

Programme prior to the conference -Thursday 12 December

10:00 - 11:00 IOPS Board meeting (room B.5004)

Programme Thursday 12 December (Tinbergen room)

10:30 - 11:00 **Receipt with coffee and tea**

11:15 - 11:30 **Official opening** by Joris Mulder and welcome by Reinoud Stoel (organizer)

11:30 - 12:00 **Valedictory address by Rob Meijer** (University of Groningen)
10 years of IOPS and a sketch of the development of my own interests

12:00 - 12:30 **Presentation Ilse Peringa** (University of Groningen)
Bridging Test Validity and Practical Application in Decision-Making: Insights from the Lens Model
Discussants: Sanne Peereboom & Ulrich Lösener

12:30 - 13:30 **Lunch**

13:30 - 14:00 **Presentation Daniëlle Remmerswaal** (Utrecht University)
How many diary days? On the trade-off between the quantity of collected data and response burden in the context of smart surveys
Discussants: Marie Stadel & Ilse Peringa

14:00 - 14:30 **Presentation Andres Felipe Perez Alonso** (Tilburg University)
Mixture Multigroup Structural Equation Modeling: Comparing structural relations across many groups
Discussants: Martijn Schoenmakers, Santiago Gómez-Echeverry, Zeynep Bilici

14:30 - 14:45 **Break**

14:45 - 15:15 **Presentation Marie Stadel** (University of Groningen)
Combining Personal Social Networks and Experience Sampling Methodology to Capture the Social Life of Psychotherapy Patients
Discussants: Jill de Ron & Jordan Revol

15:15 - 15:45 **Presentation Zeynep Bilici** (University of Amsterdam)
Evaluating robust variance estimation (RVE) in MASEM
Discussants: Alfons Edmar & Franziska Ruffer

15:45 - 16:30 **Poster presentations**

Edita Chvojka (Utrecht University)

We should change how we think about fit indices in SEM

Cas Goos (Tilburg University)

Assessing Reliable and Valid Measurement as a Prerequisite for Informative Replications in Psychology

Annie Johansson (University of Amsterdam)

Problem Skipping Limits the Accuracy of Ability Estimates in Online Learning

Angelina Kuchina (Tilburg University)

Added value of subscores: Can we accurately evaluate it?

Yuqi Liu (Leiden University)

A Two-step estimator for growth mixture models with covariates in the presence of direct effects

Meike Waaijers (University of Amsterdam)

Can LLMs Support Researchers in Creating Candidate Theories?

Yufei Wu (KU Leuven)

Robust Amortized Bayesian Inference of Cognitive Model Parameters

16:30 - 17:30 **Speed dating with CBS methodologists & drinks**

19:00 - 22:00 **Conference dinner**

Restaurant Bier Kluis (Plein 20, Den Haag)

Programme Friday 13 December (Tinbergen Room)

09:30 - 10:00 **Receipt with coffee and tea**

10:00 - 11:00 **Keynote presentation by Edwin de Jonge** (CBS)

A person network of The Netherlands: Methodological challenges and practical applications

11:00 - 11:30 **Presentation Jordan Revol** (KU Leuven)

Preprocessing ESM data: a step-by-step framework, tutorial website, R package, and reporting templates

Discussants: Jill de Ron & Camila Natalia Barragan Ibanez

11:30 - 12:00 **Presentation Sanne Peereboom** (Tilburg University)

Cognitive phantoms in LLMs through the lens of latent variables

Discussants: Bunga Citra Pratiwi & Martijn Schoenmakers

12:00 - 13:15 **Lunch**

12:45 - 13:15 **PhD meeting** (Tinbergen room)

13:15 - 13:45 **Presentation Santiago Gómez-Echeverry** (VU Amsterdam)

Sensitivity of Selection Estimators: A Diagnostic based on a Simulation and a Case Study

Discussants: Jordan Revol & Daniëlle Remmerswaal

13:45 - 14:15 **Presentation Lukas Nowicki** (KU Leuven)

Evaluating the Efficacy of Mixture Multigroup Factor Analysis in Handling Non-Normal and Ordinal Data: A Simulation Study

Discussants: Edita Chvojka, Lennert Groot, Karel Veldkamp

14:15 - 14:45 **Presentation Lennert Groot** (University of Amsterdam)

In Between Methods: Evaluating Approaches for Individual Participant Data Meta-Analytic Structural Equation Modeling

Discussants: Dennis Peng & Franziska Ruffer

14:45 - 15:00 **Break**

15:00 - 15:30 **Closing and Best Poster/Presentation Awards 2024**

Valedictory address

10 years of IOPS and a sketch of the development of my own interests

Rob Meijer (Groningen University)

In this talk I first provide an overview of developments within IOPS during last 10 years. Then I discuss - in my view - an underrepresented area of psychological testing: How, in practice, test scores are used to make decisions and how difficult it is to improve test use due to human nature.

