ISSN 1798-4769
Journal of Language Teaching and Research, Vol. 8, No. 6, pp. 1229-1233, November 2017
DOI: http://dx.doi.org/10.17507/jltr.0806.27

Test for English Majors-band 8 (TEM8) in China

Yang Yang
School of Foreign Languages, Chang'an University, China

***Abstract*—As a prominent test in China, TEM8 has already been paid much attention to. There are many researches about the qualities of TEM8 at home and abroad. However, few of them have a comprehensive evaluation of TEM8. Therefore, this article attempts to analyze TEM8 in a more comprehensive perspective and provide more information about the qualities of TEM8. Based on Bachman & Palmer's mode of test usefulness, this article reviews TEM8 in terms of its reliability, construct validity, authenticity, interactiveness, impact and practicality, giving implication to test developers.**

***Index Terms*—TEM8, test quality, test usefulness, review**

I. Introduction

Test purpose: TEM8 is designed as an achievement to measure the overall English proficiency of senior undergraduates majoring in English Language and Literature in China and to decide whether these students meet the requirements of English language abilities and professional knowledge of English as specified in the National College English Teaching Syllabus for English Majors (NACFLT, 2004).

Administration: TEM8 is administered by the National Advisory Committee for Foreign Language Teaching (NACFLT) on behalf of the Higher Education Department, Ministry of Education, People's Republic of China. TEM8 is administered once a year in March and the total test time is 195 minutes before 2016 (there are exceptions: the actual testing time in 2005 is 190minutes, in 2006 and 2007 195 minutes).

Scoring: TEM8 is a criterion-referenced test, so it gives test takers feedback in the form of grades (Brown & Abeywickrama, 2010). Its scores are reported to the Academic Affairs Office of the participating universities. If the test taker' s score is or above 60, he or she will receive a certificate from the NACFLT. The proficiency level of test takers which is reported on certificate includes three ranks, that is, "excellent" (score 80 or above), "good" (score between 70 and 79) and "pass" (score between 60 and 69). Since it is each participating university that is responsible for reporting scores to test takers and its scores are just used for improving the English teaching and learning in its own university, TEM8 has the diagnostic feature.

Development: TEM8 was officially launched in 1991 after the publication of the first national teaching syllabus for English majors. Due to the rapid economic development, English language had been paid much attention to for quite a long time. Therefore, as a national test for English majors, TEM8 was widely accepted among universities in mainland China. After several years of improvement and development, nowadays, TEM8 has become a very popular English test which plays important role in English learning and teaching in Chinese universities.

For TEM8, the test items have changed a lot within 25 years since it was launched. When it was first launched, the test methods included listening, reading, proofreading, translating and writing. In 2005, a new test method was added, that is, general knowledge. Actually, the changes of the test methods and content are in accordance with the development of TEM8 syllabus. There are three main changes of TEM8 syllabus. The first one was published in 1994. After its publication, the related departments made some studies and then, in 1997 after some careful revision according to the study results, the second one were designed and published. In 2000, a new edition of Teaching Syllabus for English Majors was adopted, which had great effect on the English teaching and thus for TEM8. Therefore, in order to comply with the teaching requirements set by the 2000 edition of Teaching Syllabus, in the 2004, a new edition of Syllabus for TEM8 was designed and published. This version is more comprehensive and detailed than the first two versions. Therefore, the following tests are designed according to this syllabus for many years. However, in 2016, the test methods are changed again. The general knowledge which was added in 2005 has been canceled. There are also little or big changes for the other five test tasks.

II. Test Tasks and Methods

The following evaluation will mainly focus on TEM8 in 2016, so the test tasks and methods of TEM8 of 2016 are presented in Table 1.

The total time for TEM8 is 195 minutes for former versions. In 2016, the test time is adjusted to 150 minutes and the test content and format are also changed a lot.

