

# The Impact of Measurement Non-invariance in Target Variables on Machine Learning Prediction

*Thursday 24 July 2025 10:20 (15 minutes)*

Psychology is increasingly interested in the prediction of psychological constructs via machine learning (ML) models, for example, predicting a person's personality or intelligence. To measure these psychological constructs, psychologists often draw on questionnaire data. In supervised ML, these measurements are then used as target variables (i.e., the “ground truth”) for model training. Recently, Tay et al. (2022) introduced a conceptual framework that outlines various sources of bias throughout the ML modeling process. One potential bias is non-invariance of the questionnaire data across groups that is used as target values for supervised learning. As Tay and colleagues state, if the questionnaire used to collect the target data produces different expected scores between two groups with the same true score, this might bias the predictions of the final ML model. Specifically, two groups with the same underlying true score on the construct of interest might receive different predicted scores by the ML model. The goal of this work is to assess the actual impact of a lack of measurement invariance in target variables on the predictive performance of ML models. We address and investigate the impact of non-invariance in three different ways: empirically, semi-empirically, and simulation-based. We also discuss possible solutions to counter the impact of non-invariance in target variables.

**Primary authors:** GORETZKO, David; KIM, Eunsook; STERNER, Philipp (Ruhr University Bochum)

**Presenters:** GORETZKO, David; KIM, Eunsook; STERNER, Philipp (Ruhr University Bochum)

**Session Classification:** Symposium : “Measurement and Machine Learning”