expose

Leonie

Max Planck Institute for Human Development

MSB Medical School Berlin

Supervisors:

Prof. Dr. Andreas M. Brandmaier

Prof. Dr. Martin Schultze

expose

Science fundamentally depends on the clear delineation and accurate measurement of the object of inquiry. If measurement does not correspond to the phenomena of interest, no level of methodological rigor, statistical sophistication will avert misleading conclusions (Jessica Kay Flake and Fried 2020; Jessica K. Flake, Pek, and Hehman 2017). Recent research prompted an emphasis on strengthening statistical methods (Jessica K. Flake, Pek, and Hehman 2017; Haig 2020; Altoè et al. 2020) as well as theoretical frameworks in psychology (Eronen and Bringmann 2021), hence it becomes apparent that this development by extent hinges on the evaluation of measurement quality.

Thus, my research will focus on two directions and will be structured into two major projects. The first project will focus on facilitating valid psychometric assessments, whereas the second project will aim at improving the evaluation of reliability of measurement scales. Together, these projects will contribute to advance the evaluation of measurement quality.

Project 1

Ensuring the validity of psychological assessments is crucial, yet differential item functioning (DIF) can threaten measurement invariance (MI) (Bauer, Belzak, and Cole 2020). Recent calls for improved DIF detection methods emphasize the need for more advanced statistical approaches (Lee, Han, and Choi 2024). Moderated nonlinear factor analysis (MNLFA) is a recent approach for assessing MI via parameter moderation within a single-group confirmatory factor analysis framework. MLNFA evaluates MI across multiple continuous and categorical covariates, and accounts for heteroskedasticity by modeling factor and residual variances as functions of these covariates. While MNLFA offers continuous moderation of several parameters of Structural Equation Models (SEMs) (e.g.: factor loadings, covariances etc.), it requires a priori specification of covariates and their functional relationships (Bauer 2017; Kolbe et al. 2024). In contrast, Structural Equation Model (SEM) trees and forests are data-driven, non-parametric methods that use recursive partitioning to

identify latent subgroups in which model parameters differ, without assuming specific functional forms or predefined covariate effects. These approaches allow for nonlinear moderation of factor loadings and can reveal complex interaction effects, enabling the exploratory detection of DIF (A. M. Brandmaier et al. 2016, 2013). In this study, we conduct a Monte Carlo simulation to compare the performance of MNLFA and SEM trees as well as forests in detecting DIF and assessing MI under varying conditions. Specifically, we evaluate their effectiveness in identifying non-invariance and detecting relevant covariates.

Methods

We will preregister the simulation study according to the approach by Siepe et al. (2024). Currently we finished a trial run with 100 replications and a minimal set-up of the simulation, for approximating the needed runtime and computational capacities, as well as the maximum number of conditions that can be included in the main simulation.

Project 1.2

Historically, regularization in SEM was used mainly indirectly for parsimony and robustness, but more recent work applies it directly to the detection of measurement bias and measurement non-invariance. Most existing regularized SEM applications still focus on model parsimony or robust estimation, although several newer approaches explicitly target the identification of non-invariant or biased parameters. Confirmatory MI workflows typically require manual specification of equality constraints, whereas exploratory methods such as SEM trees search for subgroups characterized by different parameter values. Regularization has recently been used to detect measurement bias, but fully standardized and widely adopted workflows that systematically bridge confirmatory MI procedures with exploratory subgroup discovery remain rare. We suggest to use   regularization not only for general sparsity but specifically to select a limited subset of parameters that are permitted to be non-invariant. In multi-group SEM, non-invariance can be expressed as violations of cross-group equality constraints, often parameterized as deviations of group-specific parameters from those of a designated reference group. In MNLFA, non-invariance appears

as moderation effects on intercepts, loadings, or residual variances.   penalization applied to MI assessment would directly limit the number of parameters allowed to differ across groups or time. This aligns with the modeling goal of treating non-invariance as sparse, even though empirical research frequently documents more extensive violations. Together, these considerations provide a conceptual foundation for applying   -based methods to MI detection, at a time when practical implementations within SEM remain limited and require further methodological development.

