Achievement of Construct Validity in Language Testing
—Avoiding Test Bias

HUO Ying, LIU Wei
Northeastern University, Shenyang, China

As an independent discipline, language testing is an important means of inspecting and assessing teaching effectiveness and learning outcomes. Construct validity is the basis for language testing, determining the purpose of the test and collecting evidences for inspecting the test effectiveness. For different test subjects, there will be a tendency to produce test bias, influencing the realization of construct validity. Scientific language testing requires testers in the test to avoid test bias reasonably and effectively, achieving construct validity and making the test more scientific.

Keywords: language testing, construct validity, test bias, scientificity

Introduction

Language testing is an important way to inspect test levels and abilities, which is not only affected by the test reliability, validity, and discrimination, but also closely related to individual characteristics of the subject. The validity of the test is to examine what the tester wants to measure, an important aspect to achieve the test. Without test bias, in other words, test performance should not be affected by individual characteristics of the subject, the subject can play up to his normal real level.

Construct Validity in Language Testing

The Concept of Construct Validity

In his book Language Testing in Practice, Lyle F. Bachman explains that “the so-called construct validity is subordinate to the sense and rationality of interpretation of the language test scores, which means this interpretation is the assessment of language skills of the subject” (Bachman & Palmer, 1996, pp. 254-271). Bachman believes that through construct validity, we can not only assess the language ability of the subject, but also estimate the reasonability of the target language adopted in the test by means of interpreting the test scores. The concept of construct validity was first jointly proposed by Cronbach and Meehl in their book Construct Validity in Psychological Test, where they considered structure, as human’s imaginary property, could be reflected in the test (Cronbach & Meehl, 1955, pp. 281-302). This structure is abstractive, reflecting intelligence, attitude, and understanding ability which cannot be directly measured. So by using the concept of structure,
language testers can design a reasonable test to measure language skills of a subject, proved by the test results, to achieve the construct validity. Briefly, construct validity is to interpret test scores in order to assess the language proficiency of the subject and test tasks.

Construction of Construct Validity

Because the ability to be measured cannot be directly observed, testers must deduce by the directly observed performance; since the ability is only theoretical, testers are required to assume in advance that this ability will affect how the subject uses the language and performs in the test, which means when building a valid structure, testers need to check out the assumed relationship between pre-defined test scores and the ability to be measured. So, to build construct validity in a language test, according to Popham, a hypothesis structure should be first proposed to explain testing performance, then a number of assumptions of testing performance are deduced from the structure, and finally these assumptions are inspected by logic and experience (Popham, 1988, pp. 25-30).

Importance of Construct Validity to Language Testing

Construct validity is the basis for all the various validities, which determines the nature of the entire examination and the system the exam belongs (LI, 1996, pp. 39-40). Construct validity determines the direction and content of language testing, vital to the authenticity and rationality of the interpretation made by testers to explain test scores. It is particularly important for language testing since it shows whether this test could measure the subject’s real language ability.

Test Bias Affecting Construct Validity

The Concept of Test Bias

In the testing process, test bias refers to the difference in performances of subjects due to individual characteristics and testing tasks, which affects testers’ assessment of subjects’ real language ability on the basis of testing scores (Bachman, 1990, pp. 81-110). For the language test subjects, their real language capacities and personal factors work together to affect their performance in the test. Personal factors include language skills (language knowledge, strategic competence, or metacognitive capacity), topic knowledge, and affective schemata. There are various test biases, such as misunderstanding of test scores, sex discrimination, race discrimination, wrong assessment of the level of the subject, priority of the test to some subjects, inappropriate selection procedures, inadequate criteria evaluation, and uncomfortable test atmosphere and the environment. In the actual language test, test bias is much more complicated because cultural and educational backgrounds cannot be clearly separated from the language skills we want to test.

Possible Test Biases

When building a valid structure, testers should take full account of personality traits of the subject and impact of test tasks on the subject, in order not to pose priority to some subjects. Testing bias may relate to cultural background, background knowledge, cognitive features, local language, ethnicity, gender, age, etc. It is also responsibility of the testers to consider whether test tasks will affect test performance of different individuals or not, such as in the test environment, the guide format, input, and expected manner of examination.