Listening: Before 2016, *Listening* includes three different task types: mini-lecture, interview and news broadcast. The first task requires test takers to listen to a lecture taking notes and then fill the ten blanks according to the notes. The second has one conversation and asks test takers to answer five multiple choice questions. The third includes several news broadcast and five MCQs. However, in 2016, only the first two task types are remained. For completing the first task, test takers are required to listen to the lecture and fill the fifteen blanks at the same time. The second asks test takers to listen to one or two conversation and finish ten MCQs. Though the test task type is changed the time allocation and the talking speed of the recording is remained as before.

Reading: Before 2016, the *Reading* part only has four passages for multiple choice questions, while in 2016 this part is changed a lot and it has three passages for two different reading tasks. That is to say, reading comprehension comprises two task types: multiple choice questions and short answer questions. Test takers are required to read three passages and then answer fourteen MCQs and 8 short answer questions. The time allocation is the same as before.

General knowledge: It was added in 2005 but cancelled in 2016. It includes ten MCQs about the culture and literature of English speaking countries and linguistics. The cancellation of this part may lead to the changes of the whole test quality.

TABLE 1
TEST TASKS OF TEM8

Task	Input	Format	%	time
Listening	1 mini-lecture, listen once, 150wpm	Filling in the blanks	15	25m
	1 interview, listen once, 150wpm	MCQ	10	
Reading	three passages totaling c. 3000 words	MCQ Short answer questions	30	45m
Language Usage	A passage of 250 words	Error identification and correction	10	15m
Translation	A text of 150 characters	Translation C-E	15	20m
Writing	Materials	An article of 300 words	20	45m

Note. wpm = words per minute; MCQ = Multiple Choice Questions; C = Chinese, E = English.

Language usage: It was used to be called “proofreading” before 2016. It requires test takers to add, delete or change one word of each sentence to correct grammatical errors. There are ten sentences that they need to correct.

Translation: Before 2016, *Translation* part has two translating tasks, that is, translating Chinese into English and translating English into Chinese, and the total time for this part is 60 minutes. In 2016, the revised version only has one task-translating Chinese into English and the time is adjusted to 20 minutes.

Writing: Before 2016, the *Writing* part gives test takers a topic and requires them to write an article about 400 words, while the revised version provides materials and enough context and asks test takers to write an article about 300 words based on the given materials.

The change of the test task and method would lead to the change of the test qualities of TEM8, therefore, it is necessary to make a study on it and provide some information to the related departments and persons in the field of English teaching and learning. For the test qualities, there are some different modes to measure it, among which the test usefulness and test fairness are very widely used. The former was proposed by Bachman & Palmer (1996) and the latter was proposed by Kunnan, A.J. (2004). The following evaluation and analysis focus on the former, that is to say, the review in this article is about the test usefulness of TEM8.

III. TEST QUALITIES

The usefulness is the most important quality of a test (Bachman & Palmer, 1996). Based on the test qualities of TEM8, many studies have been carried out in China (Xu, 2012; Zou, Peng & Kong, 2009; Zou, 2003). However, few of them have a comprehensive evaluation of the quality of TEM8. Among these studies, the content validity, construct validity and the washback are emphasized while the other qualities are always neglected. Therefore, this article attempts to analyze TEM8 in a more comprehensive perspective and provide more information about the qualities of TEM8.

Bachman & Palmer (1996) proposed a mode of test usefulness in terms of six main qualities: reliability, construct validity, authenticity, interactiveness, impact, practicality. This model will be used in this article to review TEM8. The reason why this mode is used is that this mode is very comprehensive for assessing a test and the six qualities in this mode are interrelated. For TEM8, this mode is very practical and authoritative. In other words, TEM8 is needed to be analyzed by this mode so as to provide some suggestion for further improvement.

1. Reliability

Reliability is often defined as consistency of measurement (Bachman & Palmer, 1996). A reliable test should (1) be consistent in its conditions across two or more administrations; (2) give clear directions for scoring; (3) have uniform rubrics for scoring; (4) contain items that are unambiguous to the test taker (Brown & Abeywickrama, 2010).

