UCONN

NEAG SCHOOL OF EDUCATION



Modern Modeling Methods Conference

June 26-28th, 2023

University of Connecticut

2023 Modern Modeling Methods Conference

UCONN
NEAG SCHOOL OF EDUCATION

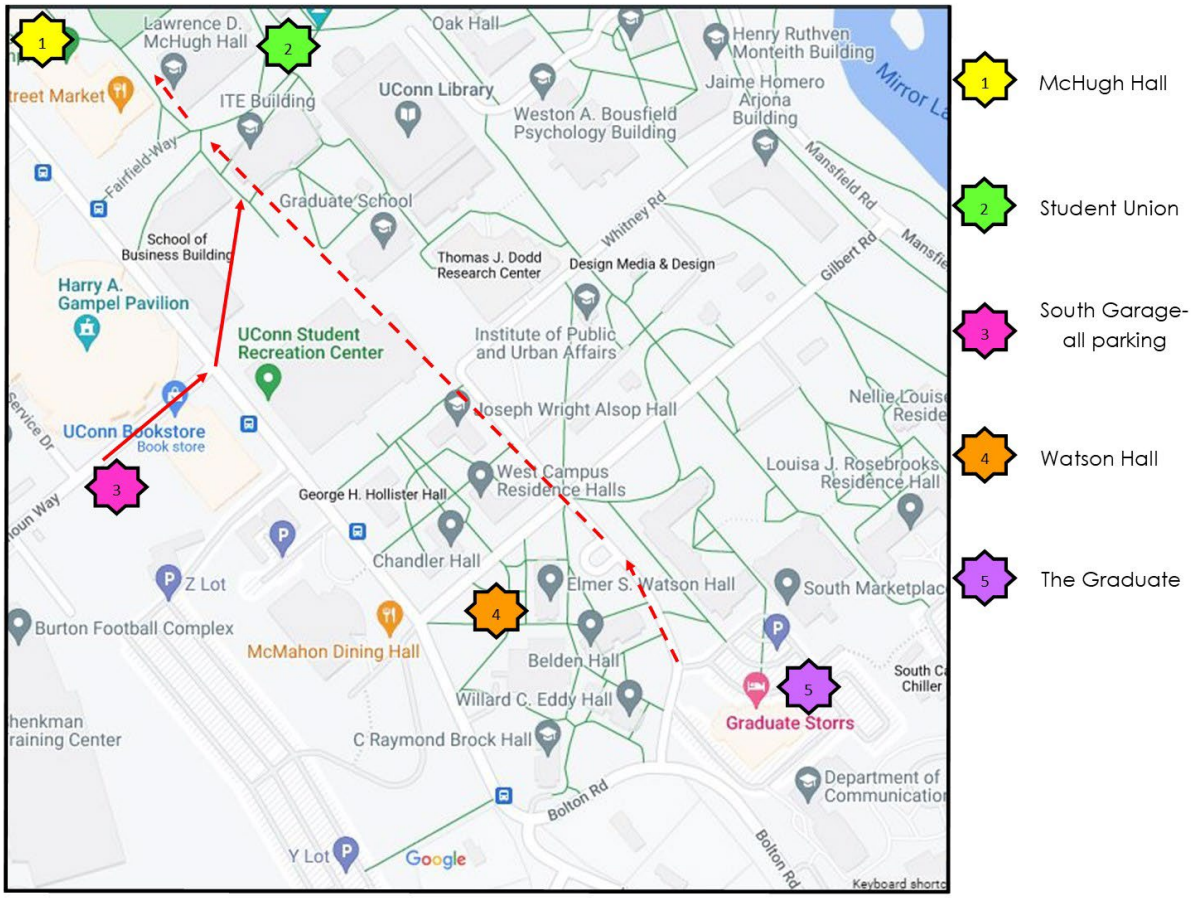
Welcome and thank you for joining us for the 9th Modern Modeling Methods Conference at the University of Connecticut. We are back after a lengthy Covid-induced hiatus, and I have been looking forward to this week all year long. Special thanks to Bengt Muthen, Tihomir Asparouhov, Ellen Hamaker, and all the presenters for making this wonderful program possible! I love seeing all the wonderful modeling work that researchers are doing, and I love getting the chance to interact with people whose work I am reading and citing throughout the year. Also, thank you to Robbin Haboian-Demircan and conference services, Anthony Gambino, Katie Gelsomini, and Joselyn Perez for providing administrative and logistical support for the conference. I couldn't keep this conference going without them!

I hope to see you all again at our tenth Modern Modeling Methods Conference June 24-26, 2024. I am still working on the keynotes for next year, but I can share that the first confirmed keynote speaker for next year's conference is ***Denny Borsboom***. Proposals for papers, symposia, and posters are due January 29th, 2024 and can be submitted online at our website: www.modeling.uconn.edu

If you have suggestions for future keynotes workshops or anything else, please be sure to fill out the conference evaluation, located on our website, www.modeling.uconn.edu . We hope you have a great conference – Happy modeling!

D. Betsy McCoach
Professor, Research Methods, Measurement, and Evaluation program
Department of Educational Psychology, Neag School of Education

Modern Modeling Methods Conference Walking Map



1 McHugh Hall

2 Student Union

3 South Garage- all parking

4 Watson Hall

5 The Graduate

Modern Modeling Methods – 2023 Schedule

Monday, June 26th

Pre-Conference Workshop
8:30-5:00
McHugh 101

New Features in Mplus Versions 8.9 and 8.10
Bengt Muthén, Tihomir Asparouhov, & Ellen Hamaker

Continental Breakfast and Registration 8:00 – 8:30 am
McHugh Atrium

Pre-Conference Workshop
8:30-5:00 – McHugh 101

This is your chance to catch up on the latest series of Mplus analysis acronyms: AESEM, ASEM, BSEM, DSEM, ESEM, PSEM, PSEM-ELGM, RDSEM, and RSEM. These techniques offer better ways to do factor analysis, growth modeling, and SEM as well as more practical ways to study measurement invariance and structural differences across groups and time. These techniques also present new ways to analyze intensive longitudinal data (ILD) using multilevel time series analysis with DSEM and Cross-Classified modeling. Novel real-data applications of these techniques will be discussed together with Mplus scripts.

The workshop starts at 8:30 and ends at 5. There will be 4 blocks, each consisting of 1 ½ hours lecturing followed by 15 minutes of Q & A's. There will be no hands-on computing sessions. Some familiarity with Mplus is required. The new applications to be discussed will draw on a combination of ideas in our papers including these:

Asparouhov, T. & Muthén, B. (2022). Residual Structural Equation Models. *Structural Equation Modeling: A Multidisciplinary Journal*, DOI: 10.1080/10705511.2022.2074422

Asparouhov, T. & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(3), 397-438.

Muthén, B. & Asparouhov, T. (2018). Recent methods for the study of measurement invariance with many groups: Alignment and random effects. *Sociological Methods & Research, 47*(4) 637-664.

Asparouhov, T. & Muthén, B. (2022). Multiple group alignment for exploratory and structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal, 30*(2), 169-191. DOI: 10.1080/10705511.2022.2127100

Muthén, B. & Asparouhov, T. (2012). Bayesian SEM: A more flexible representation of substantive theory. *Psychological Methods, 17*, 313-335

Asparouhov, T. & Muthén, B. (2023). Penalized structural equation models. Forthcoming in SEM.

Asparouhov, T., Hamaker, E.L. & Muthen, B. (2018). Dynamic structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(3), 359-388, DOI: 10.1080/10705511.2017.1406803 (Download Mplus analyses)

Asparouhov, T. & Muthén, B. (2020). Comparison of models for the analysis of intensive longitudinal data. *Structural Equation Modeling: A Multidisciplinary Journal, 27*(2) 275-297, DOI: 10.1080/10705511.2019.1626733 (Download scripts)

Hamaker, E.L., Asparouhov, T, & Muthén, B. (2023). Dynamic structural equation modeling as a combination of time series modeling, multilevel modeling, and structural equation modeling. Chapter 31 in *The Handbook of Structural Equation Modeling (2nd edition)*; Rick H. Hoyle (Ed.); Publisher: Guilford Press. (Download supplementary materials).

Tuesday, June 27th

Continental Breakfast and Registration 7:30 – 8:30 am
McHugh Hall Atrium

Concurrent Paper Session 1 Tuesday 8:30 – 10:00 am

Session 1A: Dynamic Fit Index Cutoffs in SEM
Room 202

Paper	Authors
Dynamic Fit Index Cutoffs for Generalizing and Extending Hu & Bentler	Daniel McNeish Melissa Wolf Patrick Manapat

Session 1B: Causal Moderation
Room 305

Paper	Authors
Alternative Specifications for Instrumental Variable Analysis in Structural Equation Modeling: First Steps Toward Latent Analysis of Symmetrically Predicted Endogenous Subgroups	Anthony J. Gambino
Estimating and Interpreting Heterogeneous Treatment Effects in Online Experiments	Mårten Schultzberg
Causal Language in Evaluating Moderation/Interaction Hypotheses	Amanda Kay Montoya

Session 1C: Latent Variable Modeling as a Vehicle toward Diversity, Equity, and Inclusion
Room 307

Symposium	Authors
Illustrating and Enacting a Critical Quantitative Approach to Measurement with MIMIC Models	Matthew A. Diemer Michael B. Frisby Aixa D. Marchand Emanuele Bardelli
Using QuantCrit to Advance an Anti-Racist Developmental Science: Applications to Mixture Modeling	Sara Johnson Sara Suzuki Stacy Morris
Critical Action and Ethnic-Racial Identity: Tools of Racial Resistance at the College Transition	Channing Mathews Myles Durkee Elan Hope
Development of the White Critical Consciousness Index	Michael B. Frisby Matthew A. Diemer

Concurrent Paper Session 1 **Tuesday 8:30 – 10:00 am**

Session 1D: Measurement Invariance
Room 306

Paper	Authors
Pervasive Differential Item Functioning (DIF)	Paul De Boeck William Goette
Examining SEM Trees for Investigating Measurement Invariance Concerning Multiple Violators	Yuanfang Liu Mark H. C. Lai
Methods for Intersectional Measurement Invariance Testing	Dakota Cintron

Session 1E: Applications of Spatial and Longitudinal Models
Room 108

Paper	Authors
Subjective Well-being and Social Isolation in the COVID-19 Pandemic: A 3-Wave Longitudinal Study across One Year	Tingshu Liu Rodica Ioana Damian David Francis
Measuring Self-Regulatory Development from Kindergarten to Fifth Grade: Longitudinal Psychometrics with Alignment Optimization	Emily M. Weiss
Assessing Differences in Medical Debt in CT by Racial/Ethnic Minorities Using Modern Spatial Path Analytic Methods	Emil Coman Samuel Bruder Kelly George Saachi Shah

---

10:00 – 10:15 am: Coffee Break in the McHugh Hall Atrium

Opening Keynote – Bengt Muthén 10:15 – 11:45 am
McHugh Hall 102

Recent Advances in Modeling Short and Long Longitudinal Data
Bengt Muthén

McHugh Hall 102
10:15 – 11:45 am

The first part of this talk considers repeated measures data with a small number of time points (5-10). Recent thoughts on growth modeling and cross-lagged panel models are discussed. The second part of the talk considers intensive longitudinal data with many time points (20-100). Recent analysis issues in multilevel timeseries analysis using dynamic structural equation models (DSEM), residual DSEM (RDSEM), and cross-classified models are discussed.

Lunch **Tuesday 11:45 am – 1:00 pm**

Student Union Ballroom
3rd floor, Student Union

Concurrent Paper Session 2 **Tuesday 1:00 – 2:30 pm**

Session 2A: Penalized Structural Equation Modeling
Room 202

Paper	Authors
Penalized Structural Equation Modeling	Tihomir Asparouhov

Session 2B: Over-dispersion and Extra-zeros in Multilevel Models for Discrete Count Data
Room 108

Paper	Authors
Model Typology and Demonstration for Over-dispersion and Extra-zeros in Multilevel Models for Discrete Count Data	Ann A. O'Connell Nivedita Bhaktha Krisann Stephany Winifred Wilberforce Abena Anyidoho

Session 2C: Bayesian Approaches
Room 305

Paper	Authors
Bayesian Model Averaging: A Conceptual Introduction	Tyler Hicks Graham Rifenbark Jesse Pace
A Holistic Bayesian Approach to Addressing Measurement Reactivity with a Planned Missing Data Design	Mark Himmelstein David V. Budescu
Bayesian Model Evaluation using Marginal Likelihood for Growth Mixture Models	Xingyao Xiao Feng Ji Yihong Cheng

Session 2D: SEM Fundamentals
Room 307

Paper	Authors
Structural Equation Models for Between, Within, and Mixed Factorial Designs	Alexander M. Schoemann Stephen D. Short
Reenergizing Lost Modeling Treasures: The 100+ Years Old Path Analytic Tracing Rule	Emil Coman Sabrina Uva

Concurrent Paper Session 2 **Tuesday 1:00 – 2:30 pm**

Session 2E: Modeling Dynamic Processes I
Room 306

Paper	Authors
Empirical Bayes Derivative Estimates	Pascal R. Deboeck
Modeling Dynamic Processes with Panel Data: An Application of Continuous Time Models to Prevention Research	Pascal R. Deboeck David A. Cole Kristopher J. Preacher Rex Forehand Bruce E. Compas

Session 2F: Longitudinal Applications I
Room 308

Paper	Authors
Longitudinal Prediction of Health Using the Unique Variance in Personal Suffering Assessment Items	R. Noah Padgett Richard Cowden Tyler VanderWeele
Examination of Heterogeneity of Growth Trajectory on Wage with Auxiliary Covariates: Applying BCH Approach in Growth Mixture Modeling	Hawjeng Chiou Wenyu Chiou
Do Executive Function and Effortful Control Co-Develop? Evidence from Latent Growth Curves with Structured Residuals and Multivariate Growth Mixture Models	Emily M. Weiss

---

2:30 – 2:45 pm Break – McHugh Hall Atrium

Concurrent Paper Session 3 **Tuesday 2:45 – 3:45 pm**

Session 3A: Disaggregating Level-Specific Effects in Cross-Classified Multilevel Models
Room 202

Paper	Authors
Disaggregating Level-Specific Effects in Cross-Classified Multilevel Models	Yingchi Guo Jeneesha Dhaliwal Jason D. Rights

Session 3B: Dealing with Missing Data
Room 305

Paper	Authors
The Impact of MNAR Dropout on Estimation of Latent Growth Curve Models with Binary Observed Variables	Jason T. Newsom Brian T. Keller Nicholas A. Smith Mallory R. Kroeck
Estimating Treatment Effects in Partially Clustered Randomized Controlled Trials with Missing Data: Challenges and Solutions	Manshu Yang

Session 3C: Modeling Item-Level Heterogeneous Treatment Effects with Explanatory IRT
Room 307

Paper	Authors
Modeling Item-Level Heterogeneous Treatment Effects with the Explanatory Item Response Model: Leveraging Online Assessments to Pinpoint the Impact of Educational Interventions	Joshua Gilbert James Kim Luke Miratrix

Session 3D: Comparing Latent and Composite Constructs with Nested Equivalence Testing
Room 306

Symposium	Authors
Comparing Latent and Composite Constructs with Nested Equivalence Testing	Danielle Siegel Mijke Rhemtulla

Concurrent Paper Session 3 **Tuesday 2:45 – 3:45 pm**

Session 3E: Latent Class Analysis
Room 108

Paper	Authors
Modeling Careless Responding in Ambulatory Assessment Studies Using Multilevel Latent Class Analysis: Factors Influencing Careless Responding	Kilian Hasselhorn Charlotte Ottenstein Tanja Lischetzke
Using Latent Class Models to Derive Subtypes for the Enneagram Personality Typology	Jay Magidson

Session 3F: Bifactor Measurement Models
Room 308

Paper	Authors
Confirmatory Bifactor Measurement Models: Their Utility in Scale Development, Psychometric Modeling and Research with Latent Variables and/or Multidimensional Constructs	Rafael Ramirez
Investigating the Factor Structure of Sense of Social and Academic Fit Scale: A Multilevel Bifactor Study	Lizzy Wu Gabriella Jiang Nidia Ruedas-Gracia Taiylor Rayford Shiyu Sun

