

IRT AND LATENT REGRESSION FOR DETECTING CHANGES IN READING ABILITY OVER TIME

Zoltán Lukácsi
Budapest Business School

Keywords: IRT; structural analysis of a univariate latent variable; reading comprehension

In July, 2011, Euro Examinations launched an item bank recalibration project, as the formerly estimated item parameter values endangered the fairness of logit-based score reporting. The scheme paved the way for an analysis of differences in test taker ability over time. The research focus of the present study was whether candidate ability as measured by the Reading Paper on the Euro was constant in the population over test administrations.

In a common-item nonequivalent groups design (Kolen, 2007, p. 45), eleven general English tests at level B2 taken by 17,808 candidates and built up of altogether 162 dichotomous items were jointly calibrated onto a common scale. When relating the latent variable of reading comprehension to the explanatory variable of exam period, the measurement model I applied was OPLM (Verhelst, Glas, & Verstralen, 1995), and the structural model was SAUL (Verhelst & Verstralen, 2002).

The results of the latent trait analysis led the Exam Office to conclude that test taker ability showed significant differences between administrations, thus the null hypothesis that candidates came from the same population was refuted. However, effect sizes defined as Cohen's *d* remained small and ranged between 0.07 and 0.28.

The implications of the study are twofold. First, differences in overall test difficulty were proven, and so the adequacy of setting the standard in the form of a raw score was challenged. Second, since candidate ability was found unstable over test administrations, expecting a fixed proportion of successful candidates is seen as unrealistic.

F1