Validity Considerations in the Construction 
and Use of Test for English Majors*

XU Qian
Shanghai International Studies University, Shanghai, China

Test for English Majors (TEM) is a national test for college undergraduates majoring in English in China. As a nationwide and high-stakes test, its validity has long been an issue under discussion. Therefore, this paper explores the validity of TEM, employing Bachman and Palmer’s Assessment Use Argument (AUA) model, focusing on the aspects of consequences, decisions, interpretation, and assessment records of TEM. Based on these validity considerations, some misinterpretations and misuses have been pointed out and relevant suggestions have been made on the reform of TEM.

Keywords: TEM, validity, Assessment Use Argument

Introduction

Test for English Majors (TEM) is a national test for college undergraduates majoring in English, with an annual test taking population of more than 400,000, counting both TEM-4 and TEM-8 candidates. As one of the predominant English tests in China, TEM has won extensive recognition from the test takers, relevant institutions, and society at large, playing an increasingly important role in English language teaching and learning at the tertiary level (Jin & Fan, 2011).

Therefore, in this paper, I will look into the validity of TEM, trying to justify some of its interpretations and uses, as well as pointing out some possible misinterpretations and misuses. Bachman and Palmer’s Assessment Use Argument (AUA) model will be employed in the analysis because one way to provide a consistent framework for validation efforts is to structure them in terms of arguments and the validity argument can provide an overall evaluation of the intended interpretation and uses of test scores (Cronbach, 1988).

Assessment Use Argument

Assessment Use Argument (AUA) is a conceptual framework to link inferences from assessment performance to interpretation and use, which explicitly states the interpretations and decisions that are to be based on assessment performance, as well as the consequences of using an assessment and of the decisions that are made. It can provide an overarching inferential framework to guide not only the design and development of language assessments, but also the interpretation and use of language assessment results (Bachman, 2007).

* Acknowledgements: This paper was funded by Scientific Research Project sponsored by Shanghai International Studies University (Research Grant No. Kx181123).

XU Qian, lecturer, doctor, Law School, Shanghai International Studies University.
The AUA model can be illustrated in the following Figure 1:

![Figure 1. The Assessment Use Argument model (adapted from Bachman, 2007).](image)

In the rectangles and their respective circles are questions, or claims, which are to be considered in assessment development and the use and interpretation of assessment results. Therefore, in the following analysis, I am going to answer these questions one by one, attempting to justify or disprove the use and interpretation of TEM test results.

### Validity Considerations

In using an AUA to justify the use and interpretation of TEM results, we shall first consider its consequences—what are the consequences and are they beneficial to the stakeholders? Then, we will look into the decisions to be made—are they value-sensitive, that is, do they comply with relevant educational and societal values, as well as laws, rules, and regulations? Moreover, is the decision fair to all test takers? Next, when interpreting the ability to be assessed, we have to ensure that the interpretations are meaningful with respect to the learning syllabus or some language ability theories, impartial to all test takers, generalizable to target language use domains, relevant to the decisions to be made, and sufficient to make certain decisions. Lastly, we have to consider the consistency of assessment records, such as scores and descriptions of the test takers’ performance.

### Consequences

The purpose of TEM is to assess the language proficiency of undergraduate students majoring in English, to examine whether they have met the required levels specified in the National College English Teaching Syllabus.
Meanwhile, TEM may facilitate the evaluation of English teaching, to measure how the teaching syllabus is implemented in different universities, and to promote the reform on English teaching methodology and curriculum design.

In this sense, the consequences of TEM are beneficial to the students (test takers), teachers, and schools. TEM may motivate students to work harder, guide students in their learning, and provide them with feedback to identify their strengths and weaknesses. As for the teachers, TEM may help improve their teaching, providing both guidance and feedback. Moreover, the school can improve its English teaching program accordingly. To illustrate, a new item type, general knowledge, has been introduced since 2005 in TEM-8, because it is stated in the teaching syllabus that English majors should also master some general knowledge such as culture and literature. With this modification, students will spend more time reading related books; teachers will lay more emphasis on it in their class teaching, and some schools, if they have not included these courses in their curriculum, will consider introducing them. Therefore, TEM has achieved the beneficial consequence.

However, there are also practices that may exert negative consequences.

First, it is common practice that some universities open special TEM preparation courses, which last one semester or more. It is necessary to familiarize students with the test format so that they are able to demonstrate their true language ability in the test. However, spending too much time on test preparation will bring some negative consequences. First, students’ language ability can hardly be improved by special training on test taking techniques, which are not the required skills in the teaching syllabus at all. Moreover, the purpose of TEM is to measure whether students have met the required levels specified in the teaching syllabus. If the English teaching program in the universities can meet this standard, the preparation course is by no means necessary; while if the students cannot meet the requirement, but pass the test just by technique training, the results of TEM will provide false feedback to students, to teachers, and to the school. Lastly, the preparation course may affect the English teaching program because it takes the time that should be set for the other courses specified in the teaching syllabus; it may affect the teachers in that it will bring them unnecessary burden and pressure; and it may affect the students because monotonous skill training will only bore them. In a word, special technique training is not the instructional practice that TEM intends to promote at all.