---

3:45 – 4:00 pm Break

Concurrent Paper Session 4 **Tuesday 4:00 – 5:00 pm**

Session 4A: Partially Nested Designs
Room 202

Paper	Authors
Statistical Power for Multisite Partially Nested Regression Discontinuity Designs	Fangxing Bai Benjamin Kelcey
Estimation, Statistical Power in Partially Nested Multisite Clustered-Randomized Trials	Kyle Cox Yanli Xie Benjamin Kelcey

Session 4B: Longitudinal Causal Effect Estimation
Room 305

Paper	Authors
Estimating Longitudinal Causal Effects: A Comparison of Marginal Structural Models, and Structural Equation Modeling Approaches	Jeroen D. Mulder
Asking and Answering Causal Questions Using Longitudinal Data	Rafael Quintana

Session 4C: Characterisation and Identification of Multivariate Latent Manifolds: Analytically Resolving Factor Score Indeterminacy
Room 307

Paper	Authors
Characterisation and Identification of Multivariate Latent Manifolds: Analytically Resolving Factor Score Indeterminacy	Landon Hurley

Session 4D: Latent Variable Models for Location, Shape, and Scale Parameters
Room 306

Paper	Authors
Latent Variable Models for Location, Shape, and Scale Parameters	Camilo Cardenas-Hurtado Irini Moustaki Giampiero Marra Yunxiao Chen

Concurrent Paper Session 4 **Tuesday 4:00 – 5:00 pm**

Session 4E: Interactive, Automated, and Dynamic SEMs for Maximal Productivity
Room 308

Paper	Authors
Interactive, Automated, and Dynamic SEMs for Maximal Productivity	Laura Castro-Schilo

Session 4F: Mixed Effect Location Scale Models: Methodological and Substantive Considerations
Room 108

Paper	Authors
Mixed Effect Location Scale Models: Methodological and Substantive Considerations	Jennifer Richardson D. Betsy McCoach

---

Poster Session and Reception

Student Union Ballroom
5:00 – 7:00 pm
Tuesday, June 27th
Third Floor, Student Union

Wednesday, June 28th

Continental Breakfast
7:30 – 8:00 am
McHugh Hall Atrium

Concurrent Paper Session 5 Wednesday 8:00 – 9:00 am

Session 5A: A Two-Step Robust Estimation Approach for Inferring Within-Person Relations in Longitudinal Design
Room 102

Paper	Authors
A Two-Step Robust Estimation Approach for Inferring Within-Person Relations in Longitudinal Design	Satoshi Usami

Session 5B: Model Fit Issues in Measurement Models
Room 202

Paper	Authors
ESEM, CFA, and Somewhere In-Between: The Effect of Measurement Quality on Model Fit Sensitivity	Tim Konold Elizabeth Sanders
Performance of Model Fit Indices in Confirmatory Factor Analysis on Misspecified Item Response Models with Local Dependence	Jiangqiong Li Dubravka Svetina Valdivia

Session 5C: Causal Inference I
Room 305

Paper	Authors
Synthetic Control Models for Causal Inference in Observational Social and Behavioral Sciences	John M. Felt Zachary Fisher Chad Shenk
A Tutorial on Propensity Score Analysis with Semi-Continuous Treatment	Huibin Zhang Walter L. Leite

Concurrent Paper Session 5 **Wednesday 8:00 – 9:00 am**

Session 5D: Partially/Disparately Nested Structures
Room 306

Paper	Authors
Croon's Estimation of Multilevel Structural Equation Models with Partially Nested Data	Kyle Cox Benjamin Kelcey
Structural After Measurement Estimation of Disparately Nested Structures	Fangxing Bai Ben Kelcey Yanli Xie Kyle Cox

Session 5E: Difference-in-Differences: A Methodological Illustration
Room 107

Paper	Authors
Difference-in-Differences: A Methodological Illustration	Meghan Cain

Session 5F: Designing Against Bias: Identifying and Mitigating Bias in Machine Learning and AI
Room 108

Paper	Authors
Designing Against Bias: Identifying and Mitigating Bias in Machine Learning and AI	David J. Corliss

---

9:00 – 9:15 am Break –McHugh Hall Atrium

Keynote – Ellen Hamaker **Wednesday 9:15 – 10:30 am**

Opportunities and Challenges of Intensive Longitudinal Data

Ellen Hamaker

McHugh Hall 102
9:15 am – 10:30 am

Technological developments like smart phones and activity trackers have made it relatively easy to obtain many repeated measures from large samples of people while they are living their daily life. Measures may include self-report on affect, behaviors, cognitions, and the environment, but also physiological and/or non-intrusive measurements throughout the day. Such intensive longitudinal data offer new opportunities for studying the dynamics of everyday processes, and allow researchers to pose new research questions. However, with these new opportunities also come new challenges: How should we measure a process—e.g., how often and at what rate should we measure it—and what model should we use to analyze the data—e.g., how can we link our model to our research question? In this talk I will discuss opportunities and challenges associated with this exciting new methodology and sketch various ways in which we may move forward.

10:30 – 10:45 am Break –McHugh Hall Atrium

Concurrent Paper Session 6 Wednesday 10:45 am – 12:15 pm

Session 6A: Transparency, Reproducibility, and Replicability of Modeling Methods: Problems and Solutions
Room 102

Symposium	Authors
Transparency, Reproducibility, and Replicability of Modeling Methods: Problems and Solutions	Jessica Kay Flake D. Betsy McCoach Andrea Howard Amanda Montoya

Session 6B: Complex Designs which Stretch the Boundaries of Conventional IRT
Room 108

Symposium	Authors
Scoring the Complexities of Alphabet Knowledge	Jason Anthony Janelle J. Montroy Jeffrey M. Williams
Multinomial Models for Unpacking Linguistically Informative Response Choices: Test Structure and Item Function	Lee Branum-Martin Julie A. Washington Katherine Rhodes
Picturing the Impossible: Pictures, Repetitions, and Item-Set Card Effects in Measuring Vocabulary	Eleanor Fang Yan Lee Branum-Martin
Assessment of Children with Autism Using Eye Gaze and Generalized Additive Logistic Regression	Ryan P. Bowles Emily Lorang Courtney E. Venker Madeline Klotz

Session 6C: Reconsidering Lord's Paradox
Room 202

Paper	Authors
Robust and Pseudo-Robust Solutions to Lord's Paradox	Robert E. Larzelere Hua Lin
Does Group-Mean Centering Always Inflate Type I Error Rates in Multiple Regression Analyses?	Robert E. Larzelere Hua Lin
Dual Centered-ANCOVA with interaction, a Solution for Lord's paradox with Implications for Valid Causal Inferences in Longitudinal Analyses	Hua Lin Robert E. Larzelere

Concurrent Paper Session 6 — Wednesday 10:45 am – 12:15 pm

Session 6D: Modeling Dynamic Processes II
Room 305

Paper	Authors
Consequences of Sampling Frequency on the Estimated Dynamics of AR Processes using Continuous Time Models	Rohit Batra Simran Johal Meng Chen Emilio Ferrer
Summing the up and downs of life: The Bayesian Reservoir Model of Psychological Regulation	Mirinda M. Whitaker Cindy S. Bergeman Pascal R. Deboeck

Session 6E: Dyadic and Group Applications
Room 306

Paper	Authors
Generalizability Theory Applied to Daily Relationship Quality: Substantive and Statistical Directions	Madison Shea Smith Susan C. South
The Dynamics of Opinion Expression During Group Discussion	Joseph A. Bonito Stephen A. Rains
Boosting Powers by Combining Spatial Econometrics with Dyadic Analysis and SEM: Racial/Ethnic Differences in Life Expectancy across the US States	Emil Coman Peter Xiang Chen Sandro Steinbach Adrian-Gabriel Enescu Monica Raileanu Szeles

---

Lunch — Wednesday 12:15 – 1:15 pm

Student Union

Please be sure to return your lunch card!

Please note: Dining cards can be used *only* on Wednesday, June 28th and *only* at the Union Street Marketplace. When you return your dining cards, be sure to hand them to a member of the conference staff so they can cross your name off the list. Please do not just leave it on the registration table—those who return their cards will be entered in a raffle to win a free conference registration for the 2024 conference.

Concurrent Paper Session 7 **Wednesday 1:15 – 2:15 pm**

Session 7A: Tutorial: Methods for Sensitivity Analysis to Omitted Confounders in SEM
Room 102

Paper	Authors
A Tutorial on Methods for Sensitivity Analysis to Omitted Confounders in Structural Equation Modeling	Walter L. Leite Zuchao Shen Charles L. Fisk Eric A. Wright Jeffrey Harring Katerina M. Marcoulides

Session 7B: Issues in Multilevel Modeling
Room 202

Paper	Authors
Modeling Options for Clustered Data: An Empirical Comparison of Hierarchical and Population Average Models	Bethany A. Bell Jason Schoeneberger Anthony A. Mangino
Understanding the Consequences of Collinearity for Multilevel Models: The Importance of Disaggregation across Levels	Haley E. Yaremych Kristopher J. Preacher

Session 7C: Modeling Single Subject Data
Room 305

Paper	Authors
GLMMs for Overdispersed Count Data in SCED Studies: Does Autocorrelation Matter?	Haoran Li Wen Luo Eunkyeng Baek Kwok Hap Lam Wenyi Du Noah Koehler
Individual Participant Data Meta-Analysis Including Moderators: Empirical Validation	Mariola Moeyaert Yukang Xue Panpan Yang

Concurrent Paper Session 7 **Wednesday 1:15 – 2:15 pm**

Session 7D: Enacting Critical Quantitative Methodology: Leveraging IRT to Advance Critical Consciousness Measurement
Room 108

Paper	Authors
Enacting Critical Quantitative Methodology: Leveraging IRT to Advance Critical Consciousness Measurement	Matthew Diemer Michael B. Frisby Andres Pinedo Emanuele Bardelli Elise Wilkerson Sara McAlister

Session 7E: The Construction and Estimation of Multidimensional Linear Factor Models without Parametric Assumptions
Room 306

Paper	Authors
The Construction and Estimation of Multidimensional Linear Factor Models without Parametric Assumptions	Landon Hurley

---

2:15 – 2:30 pm Break – Refreshments in McHugh Atrium

Concurrent Paper Session 8 **Wednesday 2:30 – 4:00 pm**

Session 8A: Dyadic and Group Designs
Room 102

Paper	Authors
Understanding Group Effects Using the Co-Partner Design	David A. Kenny
Sample Considerations for Detecting Person, Dyad, and Contextual Effects Using the Common Fate Model for Dyadic Analysis	Robert E. Wickham

Session 8B: Optimal Design and Cost Considerations
Room 202

Paper	Authors
Power Analysis and Sample Size Planning in the Design of Two-Level Randomized Cost-Effectiveness Trials	Wei Li Nianbo Dong Rebecca Maynard Benjamin Kelcey Jessaca Spybrook Yue Xu
Optimal Design of Experimental Studies Under Condition- and Unit-Specific Cost Structures	Zuchao Shen Benjamin Kelcey
Optimal Design of Multisite-Randomized Trials Investigating Mediation Effects Under Unequal Costs	Zuchao Shen Wei Li Benjamin Kelcey Walter Leite Huibin Zhang

Session 8C: Longitudinal Applications II
Room 108

Paper	Authors
An Application of Random Changepoint Models to Cognitive Aging Research	Zachary J. Kunicki Yi Feng Douglas Tommet Sharon K. Inouye Richard N. Jones
Application of a Novel Model for Analyzing Data from Randomized Pretest, Posttest, Follow-up Designs: Results from a Pediatric Randomized Behavioral Clinical Trial	Constance Mara
Health in All Polices Approach: A Dynamic Modelling of Social Policies' Effect on Mental Health	Ekaterina Melianova

Concurrent Paper Session 8 **Wednesday 2:30 – 4:00 pm**

Session 8D: Sample Heterogeneity in Dynamic Psychological Processes
Room 305

Symposium	Authors
A Bayesian Multilevel Mixture Autoregressive Model	Xingyao Xiao Feng Ji
Identifying and Explaining Sample Heterogeneity in Dynamic Psychological Processes Using ml-VARTree	D. Jody Zhou Emilio Ferrer Siwei Liu
Impact of Temporal Order Selection on Clustering Intensive Longitudinal Data Based on Vector Autoregressive Models	Hairong Song Yaqi Li
Penalized Subgrouping of Heterogeneous Time Series	Christopher Crawford Jonathan Park Sy-Miin Chow Zachary Fisher
Clustering Analysis of Time Series of Affect in Dyadic Interactions	Samuel D. Aragones Emilio Ferrer

Session 8E: Modeling COVID-19
Room 306

Paper	Authors
Unobserved Components Models: Applications in Post-COVID Analysis	David J. Corliss
Using Cross-classified Multi-level Modeling to Identify COVID-19 Period Effects on Race and Gender Differences in Training Effectiveness	Youngmin Kim Bridget McHugh Rebecca Berenbon Abena Anyidoho
Daily Associations of Emotion and Fatigue in College Students during the Early Stages of the COVID-19 Pandemic: An Application of Dynamic Structural Equation Modeling	Parisa Rafiee Elizabeth Pauley Melissa Rothstein Amy L. Stamates Manshu Yang

Paper Abstracts

Concurrent Paper Session 1 Tuesday 8:30 – 10:00 am

Session 1A: Dynamic Fit Index Cutoffs in SEM (Room 202)

Dynamic Fit Index Cutoffs for Generalizing and Extending Hu & Bentler
Daniel McNeish, Melissa Wolf, Patrick Manapat
Abstract: Fit indices like RMSEA and CFI are effect sizes quantifying misspecification magnitude. There are no inherent values that indicate "good" fit, so indices are typically compared to cutoffs suggested by simulations in Hu and Bentler (1999). However, methodological research has shown that these cutoffs do not generalize across models whose characteristics differ from those studied by Hu and Bentler. Specifically, cutoffs change widely with different numbers of items, factors, or strength of loadings; which arbitrarily rewards or punishes some types of models. Prior research has suggested deriving unique cutoffs via customized simulations that substitute the characteristics of the model for Hu and Bentler's conditions. However, this idea has not been widely adopted, presumably because many researchers are not well-versed with simulation methods. In this talk, we discuss the dynamic fit index method, which features an algorithm to design and execute a Hu and Bentler-style simulation based on estimates from the user's model. The main idea is to generalize the logic of Hu and Bentler's simulation across any set of model characteristics such that optimally sensitive cutoffs are derived for any arbitrary model. We describe the foundations of the method, how to implement it with accessible software, and provide simulation results to show how the method can improve fit evaluation relative to Hu and Bentler cutoffs.

Session 1B: Causal Moderation (Room 305)

Alternative Specifications for Instrumental Variable Analysis in Structural Equation Modeling: First Steps Toward Latent Analysis of Symmetrically Predicted Endogenous Subgroups
Anthony J. Gambino
Abstract: Structural equation modeling (SEM) is capable of estimating models with maximum likelihood that function like instrumental variable analyses, which are typically estimated via two-stage least squares (TSLS). This is usually done by including an instrumental variable in the model that is not allowed to cause the outcome of interest but causes an endogenous causal predictor which does cause the outcome of interest (and a correlation between the residuals of the endogenous predictor and the outcome is freely estimated). This research explores an alternative specification possible in SEM that also replicates the traditional TSLS instrumental variable analysis. Subsequently, this alternative specification is explored as a way to perform analysis of symmetrically predicted endogenous subgroups (ASPES) in the SEM framework. Finally, simulation studies are used to evaluate the potential utility and validity of this approach.

Estimating and Interpreting Heterogeneous Treatment Effects in Online Experiments
Mårten Schultzberg
Abstract: At Spotify, we run thousands of randomized experiments every year on hundreds of millions of users. With users across the world, we are expecting to see heterogeneity in the treatment effects. Estimating and interpreting heterogeneous treatment effects at this scale is a challenge, and the literature on this topic is expanding rapidly. This talk presents some of the modeling techniques that we use in the Experimentation Platform at Spotify to learn about heterogeneous treatment effects. I focus on the trade-offs between model complexity and ease of interpretation, and the challenges that arise from working with very large samples.

