (Henning, 2001, 97). For example, in evaluating the predictive validity of a test, one can refer to a later test that measures the ability the previous test was intended to predict.

## **2.7 Construct Validity.**

The notion of construct validity is rather abstract. The term construct refers to a psychological construct, a theoretical conceptualization0 about an aspect of human behavior that cannot be measured or observed directly. Examples of constructs are intelligence, achievement motivation, anxiety, achievement, attitude, dominance and reading comprehension (cited from Alderson & Clapham, 2000:183). According to Bachman (1990), construct validity concerns the0 extent to which the performance on tests is consistent with prediction that we make on the basis of theory of ability, or construct (Bachman, 1990:225). That is to say, if the test-takers' performance in a test share alike with their true ability, the test is said to have construct validity.

## **2.8 Judgment for Content Validity.**

Content consists of the major part of a test. The consideration of the test content is thus an important part of both test development and test use (Bachman, 1990:244). And correspondingly; the content validity contributes much to the integrated quality of a test. Content validity should therefore be taken into particular consideration to guarantee a qualified test, especially those prestigious ones such as TEM4. However, it is not an easy job to develop the content of a test. As we all know, language learners' knowledge of the target language is infinite, and hence can't be tested for every aspect. As Henning (2001) explained, usually a test cannot be exhaustive but must be selective in content. That is to say, the test developers must choose a representative sample from the hypothesized knowledge database of language learners. Hughes (2000) pointed out that a test is said to have content validity if its content constitutes a representative sample of the language skills, structures, etc. To judge whether the content of a test is representative, we need to compare the test content with the test specifications. Bachman (1990) mentions two aspects to the part of validation of test content: content relevance and content coverage. The former concerns the0 extent to which the content is related to the behavior domain in question; while the later is to examine the extent to which the tasks required in the test adequately represent the behavior domain in question (Bachman, 1990: 244). Actually, we should check these two aspects of the test content in light of the test purpose, usually by comparing the items of a test with the macro teaching syllabus, test syllabus and test specifications.

### **3. Analysis of the Content Validity of the Reading Part of 2019 TEM4**

2019 TEM4 has been successfully implemented. But still, the quality of this test, both its advantages and disadvantages, is worthy of rechecking to facilitate later test development. Altogether, the 2019 TEM4, like others, consists of six parts: dictation, listening comprehension, cloze, language use, reading comprehension and writing. Since the reading comprehension part takes up a big share of the whole test, the validity, especially content validity, of reading comprehension part directly exerts great influence on the quality of the test. This paper tries to make a reasonable evaluation of 2019 TEM4 though probing into the content validity of the reading comprehension part.

As mentioned previously, to measure the content validity of a test, a careful comparison between test content and the teaching syllabus, test syllabus and test specifications must be made. So it's necessary to make a brief review on the syllabus and specifications before a detailed analysis. According to the teaching syllabus for English major of Higher Education, at the foundation stage, the teaching task is to impart the basic knowledge and cultivate students' basic language ability, that is, to enhance students' ability to use the target language in practice, to enrich students' social cultural knowledge and increase students' intercultural sensitivity so as to prepare the students for further study in advanced stage. And the test syllabus for TEM4 states that the aim of the test is to check whether and to what degree have the students, who have finished all the courses required by English major level four, mastered the required language ability including grammar, vocabulary and pragmatic ability, etc. Moreover, the test aim for the reading comprehension part is to measure the students' ability to retrieve and analyze information through reading with accuracy and a certain speed.

Combining the above principle of teaching and testing with other factors, this paper tries to discuss the content validity of the reading comprehension part of 2019 TEM4 in terms of the following aspects.

#### **3.1 Instructions.**

According to Weir, the test instruction should be candidate-friendly, intelligible, comprehensive, brief and simple, and more importantly as explicit and clear as possible (Weir, 2005:57). For the reading comprehension part of 2019 TEM4, it states "In this section there are three passages followed by ten multiple choice questions. For each multiple choice question, there are four suggested answers marked A, B, C and D. Choose the one that you think is the best answer and mark your answers on ANSWER SHEET TWO." for section A and "In this section there are five short answer questions based on the passages in Section A. Answer the questions with NO MORE THAN TEN WORDS in the space provided on ANSWER SHEET TWO." for section B. In this short paragraph with common words and easy grammars, information about the reading comprehension test, including the structure of the test ("three passages followed by ten multiple choice questions" as well as "five short answer questions based on the passages in Section A"), ways of answering ("Choose the one that you think is the best answer and mark your answers on ANSWER SHEET TWO" as well as "Answer the questions with NO MORE THAN TEN WORDS in the space provided on ANSWER SHEET TWO"), etc. is clearly presented. Furthermore, this kind of instruction is very familiar to the students because it is organized in the same way of other tests. So it can be ascertained that the students can get a clear understanding of the requirements in doing the reading comprehension.