Second, the assessment reports of TEM are not clear enough, which may reduce the beneficial effect of TEM on providing feedback. The test takers only receive certificates reporting their performance level, such as excellent, good, and pass, while no section scores or ability descriptions are reported. With such a report, students can only obtain a rough idea of their general language ability. However, it is difficult for them to tell their strengths and weaknesses, not to mention directing their efforts in language learning. It is the same with teachers and universities, who may need more detailed description to adjust their teaching and language program accordingly.

Last, it is inevitable that some classification errors will be made—to certify a student who has not met the required level or to fail a student who should be able to pass. To mitigate this detrimental consequence, it is advisable that some complementary measures should be taken, such as teachers’ evaluation and records of class performance, if any critical decisions are to be made. This is just the practice intended by TEM, since decisions concerning penalties for those failing TEM are left in the hands of participating universities (Jin & Fan, 2011), who will make a more comprehensive evaluation of the students.
Decisions

The chief decision to be made on TEM is certification—to certify students’ level of language ability at the end of foundation and advanced stages in their college education program with TEM-4 and TEM-8 respectively. As a national test, TEM is designed by specialists in English, most of whom are teachers and administrators in the participating universities, so it can ensure the consideration of existing educational and societal values and relevant legal requirements in the test development. Moreover, the test syllabus has been published and sample tests are provided, so the test takers should be fully informed of how the decisions will be made. The test takers will be classified only according to the cut score and decision rules of TEM. In this sense, the decision made is equitable to all test takers.

Interpretations

When interpreting the language ability assessed by TEM, we will consider its meaningfulness, impartiality, generalizability, relevance and sufficiency. The definition of the construct for TEM is based on a teaching syllabus—the National College English Teaching Syllabus for English Majors—which serves as a major criterion for its meaningfulness. Besides, the interpretation is impartial in that the test content and format are carefully designed to avoid bias against any group of test takers, who are treated impartially during all aspects of test administration, such as the registration and test taking process. Since the test takers are university students, their target language use domains seem to be more of language teaching and learning situations rather than real life situations. In this sense, the characteristics of TEM tasks should correspond to those general language use tasks, instead of any specific task in a certain field. Moreover, the information provided by TEM is relevant to the decision to be made, because it is designed according to the teaching syllabus to measure whether test takers have met the required levels set by it. This also proves its sufficiency.

However, there are also aspects that may affect the interpretations about the ability assessed by TEM.

First, the fact that speaking is not a compulsory item in TEM may reduce the meaningfulness of its interpretation. In the teaching syllabus, it is clearly stated that the required skills are pronunciation, grammar, vocabulary, listening, speaking, reading, writing, and translation. However, since TEM-4 Oral and TEM-8 Oral are optional, with a test taking population of only 20,000 and 10,000 respectively, the majority of TEM candidates have not had their oral proficiency evaluated, thus making the construct assessed in TEM incomplete. Therefore, when interpreting the test results of TEM, it is difficult to determine whether the test taker has met the required level specified in the teaching syllabus, since one important element, speaking, has been left out.

Second, the wide use of multiple choice questions may pose some problems to the interpretation of TEM test results. In TEM-4 and TEM-8, multiple choice questions account for 60% and 40% of the overall questions. However, multiple choice questions can hardly generalize to real life situations. Beside, the method effect of it will make the test results less meaningful and impartial, since the test takers may get the correct answer by guessing and there are some students who might be better at guessing. It also leads to the training of test taking techniques, which is highly undesirable.

Third, most of the tasks employed in TEM only assess one single skill, such as listening, reading, and writing, which might be replaced by integrated tasks. The only integrated tasks employed in TEM-4 and TEM-8 are dictation and mini-lecture, which assess the students’ ability to take notes while listening to some lectures. It
is a task highly generalizable to classroom teaching. However, the other skills, such as reading and writing, are tested separately. It is advisable that more integrated tasks be employed, because in real life situations, most tasks will involve several language skills. For example, while communicating with someone orally, we have to listen and speak at the same time; and when writing an academic report, we will both read and write.

Last, the relevance and sufficiency of interpretation can hardly be ensured if some other decisions are to be made, such as whether the test takers are qualified for some jobs. Since TEM is a highly recognized English test in China, it is quite likely that it may be used for other purposes, such as selection and placement, both for students and teachers. However, these are not the inferences that can be drawn from the test results.

**Assessment Records**

Tremendous efforts have been made to ensure the consistency of TEM assessment records, such as to follow consistent administrative procedures, to carry out strict rater training, and to employ double marking. Statistics from TEM test centre show averaged internal consistency coefficients from 2008 to 2010 of 0.836 for TEM-4 and 0.815 for TEM-8, which are reasonably high for most test uses (Jin & Fan, 2011).

**Conclusion**

As a large-scale standardized test, TEM has been constantly under improvement, which can be demonstrated by the recent revisions of the teaching syllabus, test syllabus, and the test, as well as the numerous researches conducted on its validity and reliability. However, due to the increasing test taking population and the limited resources, practicality has become another factor that we have to consider. Therefore, when designing and using the test, we have to take all these into consideration.

**References**