Causal Language in Evaluating Moderation/Interaction Hypotheses
Amanda Kay Montoya
Abstract: Moderation analysis is ubiquitous throughout many scientific fields. Causal inference has been a major area of focus in mediation analysis, yet is described very little in the context of moderation analysis. A lack of precision in language limits the specificity and generalizability of our theories related to moderation. Previous research has proposed some language to differentiate causal structures (e.g., moderation vs. interaction), but these recommendations have been conflicting, and are not comprehensive or widely adopted. In this talk, I propose a complete nomenclature for differentiating all possible cases for hypothesized causal effects in moderation. In particular, I distinguish whether the focal predictor is hypothesized to have a causal effect on the outcome (effect vs. relationship), and whether the moderator is hypothesized to have a causal effect on the effect/relationship between the focal predictor and the outcome (moderation vs. heterogeneity). I demonstrate using real psychological examples how this distinction can be leveraged to generate clearer theories about the underlying causes of statistical interactions. The hypothesized causal framework has implications for covariate inclusion and experimental design. This work provides a framework from which researchers can describe their theories with greater specificity, allowing for clearer tests of these theories and descriptions of results.

Session 1C: Latent Variable Modeling as a Vehicle toward Diversity, Equity, and Inclusion (Room 307)

Illustrating and Enacting a Critical Quantitative Approach to Measurement with MIMIC Models
Matthew A. Diemer, Michael B. Frisby, Aixa D. Marchand, Emanuele Bardelli
Abstract: Quantitative methodology and the field of measurement have racist, sexist, and eugenicist histories. These histories have led many to abandon quantitative methods, believing that achieving equity is not possible with methods developed to propagate oppression. However, more critical and emerging scholarship has begun to articulate a Critical Quantitative (CQ) perspective. This paper aims to move beyond the critique of quantitative methods and to consolidate this emerging CQ literature, by articulating specific applications and enactments of a (CQ) framework. To do so, this paper articulates five guiding principles (i.e., foundation, goals, parity, subjectivity, and self-reflexivity) and details how CQ can address racism and inequity. MIMIC models are considered as one (underutilized) strategy to identify and mitigate the impacts of racism on measurement, illustrated with empirical examples. Finally, we note that while CQ can identify and mitigate how racism permeates measurement, this is not the same as

directly intervening on the interlocking systems of inequality that perpetuate racism, anti-Blackness, and other forms of oppression that constrain humanity and opportunity.

Using QuantCrit to Advance an Anti-Racist Developmental Science: Applications to Mixture Modeling

Sara Johnson, Sara Suzuki, Stacy Morris

Abstract: How researchers use statistical analyses shapes their research toward or away from an anti-racist agenda. This presentation discusses how social and behavioral scientists can use the QuantCrit framework to critically examine the process of conducting mixture model analyses (e.g., latent class analysis). We first summarize the tenets of QuantCrit and how it has turned the lens of critical race theory onto quantitative methodology. Second, we will briefly explain the key concepts of mixture modeling. The majority of the presentation will focus on applying QuantCrit principles in three "moments" in the mixture model process: (1) development of the research question(s) and identification of analysis variables; (2) decision-making about the role of race in planned analyses; and (3) interpretation of the results through a theoretical framework. We describe each moment, illustrate how researchers can use QuantCrit principles within it, and offer empirical examples. We also discuss how to apply a QuantCrit perspective on extensions to mixture modeling, such as relating latent classes or profiles to distal outcomes.

Critical Action and Ethnic-Racial Identity: Tools of Racial Resistance at the College Transition

Channing Mathews, Myles Durkee, Elan Hope

Abstract: Ethnic-racial identity (ERI) and Critical Consciousnesses (CC) are two developmental processes that are salient for Black youth, particularly as they transition across school contexts. What remains unclear is how these two processes overlap and change over time in Black college students. Using a cross lagged longitudinal model, we highlight the relations between ethnic-racial identity exploration and critical action–two salient dimensions of ERI and CC respectively–across 4 timepoints. Participants were Black students (N = 237; Mage = 18.2, 74% female) from a longitudinal study of college transition. Critical action positively predicted ERI exploration over each year of college, and ERI exploration positively predicted critical action in a reciprocal fashion over the same years. These findings underscore theoretical assertions that critical action and ERI are intertwined in Black youths' development and provide insight into how critical action and ERI overlap beyond adolescence. Specifically, investigating the reciprocal nature of ERI exploration and critical action can help inform university stakeholders of best practices to support the unique experiences of Black students attending PWIs. Such practices could include providing stronger support for Black-oriented spaces and courses that honor the historical and current legacies of Black students enacting systems change.

Development of the White Critical Consciousness Index

Michael B. Frisby, Matthew A. Diemer

Abstract: Critical consciousness (CC) describes the analytic awareness of systemic oppression and the informed actions taken to combat it. It is rooted in the philosophy of scholar and educator Paulo Freire and theorized to be an instrument of liberation. Psychometric instruments have been developed with the expressed aim of measuring CC for more marginalized individuals. Though this is an obvious starting point, we argue that more privileged people should also develop CC, and considerably less attention has been paid to its measurement. Using

an all-white U.S. sample as a baseline for privilege, this study applies factor analysis and item response theory to create the first psychometric instrument measuring CC in white American communities: The White Critical Consciousness Index (WhiCCI). This instrument offers reliable fit statistics and is designed from the perspective of racially privileged individuals to explicitly incorporate various systems of privilege and oppression into its critical reflection and action subdomains.

Session 1D: Measurement Invariance (Room 306)

Pervasive Differential Item Functioning (DIF)
Paul De Boeck, William Goette
Abstract: Most DIF detection methods minimize DIF in favor of group mean differences and would not detect DIF as the major source of group mean differences. We believe that pervasive DIF can go unnoticed and may cause group mean differences. We propose a method to deal with DIF that possibly pervades a large part of the test. For illustrative purposes, we present an application with real data on a neuropsychological test (Boston Naming Test) applied to two groups: Caucasian and African American respondents.

Examining SEM Trees for Investigating Measurement Invariance Concerning Multiple Violators
Yuanfang Liu, Mark H. C. Lai
Abstract: Measurement invariance is needed for accuracy and meaningful interpretations of statistical results. In addition, measurement invariance status can be associated with multiple covariates. This study explored a novel use of the structural equation model (SEM) tree to detect measurement noninvariance concerning multiple covariates under a Monte Carlo simulation. Preliminary results showed that likelihood comparisons under SEM tree had Type I error rates ≤ .052 when n ≤ 1000 and statistical power rates of .964–1.00 in detecting both linear and quadratic nonlinear intercept noninvariance when $n = 1000$. The SEM tree performed well in distinguishing true violators from noise covariates related to intercept noninvariance in terms of split rates ≥ .928, $n = 1000$, especially for a dichotomous violator. The SEM tree can identify covariates associated with parameter estimate heterogeneity, and the invariance testing can be applied to datasets with many covariates to uncover relations between items, covariates, and constructs and facilitates instruments and theory development.

Methods for Intersectional Measurement Invariance Testing
Dakota Cintron
Abstract: Intersectional measurement invariance testing requires evaluating the psychometric properties of a scale across potentially many social and political identities (e.g., gender x race x educational attainment). This paper covers three methods for intersectional measurement invariance testing: 1) the alignment method, 2) mixture multiple group factor analysis, and 3) moderated nonlinear factor analysis. The paper demonstrates the use of the three methods with an empirical analysis of the PHQ-8 depression instrument from the 2019 National Health Interview Survey. I also discuss the pros and cons of each method for intersectional measurement invariance testing.

Session 1E: Applications of Spatial and Longitudinal Models (Room 108)

Subjective Well-being and Social Isolation in the COVID-19 Pandemic: A 3-Wave Longitudinal Study across One Year

Tingshu Liu, Rodica Ioana Damian, David Francis

Abstract: The COVID-19 global pandemic has posed a great challenge to people's physical and mental health, but the extent to which increasing social isolation relates to decrements in subjective well-being (SWB) has been understudied. Prior research addressing this question has been limited to short periods of time or only one or two indicators of SWB, and the findings have been inconsistent. To address these issues, this study ($N = 972$) tracked five SWB indicators (i.e., life satisfaction, positive and negative emotions, depression, and anxiety) over three waves of data that covered one year following the pandemic declaration and included social isolation as a time-varying covariate. With latent growth curve models, all indicators of SWB remained stable during the study period, indicating that most people showed resilience. With multi-level models, social isolation was associated with SWB differently at the between-person and within-person levels. Specifically, more isolated people were emotionally (i.e., lower positive and negative emotions) and cognitively (i.e., lower levels of life satisfaction and depression) disengaged compared to their less isolated counterparts. For the same person, when they reported more isolation, they experienced lower levels of negative emotions and depression than when they reported less isolation.

Measuring Self-Regulatory Development from Kindergarten to Fifth Grade: Longitudinal Psychometrics with Alignment Optimization

Emily M. Weiss

Abstract: Measuring children's self-regulation is of great import to parents, teachers, and administrators who seek to support students' development of crucial school-related skills. However, grade-level changes in classroom structure and behavioral expectations across elementary school necessitate measurement tools that are developmentally appropriate and sensitive to individual growth. This paper presents psychometric analyses of data from the Early Childhood Longitudinal Study, Kindergarten Class of 2010-2011 ([ECLS-K], Tourangeau et al., 2019; analytic $N = 10{,}345$). Teachers reported on children's self-regulatory behaviors in the classroom with two partially overlapping sets of items for children, respectively, in kindergarten and first grade and 2nd-5th grades. First the dimensional structure of classroom self-regulation items was established separately for younger and older children. Then the alignment optimization method (Muthén & Asparouhov, 2014) was used to detect longitudinally-invariant anchor items and produce a developmental scale of self-regulation that applies to students across the whole of elementary school. Results are discussed in light of methodological and practical applications.

Assessing Differences in Medical Debt in CT by Racial/Ethnic Minorities Using Modern Spatial Path Analytic Methods

Emil Coman, Samuel Bruder, Kelly George, Saachi Shah

Abstract: This is an applied demonstration of the perils of analyzing spatial data while ignoring the inherent data similarities due to spatial proximity. We review the 'auto'-correlation (from spatial econometrics) and nonindependence (from social psychology) concepts, and provide intuitive visuals to illustrate it. We then show how to partial out this 'nuisance' (the spatial

similarity) with simple spatial lag models, in which a new variable is created for any outcome, representing the average values of each region's neighbors. A statewide data on small claims medical debt, from CT, USA, aggregated at census tract level ($N = 427$) and higher at state senate district level ($n = 36$), is used to investigate health disparities research questions involving differences in the small claims medical debt rates between regions with more/fewer racial/ethnic minority residents, and how income affects this relation. We investigate to what extent ignoring the spatial structure, the clustered/multilevel structure, or both may lead to a more biased view of the true effects. We provide results from spatial nonrecursive path analytic models, spatial mediation models [% Minority → Income → Debt rate], and spatial multilevel SEM models, with spatial effects between lower-level units and higher-level units.

Concurrent Paper Session 2 Tuesday 1:00 – 2:30 pm

Session 2A: Penalized Structural Equation Modeling (Room 202)

Penalized Structural Equation Modeling
Tihomir Asparouhov
Abstract: Penalized structural equation models (PSEM) is a new powerful maximum-likelihood based estimation technique that can be used to tackle a variety of difficult structural estimation problems that cannot be accomplished with previously developed methods. We describe the PSEM framework and illustrate the quality of the method with examples and simulation studies. We show that traditional EFA models as well as multiple group alignment (MGA) are examples of PSEM. This provides unlimited opportunities to customize and combine EFA and alignment models. The PSEM framework also extends standard SEM models with the possibility to structurally align various model parameters. We also show that this new framework is a maximum-likelihood based alternative to the small-variance-priors BSEM models, where a set of parameters are viewed as approximately zero or approximately equal.

Session 2B: Over-dispersion and Extra-zeros in Multilevel Models for Discrete Count Data (Room 108)

Model Typology and Demonstration for Over-dispersion and Extra-zeros in Multilevel Models for Discrete Count Data
Ann A. O'Connell, Nivedita Bhaktha, Krisann Stephany, Winifred Wilberforce, Abena Anyidoho
Abstract: Generalized linear mixed models (GLMMs) are a class of models designed to accommodate correlated data, such as arise from nested or clustered designs. For discrete count outcomes, the most familiar form of GLMM is the multilevel Poisson. However, Poisson models are based on an assumption of equidispersion, where the variance of observations is equal to their mean. Equidispersion is often violated, with actual count data exhibiting variance greater than the mean (over-dispersion). An excess number of zeros (zero-inflation) is another issue in modeling count data given the assumed underlying distribution for the counts. Although there are many models designed to address these three issues (clustering, over- (or under-) dispersion, and excess zeros), addressing all three and understanding the differences between potential models can be quite challenging. We present a typology and tutorial of GLMMs focusing on multilevel Poisson, Negative Binomial, Generalized Poisson, Zero-Inflation, and

Hurdle models. Model properties and parameter interpretations for each of these are demonstrated through real-world and/or simulated data, with guidance on their selection and application for education researchers. Our presentation uses R, but we also include code and discussion for additional software packages.

Session 2C: Bayesian Approaches (Room 305)

Bayesian Model Averaging: A Conceptual Introduction
Tyler Hicks, Graham Rifenbark, Jesse Pace
Abstract: It is common to use a search algorithm to find the "best" subset of predictors and then assume the correctness of the recovered set. Bayesian model averaging (BMA) offers an alternative solution, which naturally folds uncertainty about the correctness of the recovered set into inferences for a more robust analysis. Although an R package for implementing Bayesian Model Averaging has been available for a few years (Raftery, 1995), this package assumes modelers do not have missing data on predictors. However, as the number of predictors rise, the chance of missing data proportionately increases. Implementation of MBA then is still hindered by the need to give researchers more practical guidance, especially when there is some missing data among predictors. This paper documents strategies for overcoming missing data in the context of BMA analysis, demonstrates the potential benefits of applying this technique with an example involving real data, and reports preliminary simulation results.

A Holistic Bayesian Approach to Addressing Measurement Reactivity with a Planned Missing Data Design
Mark Himmelstein, David V. Budescu
Abstract: Missing data imputation methods are typically aimed at addressing situations where missingness is an incidental, unanticipated, or unwanted nuisance. If data can be assumed to be missing either fully or conditionally at random, many methods are available for imputation without bias, allowing researchers to estimate parameters of theoretical interest. However, one potentially unexplored application of missing data methods is for addressing problems where measurement itself creates a confounding effect, sometimes referred to as measurement reactivity. Measurement reactivity often occurs in pretest-posttest designs, in which researchers want to understand the effect of a treatment or intervention, but the pretest modifies the effect of the intervention on posttest results, limiting generalizability. We present a new application of planned missing data experimental design and a holistic fully Bayesian imputation technique that can allow researchers to omit the pretest, allowing for theoretical inference about treatment effects in the absence of measurement reactivity. We illustrate via simulation and empirical study based on a common advice taking experimental paradigm, where we show it is possible to measure the influence of advice on decision making without requiring people to report their beliefs prior to receiving advice.

Bayesian Model Evaluation using Marginal Likelihood for Growth Mixture Models
Xingyao Xiao, Feng Ji, Yihong Cheng
Abstract: Growth mixture models include two layers of latent variables, namely a discrete (a.k.a., latent class variable) and one or more continuous (a.k.a., random effects) latent variables. In a frequentist setting, the likelihood function has the local maxima problem, and the regularity conditions are violated when using Akaike's Information Criterion (AIC) and Bayesian

Information Criterion (BIC). In a Bayesian setting, the two layers of latent variables result in different versions of likelihood specification: 1) fully conditional on the latent classes and random effects, 2) hybrid conditional on random effects and marginal over latent classes, and 3) marginal over both latent classes and random effects. Consequently, Bayesian model comparison indices, such as Watanabe-Akaike Information Criteria (WAIC) and leave-one-out cross-validation (LOO-CV), based on these different kinds of likelihood imply different prediction intentions, or in the case of the hybrid likelihood, even lead to self-contradiction. Unfortunately, the hybrid likelihood is the default approach in Bayesian software such as Stan that does not allow the sampling of discrete parameters. This paper shows that marginal WAIC and LOO-CV perform well in evaluating a model's predictive accuracy for future subjects' developmental trajectories and in selecting the number of latent classes via a simulation study.

