
Statistical issues in the item bank development for adaptive testing

Maria Eugénia Ferrão^{1,2} and Paula Prata^{1,3}

¹ University of Beira Interior, Portugal

² Centre for Applied Mathematics and Economics (CEMAPRE), Portugal

³ Instituto de Telecomunicações (IT), Portugal
meferrao@ubi.pt, pprata@di.ubi.pt

Any conceptual assessment framework faces challenges such as the complexity of knowledge, capacities and skills to be assessed, and the increasing usability of web-based assessments, which requires innovative approaches to the development, delivery and scoring of tests. Considering the benefit for educational improvement, the next generation of educational assessment will capitalize on the power of the web to deliver more efficient tests. The prior optimization of the test size conditional on a given level of measurement error makes the simulation of adaptive tests as a crucial step.

We present a simulation study implemented in SimuMCAT. The results suggest the considerable reduction of bias as the number of items administered increases, the occurrence of ceiling effect in very small size tests, the full agreement between true and empirical distributions for computerized tests of length smaller than the paper-and-pencil tests. The extreme reduction of test size provides overall scores poorly estimated and with ceiling effect. This may be due to the insufficiency of the item pool in providing items properly mentored for outstanding students. Even in this situation, the comparison between the true and estimated score distributions suggests their agreement at the level of significance of 5%.