Importance of Avoiding Test Bias to Achieve Construct Validity

Test bias should be reasonably evaded for a fair test environment, ensuring the validity of test results,
increasing the impact of language proficiency on test scores, assessing a subject’s real language skills through reasonable interpretation of test scores, and enhancing scientific nature of language testing.

Reasonably Avoiding Test Bias for Implementation of Construct Validity

Personality Traits of Subjects

Avoid bias on cultural background. When studying option reflection theory, Chen and Henning have found that: In vocabulary test, some options would be favorable to some individuals of specific language and cultural background (Chen & Henning, 1985, p. 272). Take options with the phrase “meet one’s Waterloo” as an example. The test favors some testers with European and American cultural background because they are familiar with Battle of Waterloo history, which, like the Chinese “the failure of Guan-yu”, means failure in a lawsuit or match. Therefore, in the design of test content, test organizers should first research on basic information of subjects and have some insight into their cultural background, to avoid the test bias on subjects of certain cultural background.

Avoid bias on background knowledge of test contents. If subjects learn and understand some relevant knowledge of some contents before the test, they obtain different scores, particularly in reading comprehension and cloze questions. To check the comprehensive reading abilities, testers design a reading comprehension on mechanical design principles, questions favorable to majors in mechanical engineering, who may answer some of the options depending on the learned expertise. There comes the test bias. For examining reading ability in general, the knowledge to be tested should be defined as the background knowledge. So in the test design and interpretation of test scores, testers must differentiate language proficiency from background knowledge, and believe that language skills of the subject can be reflected in the specific context.

Test Task Features

Reasonable test task settings. If the test environment favorable to certain subject, construct validity is undermined. For example, in the computer-based test (TOEFL), certain subjects are not familiar with the operation or the program; the final test results not only reflect their language ability but are also affected by the test setup. So testers should inform subjects of the test formality in advance, to ensure the test is fair and subjects will all give full play of their language abilities.

Appropriate guide format. First, testers should notice in advance the subjects of the test structure, like composition, the number of questions, the characteristics, the order, the importance, value and time allocation of the various parts; the instruction should be brief and clear, since instructions which are too complex or in target language will be favorable to subjects with strong language ability and leave others at an unfair disadvantage; subjects need to know the testing process and the specific time schedule; they should understand the scoring mechanism, standard answers, and scoring process.

Proper input. If the test were to test writing skills, different essay topics will affect the performance of the subject, and produce test bias. On the input format of test tasks, testers must select proper means (audio, video), form (language, non-verbal, or both), language (native language, target language, or both), length, type, and speed in accordance with the test content and language abilities of all subjects; in language designed to test tasks, the difficulty of the language depends on the analysis of language abilities of all subjects and purpose of the test (placement test, level tests, or graduation test); at the same time, it is important to assess the familiarity of
subjects to the tested topic, like in the oral test, it will produce a test bias for some subject if they have learnt about the house rental services and relative key sentences.

**Reasonable expected examination method.** In the listening test, if not all of the subjects are able to write the answers to the questions, there will be test bias on those who are not proficient in writing in target language. It is required to design reasonable expected examination method in compatible with the test content and language abilities of all subjects, similar to the above reasonable input.

**Relationship between reasonable input and expected examination method.** Topic knowledge will affect the performance of the subject, as requested in oral test to talk about the theory of Western psychology rather than simple weather, which affects the performance of some of the subjects. So testers, when handling the relationship between input and expected examination approach, need to design reflecting forms (correlation of all questions, if the performance of each subject interacting with each other and test continuity), ranges (amount of input), and the mode (direct or indirect relationship between the expected examination method and test tasks).

**Conclusion**

Scientific language testing enables test an important means to detect and evaluate the teaching effectiveness. With scientific analysis of the test results, we can assess students’ mastery of language knowledge, teaching effectiveness, and soundness of the entire education system (DANG, 1999, pp. 58-61). Achievement of scientific language testing is based on reasonable construct validity, which is continuous and repetituous throughout the test. So in the test, testers should observe and evaluate language proficiency of all subjects and individual differences, to achieve the purpose of language testing without test bias. In this way, all subjects can give full play to their language ability.

**References**