Session 2D: SEM Fundamentals (Room 307)

Structural Equation Models for Between, Within, and Mixed Factorial Designs
Alexander M. Schoemann, Stephen D. Short
Abstract: Factorial designs are a popular experimental design in psychology and behavioral science. The analysis of variance (ANOVA) framework has been the traditional method for examining mean differences in factorial designs, but ANOVA requires several assumptions, e.g., homogeneity of variances, sphericity, measurement invariance, lack of measurement error in the dependent variable, that are minimized when structural equation modeling (SEM) techniques are used. Furthermore, SEM is not limited to testing group means but it can be used to investigate group differences across a range of parameters, e.g., factor loadings, variances, or covariances. We introduce a general technique to analyze between, within, and mixed factorial designs with SEM using multiple group modeling and contrast codes. Contrast codes are used to constrain parameters across between and/or within subjects' factors to allow for tests of main effects and interactions. We demonstrate this technique through a series of simulations and real data examples. Examples include simulation studies testing for factorial invariance and include real world data examples of theory driven tests for differences in latent covariances across between and within subjects' factors.

Reenergizing Lost Modeling Treasures: The 100+ Years Old Path Analytic Tracing Rule
Emil Coman, Sabrina Uva
Abstract: We provide a modern repositioning of Sewall Wright's century old method of path analysis, and recast his original tracing rule as: (1) A foundational statistical tool that provides algebraic-alternative calculations; (2) An investigational lens probing the structure behind mere associations; (3) The backbone of modern causality analytic methods; (4) A stellar application of graph theoretical concepts in statistics. Wright proposed that correlations have their own causes, and are the observable manifestations of underlying causal effects, and that one can derive an association formally from the sum of the product of (standardized) path coefficients along all pathways (or trek, or walk) connecting (causally) two variables (tracing rule). We review three tracing rule's proofs, from Harri Kiivery in 1986, from systems theory (engineering, William Huggins), SEM classics (Les Hayduk), and from basic math (chain derivatives rule) informed by Judea Pearl's review of SEM as a 'linear microscope' for causality. We show how to use the tracing rule to calculate statistical parameters, like regression coefficients, or causal effect from instrumental variable (IV) models. We finally show why path analysis is better suited logically

as the #1 topic in introductory statistics courses, because of its intuitive visual benefits, and broader reach into various statistical models.

Session 2E: Modeling Dynamic Processes I (Room 306)

Empirical Bayes Derivative Estimates
Pascal R. Deboeck
Abstract: A dynamic system is a set of interacting elements characterized by changes occurring over time. The estimation of derivatives is a mainstay for exploring dynamics of constructs, particularly when the dynamics are complicated or unknown. The presence of measurement error in many social science constructs frequently results in poor estimates of derivatives, as even modest proportions of measurement error can compound when estimating derivatives. Given the overlap in the specification of latent differential equation models and latent growth curve models, and the equivalence of latent growth curve models and mixed models under some conditions, derivatives could be estimated from estimates of random effects. This article proposes a new method for estimating derivatives based on calculating the Empirical Bayes estimates of derivatives from a mixed model. Two simulations compare four derivative estimation methods: Generalized Local Linear Approximation, Generalized Orthogonal Derivative Estimates, Functional Data Analysis, and the proposed Empirical Bayes Derivative Estimates. The simulations consider two data collection scenarios: short time series (10 observations) from many individuals or occasions, and long individual time series (25–500 observations). A substantive example visualizing the dynamics of intraindividual positive affect time series is also presented.

Modeling Dynamic Processes with Panel Data: An Application of Continuous Time Models to Prevention Research
Pascal R. Deboeck, David A. Cole, Kristopher J. Preacher, Rex Forehand, Bruce E. Compas
Abstract: Many interventions are characterized by repeated observations on the same individuals (e.g., baseline, mid-intervention, two to three post- intervention observations), which offer the opportunity to consider differences in how individuals vary over time. Effective interventions may not be limited to changing means, but instead may also include changes to how variables affect each other over time. Continuous time models offer the opportunity to specify differing underlying processes for how individuals change from one time to the next, such as whether it is the level or change in a variable that is related to changes in an outcome of interest. After introducing continuous time models, we show how different processes can produce different expected covariance matrices. Thus, models representing differing underlying processes can be compared, even with a relatively small number of repeated observations. A substantive example comparing models that imply different underlying continuous time processes is fit using panel data, with parameters reflecting differences in dynamics between control and intervention groups.

Session 2F: Longitudinal Applications I (Room 308)

Longitudinal Prediction of Health Using the Unique Variance in Personal Suffering Assessment Items

R. Noah Padgett, Richard Cowden, Tyler VanderWeele

Abstract: Suffering is prevalent in all human experience to some degree, and evidence is accumulating that individuals' perceived suffering can be uniquely predictive of future health, especially mental health. In this study, we present results from a cohort study of chronically ill individuals in the U.S. ($N = 1036$) using the Personal Suffering Assessment to predict later time point mental health. After controlling for the common variance across items, we used the item unique variances as predictors of future mental health scores to determine which items provide unique information about mental health status. The results showed that two items on the Personal Suffering Assessment were consistently related to future mental health scores between waves two and three. However, the common information across items was differentially predictive across waves. Our finding implies that jointly using all items on the PSA as a measure of suffering may misrepresent the relationship suffering has with future mental health.

Examination of Heterogeneity of Growth Trajectory on Wage with Auxiliary Covariates: Applying BCH Approach in Growth Mixture Modeling

Hawjeng Chiou, Wenyu Chiou

Abstract: The major advantage of mixture model applying to longitudinal data analysis is to identify the unobserved heterogeneity in the development of an outcome over time. By including the covariates, the dependency between time-(in)variant predictors and the heterogeneity represented by latent classes as well as the observed variables could be empirically examined. However, the identification of latent classes may shift due to the inclusion of the covariates in the measurement model simultaneously. To remedy the shift problem of latent class with covariates, Bolck, Croon, Hagenaars (2004) proposed a sequential modeling by introducing the weighted multiple group analysis in the final stages (i.e., BCH approach). To demonstrate the GMM with BCH analysis, we use a longitudinal survey of 4153 Taiwanese adults. Five waves of wage and working hours per week along with background variables were collected. To illustrate the estimation of wage trajectory with heterogeneity, we fit the two-class and three-class unconditional linear GMM, followed by the BCH approach to examine the effects of covariates. For the two-class GMM, the majority class (97.7%) had a linear growth trajectory with a higher intercept and flatter slope than the minority class (2.3%). The three-class GMM identified a similar flat-trajectory majority class and two minority classes. The follow-up BCH procedure revealed that the same identification of latent classes established in the unconditional stage. For the major class of both models, sex, educational year, and age all have statistically significantly positive influences on intercept and slope; in contrast, no statistically significant effects on the same set variables reported for minor classes, showing a strong homogenous feature of the small group of subjects. This study demonstrates the advantage of BCH approach on the robustness of the examination of heterogeneity within growth modeling.

Do Executive Function and Effortful Control Co-Develop? Evidence from Latent Growth Curves with Structured Residuals and Multivariate Growth Mixture Models

Emily M. Weiss

Abstract: Self-regulation is a central developmental task of early childhood and is considered essential for children's success during elementary school. It has typically been conceptualized as Executive Function (EF) or as Effortful Control (EC), depending on the research tradition and whether cognition or behavior is emphasized. These aspects of self-regulation are theorized to emerge from an intertwined developmental process, but the nature of their relation throughout elementary school has not been established. This paper addresses this gap in the knowledge by examining the co-development of EC and EF using latent growth curve approaches. Bivariate latent growth curves with structured residuals (Curran et al., 2014) and multidimensional growth mixture models were compared. Overall, results do not reveal a systematic co-developmental relation between EC and EF when accounting for intra-individual development. Findings will be discussed with regard to developmental theory, educational application, and methodological implications.

Concurrent Paper Session 3 Tuesday 2:45 – 3:45 pm

Session 3A: Disaggregating Level-Specific Effects in Cross-Classified Multilevel Models (Room 202)

Disaggregating Level-Specific Effects in Cross-Classified Multilevel Models

Yingchi Guo, Jeneesha Dhaliwal, Jason D. Rights

Abstract: In social science research, often observations are nested within multiple types of non-hierarchical clusters, forming a cross-classified structure. This paper discusses ways that, in cross-classified multilevel models, slopes of lower-level predictors can implicitly reflect a weighted average of multiple effects (e.g., a purely observation-level effect as well as a unique between-cluster effect for each type of cluster). The issue of conflating multiple effects of lower-level predictors is well-recognized for non-cross-classified multilevel models, but has not been fully discussed or clarified for cross-classified contexts. Consequently, even when a pure lower-level effect is desired, researchers are left with little choice but to report a conflated effect. In this paper, we show why this common practice is sometimes problematic, and provide recommendations for how to disaggregate level-specific effects in cross-classified models. We provide a novel suite of options that include fully cluster-mean-centered, partially cluster-mean-centered, and contextual effect models, each of which provides a unique interpretation of model parameters. We further clarify how to avoid both fixed and random conflation, the latter of which is widely misunderstood even in non-cross-classified models. We provide corroborative simulation results showing the deleterious impact of fixed and random conflation in cross-classified models, walk through pedagogical empirical examples to illustrate the disaggregation of level-specific effects, and conclude with recommendations for practice.

Session 3B: Dealing with Missing Data (Room 305)

The Impact of MNAR Dropout on Estimation of Latent Growth Curve Models with Binary Observed Variables

Jason T. Newsom, Brian T. Keller, Nicholas A. Smith, Mallory R. Kroeck

Abstract: Missing data due to dropout (attrition) is a common problem in longitudinal studies and more research is needed to understand the biasing effects on longitudinal modeling. Using Monte Carlo simulation, we evaluated the performance latent growth curve (LGC) models with binary observed variables to examine intercept and slope mean estimation when the dropout pattern is missing not at random (MNAR) for five time points. Parameter and standard error biases for three estimation methods were compared: weighted least squares with mean and variance adjustment (WLSMV), categorical robust maximum likelihood (MLR), and Bayes. Three missing data mechanisms—conditional missing at random (conditional MAR), partial missing not at random (partial MNAR), and fully missing not at random (full MNAR)—and two sample sizes—$N = 200$ vs. $N = 1000$—were manipulated. Bayesian estimation showed lower parameter bias across all conditions than WLSMV and MLR estimation. Standard errors were generally better estimated than means except for MLR with small sample sizes. These findings provide important guidance about the preferred estimation approach when study dropout results in data MNAR.

Estimating Treatment Effects in Partially Clustered Randomized Controlled Trials with Missing Data: Challenges and Solutions

Manshu Yang

Abstract: Partially clustered randomized controlled trials are widely used in behavioral and health-related research to assess the effectiveness of intervention strategies. In a partially clustered trial, individuals are clustered into intervention groups in one or more study arms, for the purpose of intervention delivery, whereas individuals in other arms (e.g., the waitlist control arm) are unclustered. Missing data are almost inevitable in partially clustered trials and could pose a major challenge in drawing valid research conclusions. This study focuses on handling auxiliary-variable-dependent missing-at-random (A-MAR) data in partially clustered studies. Five methods were compared via a simulation study, including three multiple imputation methods that employ simultaneous joint modeling (MI-JM-SIM), arm-specific joint modeling (MI-JM-AS), and arm-specific substantive-model-compatible sequential modeling (MI-SMC-AS), as well as a sequential fully Bayesian estimation approach using either non-informative priors (SFB-NON) or weakly-informative priors (SFB-WEAK). Results suggest that the MI-JM-AS method outperformed other methods when the variables with missing values only involved fixed effects, whereas the MI-SMC-AS method was preferred if the incomplete variables featured random effects and the number of clusters was relatively large.

Session 3C: Modeling Item-Level Heterogeneous Treatment Effects with the Explanatory IRT (Room 307)

Modeling Item-Level Heterogeneous Treatment Effects with the Explanatory Item Response Model: Leveraging Online Assessments to Pinpoint the Impact of Educational Interventions

Joshua Gilbert, James Kim, Luke Miratrix

Abstract: Analyses that reveal how treatment effects vary allow researchers, practitioners, and policymakers to better understand the efficacy of educational interventions. In practice, however, standard statistical methods for addressing Heterogeneous Treatment Effects (HTE) fail to address the HTE that may exist within outcome measures. In this study, we present a novel application of the Explanatory Item Response Model (EIRM) for assessing what we term "item-level" HTE (IL-HTE), in which a unique treatment effect is estimated for each item in an assessment. Results from data simulation reveal that when IL-HTE are present but ignored in the model, standard errors can be underestimated and false positive rates can increase. We then apply the EIRM to assess the impact of a literacy intervention focused on promoting transfer in reading comprehension on a digital assessment delivered online to approximately 8,000 third-grade students. We demonstrate that allowing for IL-HTE can reveal treatment effects at the item-level masked by a null average treatment effect, and the EIRM can thus provide fine-grained information for researchers and policymakers on the potentially heterogeneous causal effects of educational interventions.

Session 3D: Comparing Latent and Composite Constructs with Nested Equivalence Testing (Room 306)

Comparing Latent and Composite Constructs with Nested Equivalence Testing

Danielle Siegel, Mijke Rhemtulla

Abstract: Confirmatory composite analysis (CCA) and the pseudo-indicator model (PIM) are two recent techniques developed to incorporate composites (e.g., sum scores) and their indicators into structural equation models. Modeling composites within a latent variable framework allows researchers to get fit statistics for their models and impose model constraints, which can be difficult or unfeasible in other composite-based modeling frameworks such as partial least squares path analysis. Decisions about whether to model a variable as a reflective common factor or as a composite are made based on theory. However, since CCA and PIM allow researchers to specify composites within a traditional SEM framework, there is the possibility of comparison between models where a variable of interest is defined as a reflective common factor versus as a composite. We illustrate that, under certain conditions, a confirmatory factor analysis is nested within its CCA counterpart. In turn, the CCA model is nested within the more general PIM. We use two empirical examples to conduct a chi-square difference test across the three models. Results suggest support for a composite version of the models in both analyses. Limitations and future directions of this technique are discussed.

Session 3E: Latent Class Analysis (Room 108)

Modeling Careless Responding in Ambulatory Assessment Studies Using Multilevel Latent Class Analysis: Factors Influencing Careless Responding
Kilian Hasselhorn, Charlotte Ottenstein, Tanja Lischetzke
Abstract: As the number of studies using ambulatory assessment (AA) has been increasing across diverse fields of research, so has the necessity to identify potential threats to AA data quality such as careless responding. The goal of the present research was to identify latent profiles of momentary careless responding on the occasions level and latent classes of individuals (who differ in the distribution of careless responding profiles across occasions) on the person level using multilevel latent class analysis (ML-LCA). We show how ML-LCA can be applied to model careless responding in intensive longitudinal data. We used data from an AA study in which the sampling frequency (3 vs. 9 occasions per day, 7 days, n = 310 participants) was experimentally manipulated. We tested the effect of sampling frequency on careless responding using multigroup ML-LCA and investigated situational and respondent-level covariates. The results showed that four Level 1 profiles ("careful", "slow", two types of "careless" responding) and four Level 2 classes ("careful", "frequently careless", two types of "infrequently careless" respondents) could be identified. Sampling frequency did not have an effect on careless responding. On the person (but not the occasion) level, motivational variables were associated with careless responding.

Using Latent Class Models to Derive Subtypes for the Enneagram Personality Typology
Jay Magidson
Abstract: The Enneagram is a popular personality typology, the primary results of an Enneagram assessment being a specification of which of nine personality types a test taker most identifies with – their highest scoring 'core' type. In this paper we use Latent GOLD to fit several latent class (LC) sequential choice models to Enneagram ranking data and use the resulting classes to supplement one's core type specification with subtypes that provide more detailed information about an individual's personality. As it turns out, the classes extracted from the 3-class model capture similar personality traits as certain theoretical 'subtypes' that have been recommended by some Enneagram authors as important supplements to the basic Enneagram typology. In contrast to theoretical approaches to obtain subtypes, the LC approach is atheoretical as it relies upon data rather than theory to derive the subtypes. It is an extremely parsimonious data-driven approach as the 3-class LC cluster model reduces the 9! = 362,880 rankings to only 3 classes. We also present results from an equally parsimonious LC model (in terms of the number of parameters) with two dichotomous discrete factors which yields 2x2=4 joint classes, and we introduce two new graphical displays to compare these models.