#### **3.2 Text Length.**

There are altogether three texts selected as the reading materials for the reading comprehension in 2019 TEM4. Practically, the test should not be too long or too short. Due to limited test time, a test that is too long will reduce the test-takers' interest and confidence, thus influencing their performance. If so, the test cannot reflect the true ability of the test-takers. While if the text is too short, there will not be enough information to measure the ability of the test-takers. According to the test syllabus, the reasonable length for the reading materials should be kept at about 2000 words in

total. Statistical calculation demonstrates that any one of the three texts in 2019 TEM4 remains a proper length between 500 to 1000 words, which well fulfills the expected requirements.

### 3.3 Text Content.

Content about the reading material is crucial to the validity of the reading comprehension test. It must be representative considering the teaching syllabus, test purpose, students' practical situation and so on. Here, we must take into consideration the following aspect.

### 3.4 Topic.

The topic of a reading passage concerns what it is written about. The topics of the reading part of 2019 TEM4 can be presented in the following table:

Table 1 Passage topics for the reading comprehension of 2019 TEM4

Passage One	Passage Two	Passage Three
Social Science (Effect of immigrants)	Humanities + personal experience (cherish moments with family)	social science (literature improves skills in life)

From the above table, we can see that the three topics involve not only common phenomenon in daily life but also social science, which can comprehensively examine students' preliminary understanding ability and intercultural knowledge. Also, the background knowledge of these topics is easy for the students to grasp. This is desirable in the test requirement. However, there is a drawback on the selection of these topics: two of the three topics (Passage one and three) are all about social science. But the test syllabus specifies the principles of selecting reading materials as follows: the topics should cover a wide range including social, culture, common sense, general science, human biography, etc. Although they start from different aspects of life, the topics of the reading comprehension materials for 2019 TEM4 could be further verified.

### 3.5 Genre.

Genre refers to the ways to organize the formal aspects of a text to reflect specific functional intentions of a group, discipline or culture. Generally speaking, there are five kinds of genre: description, exposition, narration, argumentation and practical writing. The situation of genre types of the reading texts for 2019 TEM4 is presented in the following table:

Table 2 Genre types of the passages in the reading part of 2019 TEM4

Passage One	Passage Two	Passage Three
Exposition	Narration	Exposition

Among the three passages, two are expositions and one narration. Exposition and narration are the commonest genres used in reading comprehension tests. Articles in these two kinds of genres are relatively easy to process and retrieve information and therefore are suitable for English majors at the foundation stage. However, there are no argumentation, description and practical writing. According to the test specification various genre types should be presented in a test, because it is important for the students to foster the skills to comprehend or write articles with different kinds of genre.

### 3.6 Language.

The language of the four texts in reading comprehension test of TEM4 is similar to that of the articles the students are usually taught with. So the difficult index of the language is just within the ability of the test-takers. Most of the vocabulary and grammatical structures in the selected texts are under the scope of what the teaching syllabus requires. So performance on the three passages can reflect students' real situation of language mastery.

Generally speaking, the design of the reading comprehension part of 2019 TEM4 is consistent with the requirements of the teaching syllabus, test syllabus and test specifications in terms of its instructions, text length and text content. And it can be claimed that the reading comprehension part

of 2019 TEM4 has a relatively high content validity.

#### **4. Conclusion**

High content validity of the reading comprehension part is crucial to the quality of the whole 2019 test for English majors. However, high validity does not equal to perfect validity. Shortcomings in the design of the reading test mentioned in this paper must be improved. And there are still other factors that should be taken into consideration when evaluating the content validity of a test.

#### **Acknowledgments**

This paper is supported by Xinhua College (SYSU) Education Reform Project “EAP Curriculum Development for Yat-sen Class Based on Need Analysis”.

#### **References**

- [1] J.C. Alderson, C. Clapham, Language Test Construction and Evaluation, Foreign Language Teaching and Research Press, Beijing, 2000.
- [2] J.F. Bachman, Fundamental Consideration in Language Testing, Oxford University Press, Oxford, 1990.
- [3] J.F. Bachman, A. S. Palmer, Language Testing in Practice, Oxford University Press, Oxford, 1996.
- [4] G. Henning, A Guide to Language Testing: Development, Evaluation and Research, Foreign Language Teaching and Research Press, Beijing, 2001.
- [5] A. Hughes, Testing for Language Teachers, Foreign Language Teaching and Research Press, Beijing, 2000.
- [6] C.J. Weir, Language Testing and Validation—an Evidence Based Approach, Palgrave Macmillan, New York, 2005.