**Paper Session:**

**Response accuracy in large-scale assessments**

**Chair**

Dorothée Behr *(GESIS – Leibniz-Institute for the Social Sciences, Germany)*

**Presentations**

1. **Title:** Effects of response styles on secondary analysis in International large-scale assessments.
   **Authors:** Tomasz Żółtak, Artur Pokropek & Marek Muszyński *(Polish Academy of Sciences, Poland)*

2. **Title:** Is careless responding also a problem in face-to-face mode? Analysis of PIAAC noncognitive data.
   **Authors:** Marek Muszyński, Tomasz Żółtak & Artur Pokropek *(Polish Academy of Sciences, Poland)*
1. Effects of response styles on secondary analysis in International large-scale assessments.

Authors
Tomasz Żółtak, Artur Pokropek & Marek Muszyński (Polish Academy of Sciences, Poland)

Presenter
Tomasz Żółtak (Polish Academy of Sciences, Poland)

Abstract
Noncognitive constructs (personality traits, attitudes, interests, etc.) are of great interest in every area of the social sciences. Self-report scales are the main method used to measure them, present in almost all international large-scale assessments (ILSAs), including PIAAC. However, their use does not come without problems as self-reports are prone to various response biases (Khorramdel & von Davier, 2014; Meade & Craig, 2012; Paulhus, 1991). In this presentation, we will test how one of the possible biases, namely response styles (RS), affect different types of statistical modeling. Firstly, we conduct extensive simulation studies where different types and levels of RS (extreme (ERS), middle/midpoint (MRS), and acquiescence (ARS)) are generated and applied to statistical models. Surprisingly, very few simulation studies were conducted to evaluate the influence of RS on different types of statistical modeling. Ferrando and Lorenzo-Seva (2010) and Pleninger (2017) analysed the influence of RS presence on model fit, factor loadings recovery, scale reliability, validity, and scores, but these two studies are one of the very few that compared systematically the consequences of RS for the quality of the data. Moreover, the mentioned studies focused only on basic aspects of the measurement models and not on RS consequences for models usually used in secondary analyses, like analysis of variance (ANOVA), regression analysis, multilevel models, nor for more complex measurement issues like measurement invariance. We address this gap in the proposed study which complements previous results and expands them by presenting models more closely related to substantial hypothesis testing. Secondly, we will conduct a series of analyses using PIAAC scales. Khorramdel, von Davier & Pokropek (2019) showed that response styles are present in PIAAC data, at least to a small extent. We will conduct a series of sensitivity analyses to probe the extent to which substantial conclusions based on PIAAC data might be altered by RS bias. Additionally, we will present a new statistical R package “rstyle”, designed to generate data affected by a wide variety of RS under different conditions and assumptions, allowing for handy assessment of properties of selected statistical models using Monte Carlo simulations.

2. Is careless responding also a problem in face-to-face mode? Analysis of PIAAC noncognitive data.
Authors
Marek Muszyński, Tomasz Żółtak & Artur Pokropek (Polish Academy of Sciences, Poland)

Presenter
Marek Muszyński (Polish Academy of Sciences, Poland)

Abstract
Careless or insufficient effort responding (C/IER) is one of the main causes of low data quality in noncognitive assessments (Krosnick, 1991, 1999; Meade et Craig, 2012). However, a vast majority of the literature analysing this bias is based on either simulation or self-administrative (self-completion) data (Bowling et al., 2020; Silber et al., 2019). In this paper we will analyse whether quality of the data collected in face-to-face mode is also potentially threatened by C/IER (as it was shown that it is indeed threatened by other response biases, e.g. acquiescence (Aichholzer, 2013; Rammstedt et al., 2017) or self-enhancement (Palczyńska & Rynko, 2020). Although a large comparability between face-to-face and self-completion modes was proved (Cernat & Revilla, 2020), these analyses did not comprise C/IER comparison. In order to conduct such analysis we will use PIAAC noncognitive data and an ample set of C/IER indices, including flagging potential outliers (Emons, 2008, 2009; Mansolf et Reise, 2018; Meade et Craig, 2012), testing respondents’ intraindividual variability (Dunn et al., 2018; Marjanovic et al., 2015) and gauging intraindividual consistency (Curran, 2016; Fronczyk, 2014; Huang et al., 2012) in order to identify aberrant patterns/respondents in the dataset. Moreover, we will study C/IER impact on scales’ reliability and validity. Furthermore, we will also check how participant’s gender, age, educational attainment and social status are related to pattern and amount of C/IER. In addition, we will conduct a sensitivity analysis to assess C/IER effect on substantial analyses, using latent regression or confirmatory factor analysis (CFA). We will also analyse if interviewer input variables from the observation module are related to C/IER. Last but not least, we will use interviewer data from one of the national PIAAC datasets in order to test whether C/IER indices evince any inter-interviewer variability (cf. Menold & Kemper, 2013). Additionally, basing on Monte Carlo simulations, we will assess the sensitivity of the C/IER detection tools in simulation settings reflecting methodology of PIAAC and alike studies.