Session 3F: Bifactor Measurement Models (Room 308)

Confirmatory Bifactor Measurement Models: Their Utility in Scale Development, Psychometric Modeling and Research with Latent Variables and/or Multidimensional Constructs
Rafael Ramirez
Abstract: Although introduced more than 70 years ago, the Bifactor measurement model was overshadowed for many decades by the more general and restrictive correlated factor model and

the hierarchical second order factor model. However, many researchers have rediscovered the procedure, in particular the Confirmatory Bifactor model (CBM), as a useful alternative model to consider when evaluating measurement models of a set of scale items. The CBM, the correlated factor and the hierarchical second order factor are nested, with the CBM being the most general less restrictive model. Therefore, model fit comparison is relatively simple with standard SEM software. CBM can be of much utility in solving problems in measurement and conceptualizing and refining psychological constructs. The method is illustrated using data from a sample of 500 incarcerated adults on a scale of adult ADHD symptomatic behaviors.

Investigating the Factor Structure of Sense of Social and Academic Fit Scale: A Multilevel Bifactor Study
Lizzy Wu, Gabriella Jiang, Nidia Ruedas-Gracia, Taiylor Rayford, Shiyu Sun
Abstract: Although the Sense of Social and Academic Fit (SSAF) scale is popularly used for measuring college students' sense of university belonging, its factor structure has yet to be extensively examined. Moreover, when studying the sense of university belonging in longitudinal studies, the clustered structure within data is often ignored, resulting in distorted results. The present study used multilevel bifactor analysis techniques to investigate the factor structure of the SSAF scale with longitudinal data. Moreover, the present study evaluated whether or not the SSAF scale should be conceptualized as unidimensional and examined the negative wording effect at the between-individual level. Intraclass correlation coefficients warranted the multilevel modeling. Results showed that a bifactor model with two specific factors at both levels fit the data well. Besides the negative wording effect added at the between-individual level, the factor structures differed across the two levels within a multilevel framework. Specifically, the SSAF scale was rather unidimensional than multidimensional at the between-individual level. In contrast, the SSAF scale included a general factor and two specific dimensions at the within-individual level. Our findings support using the SSAF scale to measure college students' sense of university belonging in longitudinal studies.

Concurrent Paper Session 4 Tuesday 4:00 – 5:00 pm

Session 4A: Partially Nested Designs (Room 202)

Statistical Power for Multisite Partially Nested Regression Discontinuity Designs
Fangxing Bai, Benjamin Kelcey
Abstract: When experimental design is impractical or unethical, a common alternative is the regression discontinuity (RD) design because it often buttresses high-quality inferences on local treatment effects. In the context of education, for instance, there is extensive literature on how to adapt these designs to accommodate the types of multilevel structures and programs frequently encountered in schooling (e.g., Bloom et al., 2007; Schochet, 2009). Despite the flexibility and durability of RD designs in single and multilevel contexts, recent literature has noted that RD designs have not been adapted to the complicated structures and designs that are increasingly found in contemporary research (e.g., Hahn et al., 2001; Schochet, 2009; Raudenbush & Bloom, 2015). In this study, we expand the scope of RD designs to cover an important gap in the literature—RD designs in multisite partially nested structures. We advance RD designs in multisite partially nested structures by formulating models, developing principles of estimation,

sampling variability, and inference as well as expressions to estimate the statistical power to detect the main effects. The results provide a set of models, expressions, and software tools intended to inform and guide researchers in planning and analyzing studies with multisite partially nested RD designs.

Estimation, Statistical Power in Partially Nested Multisite Clustered-Randomized Trials
Kyle Cox, Yanli Xie, Benjamin Kelcey
Abstract: We consider a class of partially nested designs that induce meso- or intermediate-levels in the treatment condition (only) — partially nested multisite cluster-randomized trials. The novel feature of this design is that within each site (e.g., school), the intervention induces an additional level of intermediate nesting among people in the treatment condition through participation in a social treatment or experience that does not exist in the control condition. Although a review of prior substantive literature suggests that partially nested multisite clustered-randomized trials are present in theory, the appropriate analysis of this structure has largely gone unexamined. In this study, we develop the statistical theory that motivates this design, map out its use in practice, and develop estimation methods and closed-form expressions to track treatment effects, their sampling variability, and the statistical power to detect the main treatment effect and its variability across sites. Collectively, the results provide the core tools to effectively and efficiently design and analyze studies drawing on partially nested multisite cluster-randomized.

Session 4B: Longitudinal Causal Effect Estimation (Room 305)

Estimating Longitudinal Causal Effects: A Comparison of Marginal Structural Models, and Structural Equation Modeling Approaches
Jeroen D. Mulder
Abstract: Assessment of the causal effect of a time-varying exposure on an end-of-study outcome requires the implementation of causal inference principles in a longitudinal context. A main concern, especially when using observational data, is adjustment for time-dependent confounding specifically in the presence of confounding variables that are affected by previous exposure. This paper contrasts two approaches for estimating joint treatment effects in this context: The marginal structural model (MSM) approach, and the structural equation modeling approach (SEM). First, we compare the causal assumptions that underlie MSM and SEM approaches by discussing how both are related to each of three general aspects key to investigation of causal effects: Formulating a causal research question using the counterfactual framework, identification of the causal effect, and estimation of the causal effect using finite sample data. Second, we compare the statistical properties of a typical MSM estimator (inverse-probability-of-treatment weighted ordinary least squares) and three different SEM-related estimators (including traditional path analysis) for varying degrees of model misspecification.

Asking and Answering Causal Questions Using Longitudinal Data
Rafael Quintana
Abstract: The first step of empirical research is to define a research question. However, there is little guidance regarding which causal questions we can ask using longitudinal data, and what is the best way to answer these questions. The goal of this project is to distinguish between three causal quantities that are relevant in social and behavioral research: the contemporaneous

treatment effect, the cumulative treatment effect, and the long-term treatment effect. I will define these quantities, clarify which causal assumptions are needed to identify these effects, and present statistical models that can be used to estimate these quantities. I will illustrate the methods discussed by studying how peer victimization in school affects internalizing behaviors.

Session 4C: Characterization and Identification of Multivariate Latent Manifolds: Analytically Resolving Factor Score Indeterminacy (Room 307)

Characterisation and Identification of Multivariate Latent Manifolds: Analytically Resolving Factor Score Indeterminacy

Landon Hurley

Abstract: Factor score indeterminacy is a foundational problem of measurement in the Social Sciences. However, since its introduction in 1928, the concept has been exclusively treated as a non-resolvable fact, rather than a criterion condition to satisfy. We explore here a solution as a system of linear equations which satisfies two independently developed orthonormal criterion which satisfy a projective geometric duality. This allows us to define, uniquely, a common factor score solution which satisfies both criteria conjointly, and thus removes the indeterminacy exhibited as multiple equally well-fitting estimation problems upon finite length tests. We present the Maximum Entropy characterisation of this solution as a rho-tau embedding with gauge freedom, common topics of interest in the field of information geometry, which allows a probabilistic demarcation of the linear estimation problem. Together, although not resolving the necessary increase in total information, we do present a solution which restricts the solution factor score space to one which uniquely satisfies both rank and score ordering, thereby identifying a well-fitting factor model with unique factor scores, without the necessity of asymptotic weak convergence. Relationships to non-parametric factor models and the Rasch model are also explained using this same linear topological framework.

Session 4D: Latent Variable Models for Location, Shape, and Scale Parameters (Room 306)

Latent Variable Models for Location, Shape, and Scale Parameters

Camilo Cardenas-Hurtado, Irini Moustaki, Giampiero Marra, Yunxiao Chen

Abstract: Latent variable models (LVM) are used in the Social Sciences for measuring unobservable factors and for dimensionality reduction of multivariate data. Traditionally, the focus has been on modelling the conditional mean of observed variables that, given the latent factors, are assumed to follow a distribution in the exponential family. However, in some applications we are interested in either modelling higher order moments (variance, skewness, kurtosis), or we have variables whose empirical distributions do not satisfy the exponential family assumption (mixtures, zero-inflation, heaping). In these cases, there is a need for a distributional approach to LVM. We aim to fill this gap by proposing a general class of latent variable models for location, shape, and scale parameters (LVM-LSS), by extending the work of Rigby and Stasinopoulos (2005, J. of Roy. Stat. Soc) to models with latent variables. In essence, we model the different parameters that characterise the conditional distributions for the manifest variables as linear functions of the factors. We propose a penalised maximum likelihood estimation that uses an automatic procedure for multiple tuning parameter selection. The proposed framework is tested via simulation studies. We present empirical applications using data from a political attitude assessment survey.

Session 4E: Interactive, Automated, and Dynamic SEM (Room 308)

Interactive, Automated, and Dynamic SEMs for Maximal Productivity
Laura Castro-Schilo
Abstract: This presentation introduces JMP Pro statistical software, which hundreds of universities have licensed for faculty and students' free use. JMP is powerful statistical software designed for anyone solving problems with data. Packed with tools for data preparation, analysis, and graphing, JMP leverages its interactive user interface, dynamic visualizations, and powerful statistical techniques to enable users to work efficiently. The presentation starts with a brief overview of the software and then focuses on a new platform for structural equation modeling (SEM). The SEM platform is unparalleled in its interactivity, user-friendly features, and streamlined workflow for users. The demonstration sheds light on (1) major gains in efficiency, (2) ease of use, (3) dynamic user feedback to prevent common errors, and (4) powerful visualizations to aid the interpretation of results. Each of these points is elucidated through examples fitting confirmatory factor analysis, path analysis, and latent growth curve models.

Session 4F: Mixed Effect Location Scale Models: Methodological and Substantive Considerations (Room 108)

Mixed Effect Location Scale Models: Methodological and Substantive Considerations
Jennifer Richardson, D. Betsy McCoach
Abstract: When the assumption of variance homogeneity is violated in multilevel analysis, standard procedure is to estimate robust standard errors. Although this approach acknowledges the assumption violation, it does not fix it. Fitting the traditional multilevel model when variance heterogeneity is present may be thought of as model misspecification. In addition to methodological considerations, random variation in level-1 residual variance may be of substantive interest in certain research contexts. Research on teacher ratings of student behaviors provides one such context. Most research in this area focuses on between-student differences in overall behavior. However, between-student differences in within-student variability in behavior may provide additional insight into student behavior patterns. The mixed effects location-scale model (MELSM) is an extension of the traditional multilevel model. The MELSM can be used to model random variation in within-student behavior in addition to random variation in overall student behavior. The presentation includes a description of the MELSM, a demonstration of the substantive benefits of modeling level-1 variance heterogeneity using MELSM, and an overview of simulation results comparing the MELSM to the standard multilevel model and the multilevel model with heterogeneous variances.

Concurrent Paper Session 5 Wednesday 8:00 – 9:00 am

Session 5A: A Two-Step Robust Estimation Approach for Inferring Within-Person Relations in Longitudinal Design (Room 102)

A Two-Step Robust Estimation Approach for Inferring Within-Person Relations in Longitudinal Design

Satoshi Usami

Abstract: Psychological researchers have shown an interest in disaggregating within-person variability from between-person differences. Especially for inferring reciprocal relations among variables at the within-person level, applications of the random-intercept cross-lagged panel model (RI-CLPM) has increased rapidly. RI-CLPM is a useful analytic option, but various kinds of (SEM-based) statistical models are available. Usami (in press; Psychometrika) recently aimed to synthesize SEM-based approach traditionally used in psychology and potential outcome approaches used in epidemiology, in order to enable flexible and robust inference of within-person relations. This presentation provides an overview, simulation results and an analytic example for this new estimation approach. This method assumes a data-generating process similar to that in RI-CLPM, and has several potential advantages: (i) the flexible inclusion of curvilinear and interaction effects for WPVS as latent variables, (ii) more accurate estimates of causal parameters for reciprocal relations can be obtained under certain conditions owing to them being doubly robust, even if unobserved time-varying confounders and model misspecifications exist, and (iii) the risk of obtaining improper solutions is minimized. Simulations demonstrate that the proposed approach works well in many conditions if longitudinal data with $T \geqq 4$ are available, and that the accuracy increases as T becomes larger.

Session 5B: Model Fit Issues in Measurement Models (Room 202)

ESEM, CFA, and Somewhere In-Between: The Effect of Measurement Quality on Model Fit Sensitivity

Tim Konold, Elizabeth Sanders

Abstract: The primary advantages of exploratory structural equation models (ESEM) over CFAs are that they allow for a more realistic evaluations of simple structure by allowing low cross loadings to be estimated rather than fixed to zero, resulting in less parameter bias and generally better fitting models. However, when model fit is reasonable for a CFA (over ESEM), CFA is preferred on the basis of parsimony. The current study examines the sensitivity of GFI values in helping to adjudicate between ESEM and CFA models with particular focus on the amount of cross-loading saturation in the population, the magnitude and pattern of these cross-loadings, and increases in the magnitude of the target loadings. Results of our Monte Carlo simulation show that 1) ESEM and CFA global fit indices are largely indistinguishable in the presence of positive cross-loadings, 2) negative cross-loadings and a mix of negative and positive cross loadings result in increased separation of GFI values, respectively, and 3) the separation is less pronounced for the CFI and TLI, and more pronounced for the RMSEA, as the magnitude of the target loadings increases.

Performance of Model Fit Indices in Confirmatory Factor Analysis on Misspecified Item Response Models with Local Dependence

Jiangqiong Li, Dubravka Svetina Valdivia

Abstract: Given the theoretical connections between Item Response Theory (IRT) and Confirmatory Factor Analysis (CFA), commonly used global fit indices (GFIs) in CFA can be applied to detect IRT model misspecifications. We examine the performance of selected GFIs when the local independence (LI) assumption in IRT models is violated via simulation studies, focusing specifically on dichotomous and polytomous Rasch models. We investigate the sensitivity of GFIs in detecting local dependent (LD) under varying simulation conditions and examine the relationship between values of GFIs and deviation in IRT parameter estimates. Furthermore, by considering dichotomous and polytomous Rasch models, we hope to identify the factors, such as the sample size, the test length, the response categories, and/or the magnitude and complexity of LD, etc. that impact the performance of GFIs in CFA.

Session 5C: Causal Inference I (Room 305)

Synthetic Control Models for Causal Inference in Observational Social and Behavioral Sciences

John M. Felt, Zachary Fisher, Chad Shenk

Abstract: Synthetic control models (SCMs) are a quasi-experimental method for causal estimation in observational research. This approach generates a "synthetic" control via a weighted average of never-exposed comparison units, where weights are chosen to minimize the squared imbalance of lagged outcomes of the exposed and never-exposed unit before the exposure event. Although SCMs have been touted as the most important innovation in policy research in the past 15 years, these methods have not been widely adopted in social and behavioral sciences. Recent innovations extending SCMs to included multiple exposed units, where the exposure can occur at different times for each unit, has made these approaches more feasible to the type of data collected in areas including child maltreatment. The purpose of this presentation is to introduce the synthetic control method and demonstrate its application for studying causal associations between child maltreatment and behavior problems.

A Tutorial on Propensity Score Analysis with Semi-Continuous Treatment

Huibin Zhang, Walter L. Leite

Abstract: Propensity score analysis has been a popular method to handle observational data to estimate the average treatment effect. However, there is little work examining how to apply propensity score analysis to semi-continuous exposure. This tutorial provides potential researchers with guidelines for how to estimate propensity scores and the average treatment effect for semi-continuous treatments. The tutorial introduces two parametric methods- the hurdle model and zero-inflated model- and one machine learning method- The gradient Boosting Machine. An illustrative example is provided with data from a popular virtual learning environment, with recommendations on how and when to use each model.

