

2nd GLOBAL CONFERENCE on LINGUISTICS and FOREIGN LANGUAGE TEACHING,
LINELT-2014, Dubai – United Arab Emirates, December 11 – 13, 2014

An Analysis of Gap Fill Items in Achievement Tests

Halka Capkova^{a*}, Jarmila Kroupova^a, Katerina Young^a

^aUniversity of Economics, Prague

Abstract

This article revives the discussion over measurements of validity in criterion referenced (CR) tests. It presents how the principles of Classical Testing Theory (CTT), normally associated with norm-referenced tests, were applied to the Business English achievement tests at the University of Economics, Prague, Czech Republic. Firstly, measures of validity in criterion-referenced tests, test purpose, and test specifications are discussed. Next, a 10-item vocabulary gap fill subtest is subjected to a detailed analysis through the use of facility and discrimination indices. Key and distractor analyses of each item are then performed. The insights gained from such analyses are examined in relation to the cyclical test design process of constant review of items so that a high level of standardization is achieved. This paper thus provides teachers with simple tools to build valid language gap fill tests which reflect the criteria of accurate and equitable testing.

© 2015 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of Academic World Research and Education Center.

Keywords: distractor analysis; achievement test; gap fill; validity; construct; facility index; discrimination index

1. Introduction

1.1. Overview of Testing Approaches

Statistical analyses of reliability and validity of individual items have been mainly associated with norm-referenced testing (NRT). In norm-referenced tests, such analyses are used to ascertain that a test reliably discriminates among test takers. In other words, discrimination is the goal on the basis of which important decisions, often regarding selectivity to educational programs or professions, are made. In criterion-referenced testing (CRT), particularly in classroom-based testing situations, the goal is to show a student's progress or achievement. Researchers agree that the use of statistical analyses for such CRT purposes is futile as (Popham and Husek

* Halka Capkova Tel: +5564636

E-mail address: capkova@vse.cz

Popham, 1971, p.29) explain: “Criterion-referenced measures are validated primarily in terms of the adequacy with which they represent the criterion. Therefore, content validity approaches are more suited to such tests. A carefully made judgment, based on the test’s apparent relevance to the behaviours legitimately inferable from those delimited by the criterion, is the general procedure for validating criterion-referenced measures.”

Since many institutions now use a cut-off point in form of a percentage, i.e. often an externally imposed „standard“ to establish a student’s mastery of the content, and make their pass or fail and other administrative decisions based on this result, test designers have even greater responsibility to develop fair and accurate tests. This article intends to demonstrate that criterion-referenced tests are NOT a dead enterprise, as Fulcher and Svalberg rightfully argue (2013 p.1), but form the bulk of a teacher’s and student’s everyday educational experience. Teachers can thus empower themselves by employing simple norm-referenced (NR) statistical tools in judging appropriateness of individual items on the tests which they create. In this way, norm-referenced and criterion-referenced testing can complement each other instead of bitterly occupy opposing positions. The analysis of the following exercises is an attempt at such synthesis.

1.2. *Validity*

Test validity is a measure of how accurately a test score reflects a test-taker’s real-life language ability. For a test to be valid, all items have to measure what the test is supposed to measure, i.e. what researchers in the field call a “construct”. If a test or a section of a test measures vocabulary skills, as in the case of the example in this paper, then all items in this particular section or test need to be generated around this construct. Thus, as information about the test-taker’s ability is collected item by item, the inferences made about the test taker’s skills from the scores gained on all of these items are justified. In criterion-referenced (CR) testing, validity is demonstrated in the process of test creation. Items are continuously created, piloted, used in real test, analysed and modified based on a clearly stated purpose and test specifications in an on-going cyclical process.

To exhibit validity of the Department’s mid-term or final exam tests, this paper employs the principles of Classical Testing Theory (CTT), which are traditionally associated with norm-referenced tests. While norm-referenced tests are designed to discriminate among students, achievement tests are designed to show that students have learned what they have been taught, there is a need for both at the university level. This unique combination gives rise to a situation where teachers want their students to show what they have learned, yet they also have to “separate” out students who achieve significantly below the expected level from those who pass the test with distinction. The values of facility and discrimination indices as recommended by various researchers differ (See Green and Ebel below), but they all reflect the need for some measure of discrimination between the lower and the upper ridge of students. The purpose of university level language tests thus requires the teachers to strike the right balance. It depends on each particular university context how much they want to discriminate among these two ends of the student spectrum and decide on “discrimination” boundaries for their tests. NR methodology is a tool to make them visible.

1.3. *Local Situation, Test Purpose, and Test Specifications*

There are 19,000 students enrolled at the University of Economics. The English Department offers nearly 50 different courses, including courses for distance students and the University of the Third Age (aimed at the education and stimulation of mainly retired members of the community - those in their third 'age' of life). Analysed scores were selected from the largest compulsory course in terms of enrolment (2,000 students), English for Business Studies, which is targeted at the B2/C1 level of CEFR and is offered every semester. Because of its size and frequency, computer-based testing (CBT) seems to be the ideal option. Not only does it enable the testing of large numbers of students, it also allows the standardisation of tests and test administration conditions. The decisions made about the presentation and delivery are congruent with the test purpose, i.e. gaining information on the achieved level of Business English skills, knowledge, and abilities of large numbers of students on the basis of which they are awarded credits.

In order to prepare students for computer based testing which requires additional skills such as reading from the screen, practice exercises for home preparation in the e-learning section have been created. These exercises closely

Download English Version:

<https://daneshyari.com/en/article/1109294>

Download Persian Version:

<https://daneshyari.com/article/1109294>

[Daneshyari.com](https://daneshyari.com)