Keynote presentation

A person network of The Netherlands: Methodological challenges and practical applications

Edwin de Jonge (CBS)

Official Statistics produces many statistics on society, economy, environment, and other interesting fields. All these domains are complex phenomena, e.g. society exists by the virtue of the interactions of the inhabitants of a country, it forms its fabric. Economy exist by the virtue of its economic actors.

Most official statistical output deals with statistics of the inhabitants or economic actors but not so much on the fabric of society itself. To produce statistics on our society, Statistics Netherlands derived a network of formal relations for the whole population of the Netherlands. I will introduce the network, its derivation, its use on calculating individual segregation scores and its potential for further research. The presentation will end with a discussion on the opportunities and pitfalls of using the population network dataset for academic research.

Oral Presentation

Bridging Test Validity and Practical Application in Decision-Making: Insights from the Lens Model

Ilse Peringa (University of Groningen)

The development and validation of psychological tests often focus on metrics such as reliability, factor analysis, and predictive validity. While these elements are foundational, they represent only part of the equation in improving decision-making. The true impact of a test lies in how its results are used in practice, often diverging from optimal weighting strategies assumed in research. Drawing on Egon Brunswik's Lens Model, this presentation explores the gap between theoretical validity and practical application, with a specific focus on applicant assessment and selection. Using examples from various applied domains, it is demonstrated how the integration of test information by human judges can lead to diminished validity, primarily due to human judgment noise.

Discussants: Sanne Peereboom & Ulrich Lösener

How many diary days? On the trade-off between the quantity of collected data and response burden in the context of smart surveys

Daniëlle Remmerswaal (Utrecht University)

Diary studies are used to capture behaviour, such as time use, travel, expenditure, physical activity, and medical symptoms. Traditional self-reported diary studies, however, often impose a high response burden on respondents and are prone to measurement errors. In contrast, smart surveys make use of sensors collected on a device (e.g. geolocations on a smartphone) to compile a daily diary. These sensor-based data collection methods offer advantages, including reduced measurement errors and lower response burden. The reduced burden may also enable data collection over extended periods, provided that each additional day of data offers new information. To evaluate this potential, we separate the variance between and the variance within respondents using multilevel models. We compare the inter- and intrapersonal day-to-day variability, and the reliability, across three different seven-day diary surveys. Based on these results, and considerations of response burden, we offer recommendations for researchers on the optimal duration of diary studies.

Discussants: Marie Stadel & Ilse Peringa

Mixture Multigroup Structural Equation Modeling: Comparing structural relations across many groups

Andres Felipe Perez Alonso (Tilburg University)

Behavioral scientists often examine the relations between two or more latent variables (LV; e.g., how emotions relate to life satisfaction), and Structural Equation Modeling (SEM) is the state-of-the-art for doing so. When comparing these 'structural relations' among many groups, it is likely that some groups share the same relations, so clusters of groups emerge. Before finding these clusters, we must remember that LVs are measured indirectly by questionnaires. Thus, to validly compare the relations among groups, their measurement should be invariant across the groups (i.e., measurement invariance). We propose Mixture Multigroup SEM (MMG-SEM) as a novel solution to gather groups with equivalent structural relations in clusters while accounting for the reality of measurement (non-)invariance. In this presentation, MMG-SEM's definition and estimation will be described.