According to the teacher participated in the interview, TEM8 has clear holistic scoring for subjective test items which amount to 35% in 2016. Compared with 40% in the former tests, the decrease of the subjective items would help to improve marker reliability since it makes the scoring more objective. What’s more, from 2009 TEM8 has adopted computer-assisted online scoring, which makes the scoring more efficient and reliable. The scorers of TEM8 receive professional and specific training before scoring. For example, during the training, they all score one test paper at the same time, if someone gives very high or low score, he or she will be reminded and provided another guidance until almost all scorers have similar scores for one test paper. Besides, the scorer are required to conform to the criteria of scoring specified already to make sure the fairness of the scoring. Therefore, the inter-rater reliability and intra-rater reliability could be enhanced greatly compared with the traditional scoring. Especially for inter-rater reliability, the new machine scoring makes it possible that at least two people score the same paper. And the scorers are monitored by computer, if one's score is much higher or lower than the whole group, he or she will be reminded or stopped. All of these strengthen the reliability of scoring. For the test items, according to the teacher, they are clear enough for test takers to avoid the misunderstanding. After the reformation of TEM8 in 2016, a document about the change of test format and techniques has already been notified, so that the test-takers of 2016 are also familiar with them.

Reliability can be quantified in the form of a reliability coefficient (Huges, 2003). For this, statistics from the test center show that the averaged internal consistency coefficients from 2008 to 2010 are 0.815 for TEM8 which is reasonably high for most test uses (Jin & Fan, 2011). Though the inconsistencies cannot be eliminated entirely, the test developers could control the potential sources of inconsistency, especially the characteristics of the test tasks. As a very prominent test in China, certainly, with so many years of development, it can be expected that consistency coefficients of TEM8 would be much higher.

It can be found that as a popular and important national language test, the developers of TEM8 has been tried their best to ensure the high reliability of it.

2. Construct validity

A construct refers to any theory, hypothesis, or model that attempts to explain observed phenomena of language ability (Brown & Abeywickrama, 2010). Construct validity is used to refer to the extent to which one can interpret a given test score as an indicator of the ability or construct, one wants to measure (Bachman & Palmer, 1996).

It's not easy to measure the construct validity of a test. In China, some scholars have used the quantitative approach combined with the qualitative approach to measure the construct validity of TEM8. Zou, Peng and Kong (2009) analyzed the correlation and representativeness of *General Knowledge* and based on EQS and BILOG, and in this study, they made Exploratory Factor Analysis and Confirmatory Factor Analysis. According to their research results, the test items of *General Knowledge* have relatively high correlation and representativeness. However, some items are much easier or harder, which threatens the whole construct validity. In 2016, *General Knowledge* is cancelled. It is likely that the whole construct validity of TEM8 will be changed, which needs more studies to measure it.

Han (2014) measured the construct validity of *Reading* of TEM8 by analyzing the number of passages, total words and the type of tested items. From the analysis result, it can be found that *Reading*'s construct validity is high. However, Han thought that it could be improved by adding other test items like short answer questions. In 2016, *Reading* has been changed as Han expected. Maybe the construct validity of 2016 is higher than former versions.

As an achievement test, actually, there is something else that is also very important for TEM8, that is content validity. Content validity and construct validity are closely related. Investigation of a test's content validity provide evidence for construct validity (Huges, 2003). Therefore, it is necessary to have a look at content validity. 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 (Huges, 2003). Compared TEM8 content with its syllabus, it can be found that the former almost completely meets the requirements specified in the latter. Therefore, to some extent, it can say that TEM8 has relatively high content validity, thus the construct validity. Besides, construct validity is also related to interactiveness. According to Bachman & Palmer (1996), interactiveness provides the vital link with construct validity. It not suitable to measure the construct validity of a test from the single perspective, which makes it is difficult to measure the construct validity of a language test. Therefore, more empirical studies should be made to provide useful information to the test developers of TEM8.

