Abstract (Document Summary)

In this simulation study, we investigate the power and Type I error rate of a procedure based on the mean and covariance structure analysis (MACS) model in detecting differential item functioning (DIF) of graded response items with five response categories. The following factors were manipulated: type of DIF (uniform and non-uniform), DIF magnitude (low, medium and large), equality/inequality of latent trait distributions, sample size (100, 200, 400, and 800) and equality or inequality of the sample sizes across groups. The simulated test was made up of 10 items, of which only 1 contained DIF. One hundred replications were generated for each simulated condition. Results indicate that the MACS-based procedure showed acceptable power levels ($\geq .70$) for detecting medium-sized uniform and non-uniform DIF, when both groups' sample sizes were as low as 200/200 and 400/200, respectively. Power increased as sample sizes and DIF magnitude increased. The analyzed procedure tended to better control for its Type I error when both groups' sizes and latent trait distribution were equal across groups and when magnitude of DIF and sample size were small.