Discussants: Martijn Schoenmakers, Zeynep Bilici, Santiago Gómez-Echeverry

Combining Personal Social Networks and Experience Sampling Methodology to Capture the Social Life of Psychotherapy Patients

Marie Stadel (University of Groningen)

In recent years, experience sampling methodology (ESM) as well as personalised feedback based on the collected ESM data have become popular in clinical psychology research and practice. Current approaches are, however, limited in capturing an important multifaceted factor for psychotherapy, the social context of a patient. Important social dynamics unfold in everyday life: social interactions, such as having a fight or meeting for dinner with friends, affect the mood and symptoms of a patient. Additionally, these social interactions aggregate into different social relationships which form the personal social (support) network of the patient. Thus, there are two intertwined dynamic levels of social context which cannot be captured with ESM alone. During this talk, I will present a series of studies during which I developed and tested a combination of ESM with personal social network (PSN) data collection to fill this gap.

Discussans: Jill de Ron & Jordan Revol

Evaluating robust variance estimation (RVE) in MASEM

Zeynep Bilici (University of Amsterdam)

Dependent effect sizes in meta-analysis are quite common; studies may measure the same constructs across different time points, using different operationalization strategies or by using multiple informants. Whereas traditional meta-analysis can deal with these dependencies more easily, when researchers are trying to meta-analyze multiple relationships in a SEM model the dependencies are more complex to deal with. In the context of MASEM, when we have multiple effect sizes available for the same relationship in the same study, some of the methods used in the context of traditional meta-analysis is still applicable, such as aggregation, elimination and ignoring dependency. Previous simulation results comparing the methods of aggregation, elimination, ignoring dependency and univariate three-level modeling (Wilson et al., 2016) in the context of MASEM showed that there is not one method that performs well across different conditions and evaluation criteria. Robust variance estimation (RVE) suggests an alternative approach, whereby the covariances in sampling errors are estimated by averaging the cross-products of residuals within each study (Hedges et al., 2010). By integrating a SEM model in multivariate meta-analysis with robust variance estimation, we aim to assess the problem of dependent correlations in MASEM. This simulation study assesses the performance of RVE across conditions of varying number of studies, number of dependent effect sizes within studies, the magnitude of the correlation between the dependent effect sizes and the between studies variance.

Discussants: Alfons Edmar & Franziska Ruffer

Preprocessing ESM data: a step-by-step framework, tutorial website, R package, and reporting templates

Jordan Revol (KU Leuven)

Experience Sampling Method (ESM) studies have become a very popular tool to gain insight into the dynamics of psychological processes. Whereas the statistical modeling of ESM data has been widely studied, the preprocessing steps that precede such modeling have received relatively limited attention, despite being a challenging phase. At the same time, adequate preprocessing of ESM data is crucial: it provides valuable information about the quality of the data and, importantly, helps to resolve issues in the data that may compromise the validity of statistical analyses. To support researchers in properly preprocessing ESM data, we have developed a step-by-step framework, a tutorial website that provides a gallery of R code, an R package, and templates to report the preprocessing steps. Particular attention is given to three different aspects in preprocessing: checking adherence to the study design (e.g., were the momentary questionnaires delivered according to the sampling scheme), examining participants' response behaviors (e.g., compliance, careless responding), and describing and visualizing the data (e.g., examining distributions of variables).

Discussants: Jill de Ron & Camila Natalia Barragan Ibanez

Cognitive phantoms in LLMs through the lens of latent variables

Sanne Peereboom (Tilburg University)

CognitLarge language models (LLMs) increasingly reach real-world applications, necessitating a better understanding of their behaviour. Their size and complexity complicate traditional assessment methods, causing the emergence of alternative approaches inspired by the field of psychology. Recent studies administering psychometric questionnaires to LLMs report human-like traits in LLMs, potentially influencing LLM behaviour. However, this approach suffers from a validity problem: it presupposes that these traits exist in LLMs and that they are measurable with tools designed for humans. Typical procedures rarely acknowledge the validity problem in LLMs, comparing and interpreting average LLM scores. This study investigates this problem by comparing latent structures of personality between humans and three LLMs using two validated personality questionnaires. Findings suggest that questionnaires designed for humans do not validly measure similar constructs in LLMs, and that these constructs may not exist in LLMs at all, highlighting the need for psychometric analyses of LLM responses to avoid chasing cognitive phantoms.