Session 5D: Partially/Disparately Nested Structures (Room 306)

Croon's Estimation of Multilevel Structural Equation Models with Partially Nested Data
Kyle Cox, Benjamin Kelcey
Abstract: This study examines a structure after measurement approach with a Croon's correction (SAM-Croon's) to estimate multilevel structural equation models (MLSEM) when data are partially nested. MLSEMs estimated using maximum likelihood appropriately reflect the complex multilevel systems and latent variables common in research across the social sciences. A crucial requirement for maximum likelihood estimation is a large sample size rendering it inappropriate in many planned research studies. However, SAM-Croon's has demonstrated unbiased and efficient estimates for MLSEMs even when sample sizes are limited. We extend this promising estimator to MLSEMs with data that are partially nested and use simulation studies to demonstrate its effectiveness as an alternative to typical full information maximum likelihood estimation.

Structural After Measurement Estimation of Disparately Nested Structures
Fangxing Bai, Ben Kelcey, Yanli Xie, Kyle Cox
Abstract: Despite the flexibility of multilevel structural equation modeling (MLSEM), a practical limitation many researchers encounter is how to effectively estimate (non)linear parameters with typical sample sizes when there are many levels of (potentially disparate) nesting with typical sample sizes when there are several levels of (potentially cross-classified or disparate) nesting. In this study, we develop a method of moments corrected maximum likelihood estimator for disparately nested SEMs that can accommodate latent interactions and is well-suited to the types of small to moderate sample sizes typically seen in prospective social science research. We then probe the consistency, variability, and convergence of the estimator with small to moderate samples. The estimator emerges as a practical alternative or complement to conventional maximum likelihood (ML) because it often outperforms ML in small to moderate n-level samples in terms of convergence, bias, and variance.

Session 5E: Difference-in-Differences: A Methodological Illustration (Room 107)

Difference-in-Differences: A Methodological Illustration
Meghan Cain
Abstract: Difference-in-differences (DID) offers a nonexperimental technique to estimate the average treatment effect on the treated (ATET) by comparing the difference across time in the differences between outcome means in the control and treatment groups, hence the name difference in differences. This technique controls for unobservable time and group characteristics that confound the effect of the treatment on the outcome. DID analysis can be performed on panel data or repeated cross-sectional data in which different groups of individuals are observed at each time period. In this methodological illustration, I introduce the theory behind DID analysis and then show how to fit DID models in Stata and interpret the results. I also discuss graphical diagnostics and tests and the standard errors that are appropriate to use in different scenarios, particularly when the number of groups is low.

Session 5F: Designing Against Bias: Identifying and Mitigating Bias in Machine Learning and AI (Room 108)

Designing Against Bias: Identifying and Mitigating Bias in Machine Learning and AI
David J. Corliss
Abstract: Bias in machine learning algorithms is one of the most important ethical and operational issues in statistical practice today. This talk describes common sources of bias and how to develop study designs to measure and minimize it. Analysis of disparate impact is used to quantify bias in existing and new applications. Also, a comparison algorithm can be developed that is designed to be fully transparent and without features subject to bias. Comparison to this bias-minimized model can identify areas as bias in other algorithms. The new open-source python package *Fairlearn* measures bias using confusion matrices and offers great promise to advance the mitigation of bias. These design strategies are described in detail with examples.

Concurrent Paper Session 6 Wednesday 10:45 am – 12:15 pm

Session 6A: Transparency, Reproducibility, and Replicability of Modeling Methods: Problems and Solutions (Room 102)

Transparency, Reproducibility, and Replicability of Modeling Methods: Problems and Solutions
Jessica Kay Flake, D. Betsy McCoach, Andrea Howard, Amanda Montoya
Abstract: In this session, we discuss the problems that advanced modeling methods pose for transparency, reproducibility, and replicability from the perspectives of developing, teaching, and implementing methods. Dr. Flake discusses the apparent neglect of open science reform amongst quantitative methodologists and the unique challenges advanced modeling create. Dr. McCoach discusses issues around developing decision rules for model building and reporting in her research and teaching. Dr. Montoya presents her research on an open science innovation, registered reports, and discusses how registered reports may provide a solution to these problems. Dr. Howard discusses open science solutions in the context of secondary data analysis, for which advanced modeling methods are often used. The session includes brief presentations as well as an interactive moderated panel component. All presenters have experience developing and using open science methods. This symposium showcases unique insights into the problems modeling methods pose for open science reform and provides a forum to discuss solutions.

Session 6B: Complex Designs which Stretch the Boundaries of Conventional IRT (Room 108)

Scoring the Complexities of Alphabet Knowledge
Jason Anthony, Janelle J. Montroy, Jeffrey M. Williams
Abstract: Children's alphabet knowledge is one of the best predictors of literacy acquisition. Accordingly, tests of letter sound knowledge are included in countless studies, yet less than a handful report which sounds produced by children when shown a letter were accepted as correct

answers. This is problematic insomuch as all English vowels and many English consonants are associated with multiple sounds in text. Approximately 2300 3- to 7-year-old monolingual English-speakers completed expressive and receptive tests of letter sound knowledge. Nominal response modeling of expressive test items explored utility of all technically "correct" responses, but it did not lead to unequivocal decisions about how to score some responses. Heteroscedastic ANOVAs (SAS PROC GLIMMIX) predicted children's receptive letter sound knowledge scores from subgroups of children based on whether they provided a wrong answer, a correct answer, or questionable alternative answer to a given item on the expressive test. Collectively, results indicated that when indexing children's letter sound knowledge, some technically correct responses should be scored as wrong (e.g, X-ray), some as partial credit (e.g., long vowel sounds reflecting the letter name and most soft consonants, like Gym, Cent), and some as correct (father, who, exact).

Multinomial Models for Unpacking Linguistically Informative Response Choices: Test Structure and Item Function

Lee Branum-Martin, Julie A. Washington, Katherine Rhodes

Abstract: The influence of African American English upon language testing has been acknowledged for decades, but testing accommodations rarely account for linguistic knowledge which may be expressed in non-mainstream ways. The Diagnostic Evaluation of Language Variation-Screening was designed to have clinically informative multi-category items to evaluated dialect-based language variation as well as risk for a language disability. Items are scored as correct/expected, dialect-based/expected error, or as other/incorrect. However, item-level models have not been published for the entire test and no analyses have evaluated the multiple-choice (multinomial) nature of the scoring. Although dichotomous item models have useful fit diagnostics for factor structure, multinomial models are more difficult to evaluate for fit and item quality. We used a priori confirmatory factor models to compare the published scoring structure against models of linguistically based blocks of items. Participants included 853 African American students in grades 1 to 5. Models using the published structure for the test fit poorly while models of separate item types fit well. Some items and many responses did not function adequately. Moreover, the relations across factors suggest unplanned overlap across linguistic skills as well as possible distortions due to method effects in test design.

Picturing the Impossible: Pictures, Repetitions, and Item-Set Card Effects in Measuring Vocabulary

Eleanor Fang Yan, Lee Branum-Martin

Abstract: Pictures are often used in vocabulary tests to evaluate receptive language in multiple-choice format. In the Test of Language Development, 4th edition, Intermediate, 6 pictures are presented on a card, with 7 to 11 items per card, from a total of 9 cards (80 items total). Some pictures were repeated as answers within a card and cards had their own flooring stop after 3 consecutive errors. Thus, each item was multiple-choice from the same single pool of 6 pictures within the card, potentially repeated within card, and the 9 cards might also have had method effects, all of which could disturb estimates of general vocabulary proficiency. In a sample of 895 students from first to fifth grades from a larger project, we fit confirmatory item models attempting to account for these design effects. A single factor 2PL model failed to fit well, while multifactor and bifactor models suggested these design effects were not ignorable, but all had estimation problems. Multinomial models were impossible to estimate with so many items, but

were suggestive when fit to subsets of the items. Difficulties in structure, design alternatives, and estimation are discussed.

Assessment of Children with Autism Using Eye Gaze and Generalized Additive Logistic Regression

Ryan P. Bowles, Emily Lorang, Courtney E. Venker, Madeline Klotz

Abstract: Eye gaze offers an opportunity to assess children's knowledge when the child is unable, due to age or disability, to respond with pointing or speaking. Standard analytical approaches for eye gaze data focus on summarizing each trial with a proportion of time the child is looking at the target location. This approach ignores the rich information available from examining the pattern of gaze changes over the time course of the trial. In this presentation, we use generalized additive logistic regression (Cho et al. 2022) to examine the time course of eye gaze to identify patterns associated with vocabulary knowledge for young children with autism. We auditorily presented words with a screen displaying two pictures. Some trials included an attention-grabbing wiggle to the correct answer 2300 ms after word onset. Contrary to expectations, children were less likely to look at the correct answer in wiggle trials compared to trials without the wiggle during the time window between word onset and wiggle onset. Our results highlight the utility of the generalized additive logistic regression approach but also highlight the need for better understanding of the time course of eye gaze for children with autism.

Session 6C: Reconsidering Lord's Paradox (Room 202)

Robust and Pseudo-Robust Solutions to Lord's Paradox

Robert E. Larzelere, Hua Lin

Abstract: ANCOVA-type analyses and difference-score analyses often produce contradictory results in non-randomized pre-post studies, a problem known as Lord's Paradox (Lord, 1967). Duncan et al. (2014) called for developmental science to test for robustness across different types of analyses. Unfortunately, robust consistency across analyses of residualized and simple difference scores often occurs artifactually, which we call pseudo-robustness. This paper identifies four pseudo-robust solutions to Lord's Paradox: (1) Adding the pretest as a covariate to difference-score analyses makes the treatment effect mathematically identical to ANCOVA's treatment effect. (2) Exact (or approximate) matching on pretest scores makes the treatment effects of the two change-score analyses equal (or approximately equal) to each other and to the treatment effect of standard ANCOVA prior to matching. (3) Centering pretest and posttest scores on pretest group means also produces equivalent treatment effects, but ones equivalent to the treatment effect of difference-score analysis prior to centering. (4) Treatment effects from the two change-score analyses also equal each other when increasing variance of the outcome over time exactly compensates for the difference in the treatment effects from the two change-score equations. We need to understand these building blocks more thoroughly to interpret more complex longitudinal analysis appropriately.

Does Group-Mean Centering Always Inflate Type I Error Rates in Multiple Regression Analyses?

Robert E. Larzelere, Hua Lin

Abstract: Huitema (2011) proposed quasi-ANCOVA to increase statistical power in randomized designs by controlling for group-mean centered covariates measured after the start of treatment.

Lin (2018) extended that strategy to dual-centered ANCOVA by centering the posttest scores as well as the pretest scores on the pretest group means. Following Huitema (2011), we expected dual-centered ANCOVA to increase the statistical power for predicting difference-score estimates of treatment effects. However, Lin's simulations indicate that the standard deviation of treatment effects across 1000 replications corresponds to the standard deviation of treatment effects for the original difference-score analysis rather than the smaller SD of treatment effects under standard ANCOVA before centering. We think this occurs because group-mean centering equates pretest group means artificially. This removes the random variation around equal group means that occurs in standard ANCOVA. In a randomized pre-post design, ANCOVA's null hypothesis incorporates regression toward the mean into its expected results, thereby shrinking the predicted SD of posttest scores relative to the SD of pretest scores (given equal variances). Our question: Does group-mean centering always inflate Type I error rates in regression-type analyses? If not, what determines whether group-mean centering inflates Type I error or not?

Dual Centered-ANCOVA with interaction, a Solution for Lord's paradox with Implications for Valid Causal Inferences in Longitudinal Analyses

Hua Lin, Robert E. Larzelere

Abstract: Developmental science is fundamentally about describing and explaining between-person differences in within-person change, yet there is little consensus about how to analyze change. Lord's (1967) paradox showed that two ways of analyzing change can produce contradictory results. ANCOVA-type residualized-score analyses (e.g., ANCOVA, multiple linear regression) have been preferred for causal estimates, but they are biased by stable pre-existing differences on outcome scores, according to recent critiques. Difference-score analyses (e. g., differences-in-differences, linear growth models) have less statistical power and cannot easily test crucial Pretest X Treatment interactions. This study addressed the second disadvantage with a promising innovation called dual-centered ANCOVA to test interactions in difference-score treatment estimates. Using the Fragile Families longitudinal data, we added a Treatment X Pretest interaction to analyses, first using the original data and then the centered data. Without the interaction, raw-score analyses produce inconsistent results from residualized-score and difference-score analyses when pretest group means differ. Dual-centered ANCOVA produces consistent results by centering the pretest and posttest scores on the pretest group means (Lin & Larzelere, 2020). The results show that standard ANCOVA and dual-centered ANCOVA yield similar interaction tests, but differ in whether psychotherapy seems to make depression symptoms better or worse.

Session 6D: Modeling Dynamic Processes II (Room 305)

Consequences of Sampling Frequency on the Estimated Dynamics of AR Processes using Continuous Time Models

Rohit Batra, Simran Johal, Meng Chen, Emilio Ferrer

Abstract: Continuous-time (CT) models are a flexible approach for modeling longitudinal data of psychological constructs. When using CT models, a researcher can assume one underlying continuous function for the phenomenon of interest. In principle, these models overcome some limitations of discrete-time (DT) models and allow researchers to compare findings across measures collected using different time intervals, such as daily, weekly, or monthly intervals. Theoretically, the parameters for equivalent models can be rescaled into a common time interval

that allows for comparisons across individuals and studies, irrespective of the time interval used for sampling. Our Monte Carlo simulation examines the capability of CT autoregressive (CT-AR) models to recover the true dynamics of a process when the sampling interval is different from the time scale of the true generating process. We use two generating time intervals (daily or weekly) with varying strengths of the autoregressive parameter and assess its recovery when sampled at different intervals (daily, weekly, or monthly). Our findings indicate that sampling at a faster time interval than the generating dynamics can mostly recover the generating autoregressive effects. Sampling at a slower time interval requires stronger generating autoregressive effects for satisfactory recovery, otherwise the estimation results show high bias and poor coverage. Based on our findings, we recommend researchers use sampling intervals guided by theory about the variable under study, and whenever possible, sample as frequently as possible.

Summing the up and downs of life: The Bayesian Reservoir Model of Psychological Regulation

Mirinda M. Whitaker, Cindy S. Bergeman, Pascal R. Deboeck

Abstract: Social and behavioral scientists are increasingly interested in examining dynamics within the processes they study, yet despite the wide array of processes they study, a fairly narrow set of models are applied to characterize dynamics within these processes. The Bayesian Reservoir Model provides an example of using Bayesian and multi-level estimation to fit a dynamic model that is tailored to match the process of self-regulation of stress. Two simulations compare the performance of the original version of the Reservoir Model to versions using Bayesian estimation and a multi-level modeling approach, alongside a substantive example of this model applied to data on stress and negative affect in older adults. Using Bayesian estimation provided more unbiased estimates (compared to the original estimation approach) and combining that with a multi-level modeling approach allowed for relatively unbiased estimation at small sample sizes (e.g., $N = 15$) and/or with short time series (e.g., 15 observations). The current expansion of the Reservoir Model demonstrates the benefits of leveraging the combined strengths of Bayesian estimation and multi-level modeling with a model that has been tailored to match the process of self-regulation.

Session 6E: Dyadic and Group Applications (Room 306)

Generalizability Theory Applied to Daily Relationship Quality: Substantive and Statistical Directions

Madison Shea Smith, Susan C. South

Abstract: Peoples' daily reports of their romantic relationship quality provide rich information. Although there now exists a great deal of research on administering measures of daily relationship quality, little is known about the basic processes of consistency and change in the resulting data. We apply generalizability theory to test the level at which daily relationship quality varies, how consistent these measurements are when submitted to common conceptual models, and whether they are impacted by individual differences in attachment security. Six daily reports from 101 couples were analyzed. Results demonstrate the feasibility of daily assessments when using individuals as the unit of analysis, suggest that the relative influence of days or items is low, and imply that researchers should be mindful of the large amount of error in

daily reports of relationship quality. We discuss these findings in light of implications for planning and interpreting future studies.

The Dynamics of Opinion Expression During Group Discussion

Joseph A. Bonito, Stephen A. Rains

Abstract: This paper investigated patterns of opinion expression in small discussion groups. Drawing from the punctuated equilibrium paradigm and dynamic systems theory, we examined how the polarity and strength of opinion expression by individual members influences the opinions expressed by other members over the course of a discussion. We also considered whether and how patterns of opinion expression are related to a group's opinion profile generated prior to discussion. Opinion strength and polarity were measured using opinion mining computer software. Our analysis used dynamic structural equation modeling and revealed that opinion expression is positively associated within groups (H1) and that mean opinion expression is positively associated with the mean group opinion profile (H2). The two remaining hypotheses, which addressed variance of opinion expression, were not supported. Discussion addresses implications for theory and research on small discussion groups.

