

Abstract

Assessment of Reading in English as a Foreign Language: Investigating the Defensibility of Test Items

Celso Henrique Soufen Tumolo

Universidade Federal de Santa Catarina
2005

Profa. Dra. Lêda Maria Braga Tomitch
Supervising Professor

In the present study, the defensibility of test items is investigated in three different testing situations: proficiency tests, classroom tests, and university entrance examinations. The defensibility is judged using the concept of validity as put forth by Messick (1989) for educational testing and by Bachman (1990) for language testing, i.e., in terms of validity of the interpretations and actions based on test item responses, considering the justifications coming from evidential basis and consequential basis. For the investigation, in terms of evidential basis, for construct-related evidence, constructs of language ability and reading ability are used, and for criterion-related evidence, the criterion defined for university studies by Weir, Huizhong, and Yan (2000) and the framework proposed by Bachman (1990) for the assessment of authenticity of test tasks in relation to the criterion tasks are used. In terms of consequential basis, an appraisal of the consequences is based on Bachman's (1990) notion of fairness and Shohamy's (2001) critical perspective of the use of tests. The method used for the investigation is based on recent notion of validity as argument-based proposed by many scholars in the area of testing. It is also based on the validity table proposed by Chapelle (1994) for considerations of the forces of the arguments, both in favor, or against the validity of the interpretation of ability based on the items, with the additional feature of a third column, with the refutation of the argument against, as suggested by Davidson (personal communication, 2004). Considering the arguments provided for each item, it is possible to conclude that some items are defensible and some are not. Some defensible items focus on skills such as

identification of syntax and cohesion, in particular lexical cohesion, inference of word meaning, elementary and propositional inferences, and identification of functional value. Some non-defensible items have the construct-irrelevant aspects of assessing constructs other than reading ability, such as vocabulary knowledge, background knowledge, writing ability, illustration comprehension, and the criterion-irrelevant aspect of assessing tasks not present in the criterion, such as the tasks specific for reading non-academic texts as poems, narratives, advertisements, and technical problems impeding test takers to perform at their level of ability. Other non-defensible items have the consequence of being biased once specific background knowledge is presupposed for their answers. The political and pedagogical implications of the conclusions claim for the choice or development of items incorporating features of validity, in all the facets, considering both evidential basis and consequential basis, so as to make them more defensible.

Number of pages: 246

Number of words: 66.285