

Differential Item Functioning in Gender and Living Background Groups in the GEPT

Jessica R W Wu
The Language Training and Testing Center
jw@lrtc.ntu.edu.tw

Analyses of GEPT score data have consistently reported two phenomena in regard to GEPT test-takers' performance: the female group outperforms the male group in the listening and reading tests, and the test-takers living in the Taipei area outperform those in the other areas in the listening tests. The phenomena seem reasonable because it is commonly believed that females generally tend to be more successful than males in learning languages, and learners living in Taipei get more exposure to English learning resources, such as media and native speakers, than those living in the other parts of Taiwan. However, such conventional, intuitive inferences about variations in the GEPT performance between different groups of test-takers may not be accurate. In fact, the differences may be influenced by construct-irrelevant variance which is introduced through test-takers' characteristics, such as gender and living background (Kunnan, 2000; McNamara & Roever, 2006; Saville, 2003; 2005; Shohamy, 2000). In other words, it is possible that the male group and the non-Taipei group may have been disadvantaged by bias in the GEPT. Therefore, to detect whether such bias exists in the GEPT, the present study investigates differential item functioning (DIF) in gender and living background groups in the 2006 GEPT listening tests at the elementary, intermediate, and high-intermediate levels, using the Mantel-Haenszel approach (Holland & Thayer, 1986). The results show that the GEPT-Listening tests are not differentially easier for female test-takers or for test-takers who live in the Taipei area, with only 3% of the test items displaying medium DIF. To further maintain test quality, each of the items showing DIF is examined and the sources of DIF are hypothesized.

Investigations of DIF are integral to efforts to enhance test fairness, which has increasingly been regarded as both a professional obligation and an ethical necessity in the field of language testing. However, in local assessments, and in particular the large-scale tests, there has been very little discussion of this issue. The purposes of this present study are to address that gap and to highlight the need for further DIF studies that are aligned with efforts to enhance test fairness within the Taiwanese context.