3. Authenticity

Bachman and Palmer (1996) defined authenticity as "the degree of correspondence of the characteristics of a given language test task to the features of a target language task". An authentic test should (1) contain language that is as natural as possible; (2) have items that are contextualized rather than isolated; (3) include meaning, relevant, interesting topics; (4) provide some thematic organization to items; (5) offer tasks that replicate real-world tasks (Brown & Abeywickrama, 2010).

The language of TEM8 is natural basically. In detail, the language for *Listening* is natural and close to the real life. Besides, the situation and context of the lecture or the conversation are also similar to the real world, which meet the requirements of Syllabus for TEM8. However, in order to evaluate the kinds of translation skills and ensure the validity of test items, the language of passage in *Translation* is unnatural to some extent; most often, the passages in this part are taken from the literatures. The writing topics of TEM8 are interesting basically. And the writing task of 2016 is closer to the real world than that in the former tests. In fact, each test language and topic couldn't be as nature as the language of real life. It is impossible to do that because each test has to test some skills and abilities specified in test specifications or syllabus already. However, with the development of TEM8, the test designers have tried their best to make the language and topic as natural as possible. They tend to choose the materials whose topics are very interesting and natural enough to make the test takers be involved in. Take the writing task of 2016 as an example, test takers are asked to write an article about *Ice Bucket Challenge*, an activity initiated to raise money and awareness for the disease ALS. This activity has strong repercussions and most test takers may already hear about it. What they need to do is that expressing their opinion towards the activity, especially whether the problem found with this kind of activity will finally undermine its original purpose. Compared with the former task type, this task is more natural and free for test takers to express their feeling and thought. Therefore, from this perspective, the authenticity of TEM8 is getting higher.

4. Interactiveness

Bachman and Palmer (1996) defined interactiveness as "the extent and type of involvement of the test taker's individual characteristics in accomplishing a test task". They thought that the interactiveness of a given language test task can be characterized in terms of the ways in which the test taker's areas of language knowledge, metacognitive strategies, topical knowledge, and affective schemata are engaged by the test task.

For *Listening* of TEM8, it has relatively high interactiveness, according to the interview results. When test takers participate in TEM8, before listening materials, they are encouraged to be relaxed. There are recordings before the test which will help them to be familiar the atmosphere of the test and make sure that their equipments are on work. Therefore, social affective strategies have an influence on their results. For *Reading*, metacognitive strategies are involved in. Test takers often adjust their reading strategies and reading speed or control the time when they finish different reading tasks during the test. The time of TEM8 is very limited so the use of metacognotive strategies is very necessary for them to finish all test items within the limited time. Especially in 2016, the first reading passage is about Gatsby which is a character in literature learnt by most students majoring in English Language and Literature, their literature knowledge could help them to finish tasks. The changed version of writing in 2016 is also more interactive than the form ones, since test takers could express their feeling more freely. However, according to the interview, TEM8 does not have very high interactiveness since test takers always don't know how to use the metacognitive strategies or affective schemata. In other words, students don't recognize them when participating in the test.

5. Impact

The impact of test use operates at two levels: a micro level, in terms of the individuals who are affected by the particular test use, and a macro level, in terms of the educational system or society (Bachman & Palmer, 1996). An aspect of impact that has been of particular interest to both language testing researchers and practitioners is washback. Huges (2003) defined washback as "the effect that tests have on learning and teaching".

As a curriculum-based achievement, it is no doubt that TEM8 has some positive waskback. Universities would adjust the course plan and teaching content based on the Syllabus for TEM8 and the testing results. Specifically, the colleges pay much attention to improving student's reading speed and accuracy. Listening comprehension also is attached importance to since for this part student's scores are relatively low for several years according to the statistics from the Test Center. However, TEM 8 also has negative impact on English teaching in university yet it is much less than the positive one. For example, in order to improve the score of writing, teachers pay much attention to the writing skills, which constraints students a lot in terms of the form and thought. In China, English teaching has being always paid much attention to. Therefore, some teachers and scholars have separately studied the waskback of TEM8. For example, Xu (2012) investigated the washback of TEM8 through questionnaire surveys among 5 foreign language experts about 700 English discipline leaders. This study also suggests that in general there is a positive attitude towards the washback of TEM8 but the unfamiliarity with the marking criteria and the limited information provided in the test reports may lead to some negative washback.

