Mini Symposium

Test Equating. Overview and Trends

December, 15 2014

Ekonomikum, Room B115

Uppsala University

Program

9.30 – 10.00 Coffee

10.00 – 10.45 Jorge González, Pontifcia Universidad Católica de Chile, Chile

Statistical Model Framework for Equating: Sampling Distributions, Parameters, and all of that

Statistical models assume that observed data are the realization of random variables following some probability distribution. A statistical model consists of a class of probability distributions which we assume might have generated the data. Such distributions are characterized by parameters defined on a parameter space. A particular statistical problem in the field of test measurement consists in modeling the relationship between a score on one test form as a function of the score from another test form. Such problem is called equating. Here, the parameter of interest is the so called equating transformation, which maps the scores of one test forms into the scale of the other. In this talk, I will discuss the role of random variables, probability distributions and parameters as involved in equating. By doing so, I will discuss about the possibility of going towards a formal statistical model framework for equating.

10.45 – 11.25 Marie Wiberg Umeå University, Sweden

The concept of local observed-score equating

Test equating is used in order to determine comparable scores on different test forms. Several methods can be used depending on what assumptions are made and which data collection design has been used. Regular equating methods which have one equating transformation can be divided into traditional methods, which include methods built on classical and modern test theory as well as modern methods, e.g. kernel equating methods. A new concept of equating which allowed for several equating transformations was proposed in 2001 and was soon labelled local equating. Local equating emerged from Lord (1980) concept of equity and the fact that there are some persistent problems with regular equating methods, especially the difficultness to satisfy the criteria of equity and population invariance. If an equating does not fulfil equity this may cause the equatings to be biased. During the past five years several local equating methods have been proposed for different situations. In this presentation the equating

criteria, the concept of local equating as well as different local equating methods will be discussed in light of regular equating methods.

11.30 – 13.00 Lunch

13.00 – 13.45 Björn Andersson, Uppsala University, Sweden

Item Response Theory Observed-Score Equating

Equating ensures that test scores from separate administrations of a particular standardized test are comparable. Item response theory (IRT) equating methods have recently become more accessible with new software implementations. This presentation introduces a general framework for IRT observed-score equating using kernel equating. It is shown how equating can be conducted using dichotomous and polytomous IRT models. The IRT approach is compared to non-IRT methods using simulated data and data from standardized achievement tests.

13.45 – 14.15 Coffee

14.15 – 15.00 Alina von Davier, ETS, USA

Test Score Equating: A Scientific Process or a Craftsmanship?

Test equating is the statistical procedure for making test scores comparable across different test forms that were developed according to the same specifications or "blueprints."
In this presentation I argue that test score equating is part of a statistical modeling process. The issues of data collection, avoiding confounding effects, assumptions, and model fit are pertinent steps in the equating process as in any statistical modeling process. Moreover, taking the perspective of equating as a model leads to a modular infrastructure for the operational implementation, to an efficient training system for new psychometricians, and most importantly, to consistent, replicable, and accurate measurements. In this presentation I will mention the observed-score equating framework, the kernel method of test equating, and the various software that are available for conducting equating operationally. I also argue that test equating should be part of any statistics and measurement program, because an academic training stimulates creativity and inquiry in ways an operational context with its complex responsibilities and demands cannot.

15.00 closing