Cross-country Comparability of Questionnaire Scales in Large-Scale Surveys

OECD methodological conference

Paris, 8 and 9 November 2018

Concept note

In recent decades, an increasing number of international large-scale surveys have been developed with the goals of assessing skills and comparing attitudes and values across different countries and different cultures. The former are typically assessed on the basis of a cognitive test, whose results are then modelled using Item Response Theory in order to obtain an estimate of the proficiency of respondents. The latter usually rely on scales derived from sets of questions/statements that are treated as indicators of the targeted latent construct. These indicators are often in the Likert-type format, with several answer categories representing different levels of agreement with a statement or different level of frequencies. PIAAC, PISA, and TALIS all use scales for measuring various non-cognitive constructs.

The key methodological challenge faced by these surveys is the comparability of measurements across countries and languages and over time. The use of the same instruments and of the same administration procedures in different cultures does not on its own guarantee comparability of the scale scores. Thus, in order to validly compare scale scores across groups/countries, a necessary first step is to establish measurement equivalence (i.e. invariance). The most commonly applied procedures to establish invariance of cognitive and non-cognitive scales across countries in large-scale surveys are developed within the framework of confirmatory factor analysis (CFA) and item response theory (IRT). Commonly applied criteria to evaluate the fit of such models that work well in the comparison of a small number of countries often do not work as well in practice in large-scale applications.

This conference aims at exploring novel approaches in measurement-equivalence research and at highlighting promising areas to advance the field of invariance testing. It also aims at delineating recommendations and guidelines for dealing with the issue of measurement equivalence, both in the preparation of public-use datasets and in the analytical work based on the data collected in these surveys. The target audience for this conference includes researchers and analysts working with international datasets as well as the key suppliers in the industry of international large-scale surveys.

Practical information

Venue: The conference will be held in the Auditorium of the OECD Conference Centre 2 rue André-Pascal, 75016 Paris, France.

Registration: To register, visit http://www.oecd.org/education/oecd-methodological-conference.htm and complete the registration form by 18 October 2018. Participation is free of charge.

Contact: edu.invariance@oecd.org

Day 1: Thursday 8 November

14:00-14:20

Item 1. Welcome and introduction

Item 1.a. Welcome and opening remarks
OECD

Item 1.b. Introduction to the conference
Fons van de Vijver (Tilburg University)

14:20-16:00

Item 2. Recent developments in the context of measurement invariance testing in CFA

Item 2.a. Multigroup-CFA and Alignment
Bart Meuleman (KU Leuven)

Item 2.b. Bayesian approximate measurement invariance
Rens van de Schoot (Utrecht University) **and Kimberley Lek** (Utrecht University)

Item 2.c. A Monte Carlo Simulation Study to Assess the Appropriateness of Traditional and Newer Approaches to Test for Measurement Invariance

Artur Pokropek (Polish Academy of Science)

16:00-16:30: Coffee break

16:30-18:30

Item 3. Extensions of measurement invariance testing to categorical data and IRT

Item 3.a. An IRT-Based Item-Fit Statistic for the Analysis of Measurement Invariance
Janine Buchholz (DIPF)

Item 3.b. Bayesian Marginal Measurement Invariance Testing
Jean-Paul Fox (University of Twente)

Item 3.c. Discussion: Growth in international assessments: Measurement equivalence challenges and a look at current practice

Leslie Rutkowski (Indiana University)

Item 3.d. Discussion: Why we care about comparability and how *not* to test it

Matthias von Davier (National Board of Medical Examiners)

Day 2: Friday 9 November

9:00-10:00

Item 4. Group comparison in Latent Class Analysis

Item 4.a. Multigroup and Multilevel Latent Class Analysis
Michael Eid (Free University of Berlin) - *remote presentation*

Item 4.b. Discussion
Jeroen Vermunt (Tilburg University)

10:00-10:30: Coffee break

10:30-12:30

Item 5. Can innovative item formats lead to greater comparability?

Item 5.a. Situational Judgement Tests and Anchoring Vignettes in PISA 2012
Jonas Bertling (ETS)

Item 5.b. Situational Judgement Items in TALIS 3S
Pauline Slot (Utrecht University) **and Trude Nilsen** (University of Oslo)

Item 5.c. Comparison of Design and Analysis Approaches to Enhance Comparability
Jia He (DIPF)

Item 5.d. Discussion
Pat Kyllonen (ETS)

12:30-13:00

Item 6. Concluding remarks: recommendations for the present and future of large-scale surveys
Fons van de Vijver (Tilburg University)
OECD