

Automated Test Assembly for Criterion-Referenced Equivalent Tests

Chuan-Ju Lin

National Tainan Teachers College, 33, Sec. 2, Su-Lin St., Tainan, Taiwan, 700,
cjulin@ipx.ntntc.edu.tw

Keywords: automated test assembly, criterion-referenced equivalent tests, Weighted Deviations Model (WDM) heuristic, test-assembly targets.

Abstract

The goal of automated assembly of alternate test forms is to produce multiple test forms that are equivalent to some target. To achieve that goal, the alternate forms are usually assembled according to the content and psychometric specifications obtained from a reference test. Equivalent-test assembly techniques based on conventional test-parallelism definitions constrain alternate test forms to have the same characteristics for each test taker. The test scores derived from such assembled forms are beneficial to indicate the standing of an examinee or make inferences about an individual's performance. However, the requirement that each examinee should be provided test forms with identical test properties can be too stringent and unnecessary for certification and licensure testing. In such setting, tests are intended to classify examinees into categories (e.g., pass vs. fail, or master vs. non-master), and information about an individual's standing relative to a tested population is not the main concern. One way to assemble equivalent test forms for classification testing can be to constrain alternate test forms to have identical test properties for, for example, two contrast groups (or categories) defined by a cut score (i.e., master vs. non-master groups), instead of for every test taker. For example, multiple test forms can be assembled to match target test difficulty for two groups or two target expected p-values. The resultant alternate test forms can be regarded as criterion-referenced equivalent or parallel.

The purpose of this study is to investigate the degree of classification precision and consistency, and equivalence of test forms assembled with the weighted deviations model (WDM) heuristic (Swanson & Stocking, 1993) using a novel test-assembly target and the conventional test-assembly targets (e.g., target test characteristic or information functions). A simulation will be, under various conditions resulting from manipulating independent variables (e.g., the type of statistical targets, the location of the cut-off scores on the proficiency scale), conducted to examine classification precision and consistency of test forms. Test parallelism will be evaluated by examining average test overlap rate, percentage of content specification met, and equivalence of statistical properties.

References

Swanson, L., & Stocking, M. L. (1993). A model and heuristic for solving very large item selection problems. *Applied Psychological Measurement*, 17, 151-166.