In a word, as a prominent English Language test in China, TEM8 really has positive impact on English teaching and learning.

6. Practicality

Practicality refers to the logistical, down-to-earth, administrative issues involved in making, giving, and scoring an assessment instrument. These include costs, the amount of time it takes to construct and to administer, ease of scoring, and ease of interpreting/reporting the results (Brown & Abeywickrama, 2010).

Since TEM-8 was launched in mainland China, students from different universities have taken part in it. In order to carry out it easily and efficiently, National Advisory Committee for Foreign Language Teaching adopted the way that the Academic Affairs Offices of the participating universities are responsible for registration and test administration and are answerable directly to the Test Center (Jin & Fan, 2011). Similarly, TEM8 test scores are reported to the Academic Affairs Office of the participating university and through the English department of their own colleges, test-takers could get to know their scores and ranking. Having been administered by this way for many years, TEM8 is a widely and orderly administered test in mainland China.

As have mentioned above, the carrying out of the computer-assisted scoring helps to improve the efficiency of it. The scoring of TEM8 has become more quickly and fairly. Therefore, it can be concluded that TEM 8 has quite high practicality.

IV. Conclusion

Based on Bachman & Palmer's mode of test usefulness, this article attempts to review TEM8 in terms of its reliability, construct validity, authenticity, interactiveness, impact and practicality. From the above analysis, it can be found that from the perspective of usefulness, TEM8 has attained a relatively high standard in certain aspects of test qualities. However, there still exist some weaknesses needed to be improved. For example, the test reliability and validity could be balanced and improved.

In conclusion, the TEM8 developers have strived to make it much better and enhance the usefulness of it. Being a criterion-referenced test, the designing of TEM8 strictly conforms to the teaching requirements specified in the Syllabus (NACFLT, 2000). Besides, more attention should be paid on the effective and comprehensive evaluation of TEM8, giving implication to test developers.

REFERENCES

[1] Bachman, L. Y. & Palmer, A. S. (1996). Language testing in practice: Designing and developing useful language test. Oxford: Oxford University Press.

[2] Brown, H.D. & Abeywickrama, P. (2010). Language assessment: Principles and classroom practices (2nd ed.).New York: Pearson.

[3] Han, X. (2014). A research on the validity of reading comprehension of TEM8 in 2011. *Journal of Hubei University of Science and Technology, 2,* 124-125.

[4] Huges, A. (2003). Testing for language teachers (2nd ed.). Cambridge: Cambridge University Press.

[5] Jin, Y. & Fan, J. S. (2011). Test for English majors (TEM) in China. *Language Testing, 28 (4):*589-596.

[6] Kunnan, A.J. (2004). Test fairness. In M. Milanavic & C. Weir (Eds.). *Europe language testing in a global context: Selected papers from the ACTE conference in Barcelona* (pp.27-48). Cambridge: Cambridge University Press.

[7] NACFLT. (2000). Syllabus for University English Language Teaching. Beijing: Foreign Language Teaching and Research Press.

[8] NACFLT. (2004). Syllabus for TEM-8. Shanghai: Shanghai Foreign Language.

[9] Xu, Q. (2012). The washback of Test for English Majors-Band 8. *Foreign Language World, 3,* 21-31.

[10] Zou, S. (2003). The connection between language teaching syllabus and language testing-The designing and carrying out of TEM8. *Foreign Language World, 6,* 71-78.

[11] Zou, S., Peng, K.Z. & Kong, W. (2009). Exploring the construct validity of the general knowledge section in TEM8. *Foreign Language in China, 6(1):* 45-52.

Yang Yang was born in Gansu, China in 1992. She received her degree of Bachelor of Arts in Foreign Language and Literature from Northwest Normal University, China in June, 2015.

She is currently a postgraduate student in Chang’an University, Xi’an, China.