Project 2

Furthermore, the reliability of psychological measurement instruments is commonly estimated using coefficients such as Cronbach's Alpha and McDonald's Omega. However, both coefficients share one limitation: they rely on the assumption of uncorrelated measurement errors.

This assumption of uncorrelated measurement errors is frequently violated in psychological research. Correlated errors may arise in a range of scenarios, for example: Multi-method assessments, item redundancy, or context-dependent instruments. When correlated measurement errors are present, both Alpha and Omega tend to underestimate the true reliability of the instrument, as the unique contributions of error variance are overestimated. To address this issue, we propose an extension of current reliability modeling that explicitly accounts for correlated measurement errors within a SEM framework. By allowing for correlated residuals, our approach should yield more accurate reliability estimates, particularly for instruments used in multidimensional, mixed-method, or ecologically valid settings.

The reliability coefficient will be based on previous work on the effective error of a latent variable (Andreas M. Brandmaier et al. 2015; Andreas M. Brandmaier, Oertzen, et al. 2018; Andreas M. Brandmaier, Wenger, et al. 2018; A. M. Brandmaier 2025). Using this approach, we introduce a generalized reliability estimator that generalizes coefficients α and

ω to factor models with correlated errors. Based on Classical Test Theory, we define coefficient ξ as:

$$\xi = \frac{1}{1 + \sigma^2_{\text{eff}}}$$

with σ^2_{eff} being the effective error of a pre-specified measurement model, which may or may not contain correlated measurement errors. For tau-equivalent models, ξ is identical to Cronbach's Alpha, for congeneric models, ξ is identical to McDonald's Omega. Using prior work (**Brandmaier2018**), we can compute the effective error using an approximation via the non-centrality parameter of a chi-square distribution, which in turn can be computed from the loglikelihood ratio of a given model and the same model with the true-score variance set to zero.

Application

We will demonstrate the applicability of this SEM based approach using simulated data and empirical examples (e.g., using two-method measurement (2MM) and 3MM design models by Lawes, Schultze, and Eid (2020)), highlighting the potential of this extended reliability coefficient to improve psychometric evaluation and instrument validation in psychological research. Afterwards there will also be need to derive a Eigenwert-based solution of σ^2_{eff}.

Timeline and Feasibility

- **Year 1**: completion of Project 1 (preregistration & submission of first manuscript).

- **Year 2**: Project 2 (literature review, Simulation & empirical demonstration & submission of second manuscript).

- **Year 3**: Buffer for Project completion 1 or 2 (possibly Project 3) & consolidation, defense. (Derive an exact solution of ξ as suggested by Project 2)

Altoè, Gianmarco, Giulia Bertoldo, Claudio Zandonella Callegher, Enrico Toffalini, Antonio Calcagnì, Livio Finos, and Massimiliano Pastore. 2020. "Enhancing Statistical Inference in Psychological Research via Prospective and Retrospective Design Analysis." *Frontiers in Psychology.* https://doi.org/10.3389/fpsyg.2019.02893.

Bauer, D. J. 2017. "A More General Model for Testing Measurement Invariance and Differential Item Functioning." *Psychological Methods* 22 (3): 507–26. https://doi.org/10.1037/met0000077.

Bauer, D. J., W. C. M. Belzak, and V. T. Cole. 2020. "Simplifying the Assessment of Measurement Invariance over Multiple Background Variables: Using Regularized Moderated Nonlinear Factor Analysis to Detect Differential Item Functioning." *Structural Equation Modeling: A Multidisciplinary Journal* 27 (1): 43–55. https://doi.org/10.1080/10705511.2019.1642754.

Brandmaier, A. M. 2025. "Stop Calling Cohen's d an Effect Size." PsyArXiv. https://doi.org/10.31234/osf.io/n9ta7_v1.