Boosting Powers by Combining Spatial Econometrics with Dyadic Analysis and SEM: Racial/Ethnic Differences in Life Expectancy across the US States

Emil Coman, Peter Xiang Chen, Sandro Steinbach, Adrian-Gabriel Enescu, Monica Raileanu Szeles

Abstract: We propose the use of dyadic analytic modeling to handle spatial data analytic challenges, and apply this new approach to a life expectancy dataset from the USA, at state level. We describe the challenges, the classic solutions, the proposed new solution, and then describe the data and present results of comparative analyses. We review the nonindependence ('auto'-correlation) concept, and model it as inter-dependence between a region and its neighbors, using a method that models relationships between variables collected in dyadic format, proposed 23 years ago by Gonzalez and Griffin. This approach decomposes an association into within-dyad and across-dyads components, and for the spatial setup it can partial out the 'nonindependent' part from the 'true' co-variability part, by means of restructuring the spatial data in a specific (long/vertical) manner. We illustrate this approach with a US dataset for the 49 contiguous states from the CDC and Policymap.com containing state level %Minority, Life expectancy, and Income. We present and contrast results from bivariate spatial nonrecursive SEM models, as well as naïve/a-spatial and then proper spatial mediation models, run on summary pairwise correlations data emerged from the Gonzalez and Griffin 'dyadic' spatial procedure. We point to other potential modeling extensions, such as Kenny's actor-partner modeling approach.

Concurrent Paper Session 7 Wednesday 1:15 – 2:15 pm

Session 7A: Tutorial: Methods for Sensitivity Analysis to Omitted Confounders in SEM (Room 102)

A Tutorial on Methods for Sensitivity Analysis to Omitted Confounders in Structural Equation Modeling

Walter L. Leite, Zuchao Shen, Charles L. Fisk, Eric A. Wright, Jeffrey Harring, Katerina M. Marcoulides

Abstract: A few sensitivity analysis methods for structural equation modeling (SEM) have been developed recently based on using phantom variables to represent a potential omitted confounder. Sensitivity analysis is an important tool to probe the boundaries of the conclusions of a research study. However, these methods have not been widely disseminated in the SEM user community. We provide a tutorial of methods for sensitivity analysis in SEM implemented in the *SEMsens* package of the R Statistical Software. The sensitivity analysis shown in the tutorial is for a complex SEM of the relationship between job satisfaction and turnover. The results of the sensitivity analysis show how conclusions about explanatory theories may be susceptible to unobserved relationships with omitted confounders.

Session 7B: Issues in Multilevel Modeling (Room 202)

Modeling Options for Clustered Data: An Empirical Comparison of Hierarchical and Population Average Models

Bethany A. Bell, Jason Schoeneberger, Anthony A. Mangino

Abstract: Recently, researchers have suggested that perhaps hierarchical linear models (HLMs) are not always warranted when analyzing clustered data. Instead, researchers have proposed that population average models (PAMs) might be acceptable for research studies in the psychological and behavioral sciences. Although theoretically PAMs might suffice, suggestions provided in previous research were not based on empirical investigations. Thus, to further our understanding of the functioning of various modeling methods for clustered data, we conducted a simulation study to examine parameter estimate bias, Type I error control and statistical power of tests from two-level HLMs and PAMs. Outcomes were analyzed as a function of level-1 sample size, level-2 sample size, intercept variance, slope variance, model complexity, and modeling method.

Understanding the Consequences of Collinearity for Multilevel Models: The Importance of Disaggregation across Levels

Haley E. Yaremych, Kristopher J. Preacher

Abstract: In multilevel models, disaggregating level-1 predictors into their level-specific parts (typically accomplished via cluster mean centering) has crucial benefits for parameter estimates and their interpretation. However, the importance of level-specificity has not been addressed in the multilevel literature concerning collinearity. We demonstrate how the consequences of collinearity change across different centering specifications for level-1 predictors (i.e., the use of uncentered vs. level-disaggregated predictors). Additionally, we clarify how other data characteristics may exacerbate or mitigate the consequences of collinearity across different centering specifications. Finally, we illustrate the importance of disaggregation for collinearity diagnosis in multilevel data. Through analytic developments and a simulation study, we show

that when all or some level-1 predictors are uncentered, fixed effect point estimates are influenced by collinearity. In contrast, disaggregation of all predictors eliminates the possibility of biased point estimates due to collinearity; however, bias may arise in standard errors of fixed effect estimates and in random effect (co)variances estimates. The data conditions that exacerbate and mitigate this bias are explored. We also demonstrate the misleading nature of collinearity diagnostics applied to uncentered predictors. Overall, the necessity of disaggregation for identifying and managing collinearity's consequences in multilevel models is clarified in novel ways.

Session 7C: Modeling Single Subject Data (Room 305)

GLMMs for Overdispersed Count Data in SCED Studies: Does Autocorrelation Matter?
Haoran Li, Wen Luo, Eunkyeng Baek, Kwok Hap Lam, Wenyi Du, Noah Koehler
Abstract: Autocorrelation in single-case experimental designs (SCED) is not uncommon due to repeated measurements within each case. Previous studies found that ignoring autocorrelation in statistical analyses could lead to biased effect size estimates and inflated type I error rates when making conclusions about the effectiveness of interventions in SCEDs. Our study evaluated whether generalized linear mixed models (GLMMs) including Poisson, negative binomial (NB), and observation level random effects (OLRE) models proposed in Li et al. (2023) are robust to autocorrelated count data in SCEDs. A Monte-Carlo simulation study was conducted to examine the accuracy of estimators of immediate treatment effect and treatment effect on the trend and the associated inferential statistics with GLMMs. The results showed that the NB and OLRE models that can handle overdispersion are robust to autocorrelation when its magnitude is small (0.1 or 0.3). When the autocorrelation became large, the performance of NB and OLRE models regarding interval estimates and type I error rates was affected, but not significantly deteriorated. Implications and future research directions are also discussed.

Individual Participant Data Meta-Analysis Including Moderators: Empirical Validation
Mariola Moeyaert, Yukang Xue, Panpan Yang
Abstract: We have entered an era in which scientific knowledge and evidence increasingly informs research practice and policy. Given the exponential increase in the use of single-case experimental designs (SCEDs) to evaluate intervention effectiveness, there is an accumulating evidence base available for quantitative synthesis. Consequently, there is a growing interest in quantitative synthesis techniques suitable to meta-analyze SCED research. One technique that is developed and can be applied for this purpose is individual patient data (IPD) meta-analysis. The IPD approach provides detailed information about intervention effectiveness and intervention heterogeneity. IPD is a flexible modeling approach, allowing for a variety of modeling options such as modeling moderators to explain intervention heterogeneity. To date, no methodological research has been conducted to evaluate the statistical properties of effect estimates obtained by using IPD meta-analysis with the inclusion of moderators. To address this, we conducted a large-scale Monte Carlo simulation study. Based on the results, specific recommendations are provided to indicate under which conditions the IPD meta-analysis including moderators is suitable (i.e., resulting in unbiased, precise, and powerful estimates).

Session 7D: Enacting Critical Quantitative Methodology: Leveraging IRT to Advance Critical Consciousness Measurement (Room 108)

Enacting Critical Quantitative Methodology: Leveraging IRT to Advance Critical Consciousness Measurement

Matthew Diemer, Michael B. Frisby, Andres Pinedo, Emanuele Bardelli, Elise Wilkerson, Sara McAlister

Abstract: This presentation details an enactment of critical quantitative (CritQuant) methodology, applying Item Response Theory (IRT) and MIMIC modeling. We leveraged IRT to advance critical consciousness measurement. Critical consciousness is comprised of three dimensions, (i) critical reflection, the analysis and critique of structural inequalities, (ii) critical motivation, the motivation and perceived capacity to effect change, and (iii) critical action, social action to redress inequity. A wave of recent instruments measuring critical consciousness among minoritized youth have been validated, generally with factor analyses. Yet, whether these measures efficiently assess critical consciousness across the theta distribution or contain redundant, or non-informative, items is unknown. Using IRT methods – specifically, a 2-PL graded response model - the long-form Critical Consciousness Scale was scrutinized for redundant and/or imprecise items. This yielded a short version, the Short CCS (ShoCCS). The 13-item and multidimensional ShoCCS was internally consistent (Omega total subscale estimates ranged from .80-.92) with similar information distributions as the longer measure. The ShoCCS did not exhibit DIF across gender and racial/ethnic groups, in a series of MIMIC models, suggesting invariance across these social identity categories. This research, carried out by a diverse and multigenerational team of researchers, illustrates an enactment of CritQuant.

Session 7E: The Construction and Estimation of Multidimensional Linear Factor Models without Parametric Assumptions (Room 306)

The Construction and Estimation of Multidimensional Linear Factor Models without Parametric Assumptions

Landon Hurley

Abstract: Linear latent variable models have historically assumed multivariate normality upon both the observed and latent spaces, to simplify the problem of mathematical estimation. This is due to the complexity of embedding non-linear spaces upon linear latent factors, inducing vast requirements with respect to sample sizes and necessary model structure, and precluding desirable finite sample properties. We introduce a linear correlation Gauss-Markov estimator, precluding requiring bivariate normality and dominating the Spearman and Kendall alternatives, while also satisfying the Eckart-Young requirements. This enables a non-parametric linear factor model structure, which we demonstrate to satisfy the Kullback-Leibler divergence criterion of model adequacy. We explore several conditions which encompass the embedding of non-linear observed Euclidean space into a multivariate latent space, and demonstrate these results are identical to the Pearson factor analysis model when parametric assumptions are true, yet remain unbiased and maximally informative upon non-Gaussian data. These results further define a well-fitting Hadamard solution to factor score estimation, thereby resolving factor score indeterminacy. In turn we also address the unresolved generalisation of (non-parametric) Mokken scaling to multidimensional latent manifolds, all within a common analytic framework.

Concurrent Paper Session 8 Wednesday 2:30 – 4:00 pm

Session 8A: Dyadic and Group Designs (Room 102)

Understanding Group Effects Using the Co-Partner Design
David A. Kenny
Abstract: Group effects are typically viewed as a constant that is added to each group member's score, i.e., a random intercept. There is another additional possible effect: a partner effect or a term that is added to the score of every other member in the group. To be able to measure a partner effect, persons need to be in two or more groups. The co-partner model (Bond & Cross, 2008) for a person in a group has an actor effect, which is a term for that person in all of their groups, partner effects for all of the other members in the group, and a random intercept for the group. The variance and covariance components of the co-partner model can be estimated by either an ANOVA-like variance decomposition with balanced designs or multilevel modeling with unbalanced designs. Two different empirical illustrations are discussed.

Sample Considerations for Detecting Person, Dyad, and Contextual Effects Using the Common Fate Model for Dyadic Analysis
Robert E. Wickham
Abstract: Several interesting variations on the Common Fate Model (CFM) for dyadic analysis have emerged over the past decade. Recent work by Wickham (2023) describes two variations on the Common Fate Model (CFM) for dyadic analysis: the contextual CFM, and the Between-Within CFM (BW-CFM). The contextual CFM provides access to the pure (within) Person level coefficients, as well as a (between) Dyad level coefficient representing the contextual effect, or the discrepancy in the within- and between-dyads coefficients. Although the BW-CFM also produces pure within-dyad coefficients at the Person level, the Dyad level coefficient represents the pure between-dyads effect. The current simulation study explores the sample and design characteristics associated with power to detect regression relationships at the Person (pure-within) and Dyad (pure-between) levels, as well as contextual effects when applying the BW-CFM. Manipulated factors include sample size, effect size at the Person and Dyad levels and the intra-class correlation. Sample sizes of 150 were sufficient to detect medium effect sizes at the Person and Dyad levels, but 300 were needed to detect small effects. As expected, sample size was positively associated with power to detect Person, Dyad, and Contextual effects.

Session 8B: Optimal Design and Cost Considerations (Room 202)

Power Analysis and Sample Size Planning in the Design of Two-Level Randomized Cost-Effectiveness Trials
Wei Li, Nianbo Dong, Rebecca Maynard, Benjamin Kelcey, Jessaca Spybrook, Yue Xu
Abstract: Randomized cost-effectiveness trials (RCETs) are increasingly used to evaluate the causal effects of interventions and the cost of achieving these effects in social science. One key consideration when designing an RCET is to determine the sample size that guarantees adequate power to detect the cost-effectiveness of the intervention. This study discusses statistical power analysis methods for two-level RECTs, where for example, students are nested within schools, and the treatment is either at the school level (i.e., a cluster design) or the student level (i.e., a multisite design). We also demonstrate the application of these methods for the designs of two-

level RCETs using a free and user-friendly tool – PowerUp!-CEA and provide practical recommendations on statistical power analysis to help applied researchers plan their RCETs.

Optimal Design of Experimental Studies Under Condition- and Unit-Specific Cost Structures

Zuchao Shen, Benjamin Kelcey

Abstract: Prior work has identified the optimal sample allocation to achieve maximum statistical power under a fixed budget. These frameworks produce different types of constrained optimal designs because they typically assume that costs among treatment and control clusters/individuals are equal and only optimize the sample ratios across levels. In this study, we relax cost equality assumptions and optimize power in the presence of unequal costs across treatment conditions and levels of hierarchy. We identify sampling ratios across levels and treatment conditions. The results demonstrate that previous frameworks are special and constrained cases of the proposed framework and that the proposed cost framework can identify more efficient sample allocations. Efficiency gains are fairly robust to the misspecification of initial values of design parameters and cost structures. The solutions are implemented in the R package *odr*.

Optimal Design of Multisite-Randomized Trials Investigating Mediation Effects Under Unequal Costs

Zuchao Shen, Wei Li, Benjamin Kelcey, Walter Leite, Huibin Zhang

Abstract: Mediation analyses can investigate the mechanisms of interventions. When designing studies to investigate mediation effects, one important consideration is to determine an efficient and powerful design by leveraging the sampling cost information. This is usually addressed in an optimal design framework that derives the optimal sampling ratios across treatment conditions and levels such that optimal designs achieve the maximum statistical power under a fixed budget (or use the minimum budget to achieve a fixed statistical power). Prior research on optimal sampling has mainly focused on main effects. This study develops an optimal design framework for multisite-randomized trials investigating mediation effects. The results show that the developed framework can identify more efficient and powerful allocations than conventional statistical power analysis frameworks that do not explicitly consider sampling costs and budget.

Session 8C: Longitudinal Applications II (Room 108)

An Application of Random Changepoint Models to Cognitive Aging Research

Zachary J. Kunicki, Yi Feng, Douglas Tommet, Sharon K. Inouye, Richard N. Jones

Abstract: Random changepoint models are piecewise linear models that estimate a changepoint (i.e., knot) that separates two linear functions as a random parameter. This allows researchers to make inferences on when the individual trajectories change in functional forms during the course of longitudinal follow-up and how this varies from person to person. This methodology is highly applicable to cognitive aging research, where one of two scenarios is likely to be observed. Either the changepoint can indicate when practice and retest effects end, or if practice and retest is not accounted for, it is plausible that the random changepoint can signal the beginning of a faster pace of cognitive aging due to an acute event (e.g., a stroke) or chronic condition (e.g., dementia). This talk describes the random changepoint model and provides an applied example of the random changepoint model to cognitive trajectories after surgery.

Application of a Novel Model for Analyzing Data from Randomized Pretest, Posttest, Follow-up Designs: Results from a Pediatric Randomized Behavioral Clinical Trial
Constance Mara
Abstract: Randomized pretest, posttest, follow-up (RPPF) designs are often used for evaluating the effectiveness of an intervention. These designs typically address two primary questions: (1) Do the treatment and control groups differ in the amount of change from pretest to posttest? and (2) Do the treatment and control groups differ in the amount of change from posttest to follow-up? Many approaches have been proposed for assessing change over time. Latent change models (LCMs) in particular allow researchers to evaluate group differences in change across each timepoint in a RPPF design. This approach is different from a LGCM or a MLM, which focus on the average amount of change across all timepoints. In 2012, we published a paper introducing a novel LCM for RPPF designs and compared it to existing LCMs. The novel LCM provided increased power for evaluating group differences in change from pretest to posttest and posttest to follow-up. This presentation illustrates the use of this novel LCM for RPPF designs by analyzing data from a recently conducted randomized behavioral clinical trial. The results from the novel LCM are compared to two popular methods for analyzing data from RPPF designs: ANCOVA and longitudinal MLMs.

