Journal of Educational and Behavioral Statistics, Summer 2004, Vol. 29, No. 2, pp. 177-199.

Evaluation of the CATSIB DIF Procedure in a Pretest Setting

Ratna Nandakumar
Louis Roussos

Abstract

A new procedure, CATSIB, for assessing differential item functioning (DIF) on computerized adaptive tests (CATs) is proposed. CATSIB, a modified SIBTEST procedure, matches test takers on estimated ability and controls for impact-induced Type 1 error inflation by employing a CAT version of the SIBTEST "regression correction." The performance of CATSIB in terms of detection of DIF in pretest items was evaluated in a simulation study. Simulated test takers were adaptively administered 25 operational items from a pool of 1,000 and were linearly administered 16 pretest items that were evaluated for DIF. Sample size varied from 250 to 500 in each group. Simulated impact levels ranged from a 0- to 1-standard-deviation difference in mean ability levels. The results showed that CATSIB with the regression correction displayed good control over Type 1 error, whereas CATSIB without the regression correction displayed impactinduced Type 1 error inflation. With 500 test takers in each group, power rates were exceptionally high (84% to 99%) for values of DIF at the boundary between moderate and large DIF. For smaller samples of 250 test takers in each group, the corresponding power rates ranged from 47% to 95%. In addition, in all cases, CATSIB was very accurate in estimating the true values of DIF, displaying at most only minor estimation bias.