Brandmaier, A. M., T. von Oertzen, J. J. McArdle, and U. Lindenberger. 2013. "Structural Equation Model Trees." *Psychological Methods* 18 (1): 71–86. https://doi.org/10.1037/a0030001.

Brandmaier, A. M., J. J. Prindle, J. J. McArdle, and U. Lindenberger. 2016. "Theory-Guided Exploration with Structural Equation Model Forests." *Psychological Methods* 21 (4): 566–82. https://doi.org/10.1037/met0000090.

Brandmaier, Andreas M., Timo von Oertzen, Paolo Ghisletta, Christopher Hertzog, and Ulman Lindenberger. 2015. "LIFESPAN: A Tool for the Computer-Aided Design of Longitudinal Studies." *Frontiers in Psychology* 6: 272. https://doi.org/10.3389/fpsyg.2015.00272.

Brandmaier, Andreas M., Timo von Oertzen, Paolo Ghisletta, Ulman Lindenberger, and Christopher Hertzog. 2018. "Precision, Reliability, and Effect Size of Slope Variance in Latent Growth Curve Models: Implications for Statistical Power Analysis." *Frontiers in*

Psychology 9: 294. https://doi.org/10.3389/fpsyg.2018.00294.

Brandmaier, Andreas M., Elisabeth Wenger, Nils C. Bodammer, Simone Kühn, Naftali Raz, and Ulman Lindenberger. 2018. "Assessing Reliability in Neuroimaging Research Through Intra-Class Effect Decomposition (ICED)." *eLife* 7: e35718. https://doi.org/10.7554/eLife.35718.

Eronen, Markus I., and Laura F. Bringmann. 2021. "The Theory Crisis in Psychology: How to Move Forward." *Perspectives on Psychological Science* 16 (4): 779–88. https://doi.org/10.1177/1745691620970586.

Flake, Jessica Kay, and Eiko I. Fried. 2020. "Measurement Schmeasurement: Questionable Measurement Practices and How to Avoid Them." *Advances in Methods and Practices in Psychological Science* 3 (4): 456–65. https://doi.org/10.1177/2515245920952393.

Flake, Jessica K., Jolynn Pek, and Eric Hehman. 2017. "Construct Validation in Social and Personality Research: Current Practice and Recommendations." *Social Psychological and Personality Science* 8 (4): 370–78. https://doi.org/10.1177/1948550617693063.

Haig, Brian D. 2020. "What Can Psychology's Statistics Reformers Learn from the Error-Statistical Perspective?" *Methods in Psychology* 2: 100020. https://doi.org/https://doi.org/10.1016/j.metip.2020.100020.

Kolbe, L., D. Molenaar, S. Jak, and T. D. Jorgensen. 2024. "Assessing Measurement Invariance with Moderated Nonlinear Factor Analysis Using the r Package OpenMx." *Psychological Methods* 29 (2): 388–406. https://doi.org/10.1037/met0000501.

Lawes, M., M. Schultze, and M. Eid. 2020. "Making the Most of Your Research Budget: Efficiency of a Three-Method Measurement Design With Planned Missing Data." *Assessment* 27 (5): 903–20. https://doi.org/10.1177/1073191118798050.

Lee, S., S. Han, and S. W. Choi. 2024. "A Bayesian Moderated Nonlinear Factor Analysis Approach for DIF Detection Under Violation of the Equal Variance Assumption." *Journal of Educational Measurement* 61: 303–24. https://doi.org/10.1111/jedm.12388.

Siepe, Björn S., František Bartoš, Tim P. Morris, Anne-Laure Boulesteix, Daniel W. Heck,

and Samuel Pawel. 2024. "Simulation Studies for Methodological Research in Psychology: A Standardized Structure for Planning, Preregistration, and Reporting." *Psychological Methods.* https://doi.org/10.1037/met0000695.