Discussants: Bunga Citra Pratiwi & Martijn Schoenmakers

Sensitivity of Selection Estimators: A Diagnostic based on a Simulation and a Case Study

Santiago Gómez-Echeverry (VU Amsterdam)

With the advent of Big Data and the growing usage of administrative data in social science research, there is an increasing interest in using non-probabilistic samples in official statistics and substantive research. This development has led to a surge of measures to capture the degree of systematic error due to units' selection into the sample, commonly termed selection bias. Whether as individual-level variables or aggregated variable characteristics, population information should be available to estimate the selection bias measures currently available. However, this is a noticeably strict data requirement, and it is still unclear how these estimates would perform with non-normal variables at different levels of selectivity. We addressed this want in the present study by presenting some of the most notable selection bias frameworks in survey statistics and building a series of estimators based on them while considering the availability of historical information, auxiliary information, and a combination of both sources. We performed a simulation study with a series of sensitivity analyses to test the performance of the proposed estimators under different distributional forms, selectivity levels, and with different correlational levels between the auxiliary and the target variable. In addition, We tested the estimators on a case study using register data from Statistics Netherlands. Our results indicate that most approaches are robust to deviations from normality on the target variable as long as there is low selectivity and highly informative auxiliary information. Furthermore, the results suggest that to obtain consistent estimates of the selection error, using a combination of historical and auxiliary information is preferred over using only one of these data sources. All the selection bias estimators presented can be easily implemented and incorporated in official statistics assessment and production.

Discussants: Jordan Revol & Daniëlle Remmerswaal

Evaluating the Efficacy of Mixture Multigroup Factor Analysis in Handling Non-Normal and Ordinal Data: A Simulation Study

Lukas Nowicki (KU Leuven)

In the social sciences, a common research objective is the comparison of latent variables among different groups, such as in cross-cultural studies. Accurate interpretation of these comparisons requires measurement invariance (MI), which implies that constructs are measured consistently across populations. When dealing with multiple groups, MI often does not hold, requiring pairwise comparisons between the groups to identify the sources of non-invariance. However, such comparisons can become impractical when dealing with many groups.

Mixture multigroup factor analysis (MMG-FA) offers a novel approach to this issue by clustering groups based on their measurement parameters. This method captures between-group differences and similarities in measurement parameters without requiring extensive pairwise comparisons. However, the use of maximum likelihood estimation in MMG-FA assumes continuous items and underlying multivariate normality—assumptions that are not always tenable in real-life settings. Consequently, we investigate MMG-FA's performance with ordinal data and underlying non-normal distributions when clustering based on factor loadings, examining the robustness of this model if the previously mentioned assumptions are violated.

Using simulations, we analyze various conditions, including different numbers of clusters, group sizes, ordinal categories, and levels of skewness. We employ metrics such as the adjusted Rand index, RMSE, mean bias error, and Tucker's index of factor congruence to evaluate clustering performance, parameter recovery, and cluster enumeration. To assess model fit, we use the Bayesian Information Criterion. Our findings aim to highlight the limitations and applicability of MMG-FA, suggesting potential improvements for handling real-world data complexities in a further stage of our research.

Discussants: Edita Chvojka, Lennert Groot, Karel Veldkamp

In Between Methods: Evaluating Approaches for Individual Participant Data Meta-Analytic Structural Equation Modeling

Lennert Groot (University of Amsterdam)

Researchers conducting meta-analytical structural equation modeling (MASEM) using raw data have several analysis options to choose from. Cluster-robust estimation, two-level structural equation modeling (SEM), multivariate meta-analysis of path coefficients, and One-Stage MASEM (OSMASEM) are some of these options. Two-level SEM explicitly separates effects at the within-study level from the between-study level, and OSMASEM look at the within-study effects, while cluster-robust method estimates an overall path coefficient, which essentially is a mix of within-study and between-study effects. Of these approaches, cluster-robust estimation is often used in practice, even when the research question involves the within-study level. A comparison of these methods using real-world data, however, shows that cluster-robust estimates deviate from results of other methods. Simulations using a factor model have shown that cluster-robust estimation may not always be free of bias. This study evaluates bias in parameter estimates and standard errors of MASEM methods with raw data in the context of path analysis, using simulated data. We varied equality of variance-covariance structure over the within-study and between-study level, intraclass correlations, number of primary studies being meta-analyzed, and missing data. Results show that cluster-robust estimation method yields significant bias in estimated path coefficients in certain conditions.

Discussants: Dennis Peng & Franziska Ruffer