Multiple imputation of missing data in longitudinal analyses with many variables

Simon Grund[1], Oliver Lüdtke[2,3] & Alexander Robitzsch[2,3]

[1] *University of Hamburg, Germany*

[2] *Leibniz Institute for Science and Mathematics Education, Kiel, Germany*

[3] *Centre for International Student Assessment, Munich, Germany*

Longitudinal data are characterized by a clustered structure, in which observations from multiple time points are nested within participants. In practice, longitudinal data are often incomplete, for example, when participants drop out of the study or fail to provide responses to certain questions or at certain time points. In the statistical literature, multiple imputation (MI) has been shown to provide an effective treatment of missing values in clustered data. However, one crucial requirement of MI is that the imputation model fits the data structure and the intended analyses. In longitudinal data, fulfilling this requirement can be particularly challenging, because the data often contain a large number of variables (e.g., multiple constructs measured at multiple time points) that can exhibit complex patterns of dependency.

In the present talk, we consider different strategies for conducting MI in longitudinal data with many variables and time points. To this end, we first provide an overview of the different strategies, which differ in their representation of the data and their specification of the imputation model. Then, we present the results of a simulation study, in which we evaluated the performance of the different strategies in longitudinal data with fixed measurement occasions and with different complexity and number of variables. Finally, and based on these findings, we discuss the strengths and weaknesses of each strategy and provide recommendations for practice.