Health in All Polices Approach: A Dynamic Modelling of Social Policies' Effect on Mental Health
Ekaterina Melianova
Abstract: It has been argued that addressing health issues is not a priority of public health alone but of the whole of society. Appropriate health policies involving partnerships in numerous government segments can contour health patterns and tackle health inequities. It was shown that interventions in sectors not usually associated with health could alter the social gradient in health since they define fundamental causes of health disparities. Such a multisectoral approach, recognising the interdependence of health-related and nonmedical policies, is termed "Health in All Policies" (HiAP). However, despite HiAP's potential to shape population health, there need to be more rigorous empirical studies showing how multisectoral governance works. This research systematically evaluates the dynamic effects of a spectrum of social policies (operationalised as local government expenditure) on population mental health. The study leverages geographically granular UK-based longitudinal local authority data and cutting-edge statistical tools under the Structural Equation Modelling framework (General Cross-Lagged Panel Modelling). Overall, the paper demonstrates the value of a multisectoral approach that helps maintain synergies in the government.

Session 8D: Sample Heterogeneity in Dynamic Psychological Processes (Room 305)

A Bayesian Multilevel Mixture Autoregressive Model
Xingyao Xiao, Feng Ji
Abstract: Intensive longitudinal data have become more available in recent decades due to the accessibility of wearable devices. Modeling psychological dynamics has become increasingly important in understanding complex relations and population heterogeneity. We propose a Bayesian multilevel mixture autoregressive model using Stan and discuss how Bayesian model assessment can be helpful to evaluate model fit.

Identifying and Explaining Sample Heterogeneity in Dynamic Psychological Processes Using ml-VARTree

D. Jody Zhou, Emilio Ferrer, Siwei Liu

Abstract: With the increasing popularity of intensive longitudinal data and analysis, it is now well-recognized that individuals vary importantly in their psychological processes. Data-driven subgrouping methods provide a way for researchers to examine potential qualitative, as well as quantitative, differences in these dynamic processes. Existing subgrouping methods, however, often fall short in explaining the differences between subgroups. This talk introduces ml-VARTree, which uses a decision-tree-based subgrouping algorithm for identifying subgroups of individuals characterized by different VAR models. Importantly, the algorithm automatically searches for covariates that predict subgroup memberships. We demonstrate the utility of this novel method through an empirical analysis identifying subgroups of individuals and covariates that explain why individuals are different or similar in their dynamic coregulation of affective states and food intake.

Impact of Temporal Order Selection on Clustering Intensive Longitudinal Data Based on Vector Autoregressive Models

Hairong Song, Yaqi Li

Abstract: Clustering of intensive longitudinal data provides a meaningful way to quantify sample heterogeneity in dynamic processes, assuming that such heterogeneity reflects the distinct nature of the studied processes. In this presentation, we 1.) introduce a VAR-based clustering technique, 2.) examine the impact of temporal order selection on clustering accuracy and parameter estimation by a simulation study, and 3.) demonstrate the application of the clustering technique through an empirical analysis.

Penalized Subgrouping of Heterogeneous Time Series

Christopher Crawford, Jonathan Park, Sy-Miin Chow, Zachary Fisher

Abstract: Recent technological advances have decreased the burden associated with collecting intensive longitudinal data in the social sciences. However, the best way to model multivariate time series data arising from multiple individuals is still an open question. Recently, a number of approaches have emerged for characterizing meaningful, qualitative heterogeneity in dynamic processes. One of these approaches, the multi-VAR framework (Fisher et al., 2021), is built upon the Vector Autoregressive (VAR) model and simultaneously estimates group- and individual-level models. Importantly, the multi-VAR approach accommodates both quantitative and qualitative heterogeneity in dynamics across individuals, and is compatible with an increasing number of penalization methods for structuring how information is shared across individual-level models. To this point, multi-VAR was only capable of resolving a single group-level model, presumably shared across all individuals. In reality it may be the case that for many processes shared dynamics also exist among subgroups, or clusters, of individuals. To address this limitation we extend the multi-VAR framework to allow for data-driven identification of subgroups and penalized estimation of subgroup-level dynamics. Simulation results and an illustrative example demonstrate the feasibility, strengths, and weaknesses of the proposed approach.

Clustering Analysis of Time Series of Affect in Dyadic Interactions
Samuel D. Aragones, Emilio Ferrer
Abstract: Longitudinal data, by nature, is a series of repeated measurements that can capture dynamics of individuals over time. These dynamics, such as those found in romantic relationships, can have effects on outcomes such as relationship satisfaction. In analyzing multivariate time series data, a problem that analysis techniques focus on is characterizing the heterogeneity within the data. Traditional analysis techniques such as ANOVA and linear regression are limited to patterns that describe the population as a whole. Despite recent developments in longitudinal data analysis, there are still difficulties in addressing variability within individuals. Clustering analysis is an approach taken to capture homogeneity in various subspaces and may be promising as a method of analyzing longitudinal time series data in a manner that can detect patterns within the repeated measures of individuals. A popular clustering method, Louvain, is used to assess the viability of such techniques in longitudinal data analysis. Concerns about the technique are addressed, including: [1] is heterogeneity described through our clusters, [2] what are the interpretations of said clusters, and [3], using information theory, are we able to determine how informative these clusters of partial time series are?

Session 8E: Modeling COVID-19 (Room 306)

Unobserved Components Models: Applications in Post-COVID Analysis
David J. Corliss
Abstract: Unobserved Components (UCM) is a type of State Space model used to detect and measure changes in a long-term baseline. This method decomposes time series into components including baseline, linear trend, periodic variations, and irregular. UCM is often used to analyze variations in baseline values due to change in the state of the system. In this presentation, UCM is applied to the evolution of the COVID-19 pandemic to determine if various factors such as business conditions have returned to pre-pandemic levels. Examples include durable goods as a leading indicator, GDP, and unemployment as a lagging indicator. All major statistical software systems now support UCM; code for this presentation is given in SAS. The SAS/ETS procedure PROC UCM implements PROC UCM functionality, options, graphical output, and model interpretation.

Using Cross-classified Multi-level Modeling to Identify COVID-19 Period Effects on Race and Gender Differences in Training Effectiveness
Youngmin Kim, Bridget McHugh, Rebecca Berenbon, Abena Anyidoho
Abstract: Recent research suggests school closures led to learning loss in math and reading, especially among marginalized groups. However, there is limited research on how COVID-19 impacted career-focused training. We addressed this need by comparing career-focused post-training assessment scores administered by 4,000+ workforce development trainers in 2019 and 2022. We conducted cross-classified multilevel models with three nesting variables: learner, trainer, and specific training course. Results suggest learning loss did not occur for all demographics. Overall, test scores actually increased between 2019 to 2022. However, interaction effects with year suggested COVID-19 impacted Black and female students differently. Female and African American learners' scores increased less than White and male learners. This exacerbated existing gaps between White and Black students. It also reversed learning gaps between male and female students, with female students performing worse in

2022. To examine the effects of specific curriculum, we conducted separate models for each job domain. Race and gender interactions with year differed across models, suggesting demographic-specific learning loss may vary depending on training content. Our results indicate COVID learning loss is not universal, and may depend on the specific curriculum (e.g., job domain) and learner characteristics.

Daily Associations of Emotion and Fatigue in College Students during the Early Stages of the COVID-19 Pandemic: An Application of Dynamic Structural Equation Modeling

Parisa Rafiee, Elizabeth Pauley, Melissa Rothstein, Amy L. Stamates, Manshu Yang

Abstract: Previous cross-sectional research suggests self-reported feelings of fatigue are associated with positive and negative affect; however, little is known on the daily reciprocal associations among them. The current study used dynamic structural equation modeling (DSEM), specifically the vector autoregressive model to investigate the temporal reciprocal associations between fatigue and negative/positive affect among college students during the early stages of the COVID-19 pandemic. The current study was a 21-day daily-diary study of 54 undergraduate and graduate students ranging from 18-40 years old. Results indicated current-day negative affect predicted next-day fatigue ($\Phi = 0.126$, $p < 0.01$) and next-day negative affect ($\Phi = 0.339$, $p < 0.01$), and current-day fatigue predicted next-day fatigue ($\Phi = 0.230$, $p < 0.01$). However, current-day fatigue did not predict next-day negative affect, and the dynamic reciprocal relationship between positive affect and fatigue was not statistically significant. Findings suggest higher than usual levels of negative affect may require a prolonged process and additional effort to self-regulate, resulting in elevated fatigue the following day.

Poster Abstracts

Poster Session and Reception Tuesday 5:00 – 7:00 pm

1. Convenience Samples and Measurement Equivalence in Replication Research: A Registered Report

Lindsay J. Alley, Jessica Kay Flake

Abstract: Convenience samples are common in psychological research. University students are a historically common convenience sample, but online data collection methods have led to the increasing use of crowdsourcing platforms, such as *MTurk* (Chandler & Shapiro, 2016; Strickland & Stoops, 2019). When data from different sources are pooled or compared, measurement equivalence is a concern, and little research has examined this issue for different convenience sample sources. We employed the open data from the *Many Labs* replication projects, which pools these sources, to examine measurement equivalence across convenience samples for a variety of measures. We test for measurement equivalence across available convenience sample sources, including *MTurk*, *Project Implicit*, and university participant pools, for the 9 measures (of 14 tested) that demonstrated adequate baseline model fit. Additionally, we conduct a sensitivity analysis examining the robustness of replication results to non-equivalence by reanalysing the effects using factor scores from partial equivalence models. Our results are of broad relevance to psychological researchers because it is important for researchers to understand which subgroups are likely to contribute to measurement non-equivalence, as it is rarely feasible to model every possible source.

2. Item-Weighted Expected a Posteriori Method for Improved Latent Trait Estimation in Item Response Theory

Udi Alter, Robert Philip Chalmers

Abstract: This study proposes a novel method for estimating respondents' abilities using item response theory (IRT) models with dichotomous or polytomous items. The proposed technique extends the expected a posteriori (EAP) estimation method (Bock & Aitkin, 1981) by incorporating a standardized weight function based on either user-defined values or item-fit statistics. The standardized weight values range from 0 to 1, where responses from items with lower weight values contribute less to the ability estimates. A Monte Carlo simulation was used to evaluate the new item-weighted expected a posteriori (IWEAP) approach and compare it to the common ability estimation techniques. Simulation conditions include various IRT models, the number of items, the number of response options, and the type of item-fit statistics. Results from the simulation show that the novel IWEAP approach offers more precise ability estimates which are robust to poor item fit and misspecification. Recommendations for using IWEAP in research and applied settings are discussed. We further offer additional insights about user generated weights and when these should be preferred over item-fit statistics.

3. Collapsing Categories for Likert Type Items: Dealing with Same Items, Different Categories in Longitudinal Data

P. Abena Anyidoho, Paul DeBoeck, Dorinda Gallant

Abstract: Researchers faced with sparse data for certain categories in Likert scales, would typically collapse categories to increase the sample sizes of cells. Much has been written about collapsing response categories to deal with sparse cell data counts, however, little exists on dealing with longitudinal data with the same items but different response categories. To investigate this, we explored merging longitudinal data (3 years) for two subscales that assesses the academic and social integration of ethnic/racial underrepresented minority students majoring in STEM. The first year had five categories with "no opinion" as the midpoint. The second and third years had six categories with "somewhat agree" and "somewhat disagree" as the midpoints. The other extreme category labels were the same across the three years. We collapsed the two mid-points in the 2nd and 3rd years and conducted a multigroup measurement invariance analysis using WLSMV estimator in lavaan. Using a metric invariance model, the results showed that the distance between the second and third thresholds in the second and third years was clearly larger (≥ 0.83) than in the first year (≤ 0.66). This finding has implications for interpreting what identical category labels mean depending on the neighboring category labels and the consequences of collapsing.

4. Investigating the Factor Structure of Educational Test Data: A Practical Illustration Using Multiple-Choice Credentialing Exams for Workforce Development Programs

Bridget McHugh, Rebecca Berenbon, Young Min Kim

Abstract: Factor analysis identifies latent constructs based on response consistencies and helps clarify what is being measured, thus informing the interpretation of scores. We provide a practical example of factor analysis to characterize latent constructs present in a multiple-choice assessment, utilizing data from credentialing assessments from workforce development programs. Using data from business ($N = 7{,}342$) and construction ($N = 4{,}285$) programs, we explored the internal factor structure of work skill assessments designed to measure multiple distinct soft and technical skills. Confirmatory factor analyses showed high correlations between all factors, contradicting the framework used by trainers, which assumes a clear delineation of hard and soft skills. We then investigated measurement equivalence of the structure across gender (male and female) and race (Black, Hispanic, and White). Results suggested metric invariance for business (race and gender) and construction (gender only), indicating the structure of the construction assessment may be different depending on the gender of the test taker. This research shows how factor structure can provide valuable information to stakeholders: trainers can take a more integrated instructional approach, as the factor analysis suggests soft and hard skills are not as distinct.

5. Designing Three Level Regression Discontinuity Studies With Partially Nested Structures

Fanxing Bai

Abstract: Regression Discontinuity (RD) designs are often seen as a flexible alternative to randomized designs because they allow some degree of control in terms of the treatment assignment while maintaining a rigorous basis for inference on local treatment effects (Cook, 2008). However, recent literature has noted that the RD designs have not been well adapted for the type of complex designs frequently observed in contemporary research (e.g., Hahn et al.,

2001; Schochet, 2009; Raudenbush & Bloom, 2015). An important gap in the RD literature is its application to partially nested structures (e.g., Kelcey et al., 2020; Sterba, 2017). Prior research has widely documented the prevalence and scope of partially nested structures across a broad array of disciplines (e.g., Lohr et al., 2014; Sanders, 2011). That same research has also suggested that the majority of extant empirical studies have neglected the treatment of these structures both in the design and analysis phases, and the methodological literature has only intermittently and incompletely developed (quasi-)experimental design strategies in this domain (e.g., Sanders, 2011). This poster advances RD designs for an array of partially nested structures by formulating models, developing principles of estimation, sampling variability, and inference as well as expressions to estimate the statistical power to detect the main effects. The results provide a set of models, expressions, and software tools intended to inform and guide researchers in planning and analyzing studies with partially nested regression discontinuity designs.

6. What's the Occasion? An Algorithm for Generating Optimal Measurement Schedules

Anne-Charlotte Belloeil, Kristopher Preacher

Abstract: The perils of poor study design are numerous and important—e.g., failure to identify stage-sequential processes, misrepresentation of mediation effects, concealment of the functional form describing the relation between two variables, and replication failures across studies with different measurement schedules. Moreover, longitudinal studies are expensive and time-consuming, and well-planned measurement occasions may help attenuate these problems. Yet, temporal design is frequently overlooked by researchers—perhaps because there exist relatively few tools at their disposal to plan and support their temporal design decisions. Prior research by Timmons and Preacher (2015) produced software that empowered scientists to manually supply a measurement schedule into software customizable for a selection of six functional forms, which returned the effect of spacing on the accuracy of parameter estimates. We built on this foundation of research to produce a customizable and adaptable R program for which measurement schedules need not be manually entered. It permits a broader range of functional forms, to better capture a wider variety of change phenomena. It automatically returns a temporal design that maximizes C-optimality—that is, the criterion within the optimal design framework that minimizes the variance of a best unbiased estimator of a predetermined linear combination of model parameters.

7. An Equivalence Testing Based Version of the Standardized Root Mean Squared Residual

Nataly Beribisky, Robert Cribbie

Abstract: A popular measure of model fit in structural equation modeling (SEM) is the standardized root mean squared residual (SRMR). Recently, equivalence testing has been used to assess model fit in structural equation modeling (SEM) and the present study proposes equivalence-testing based fit indices for the SRMR (ESRMR). We introduce different variations of ESRMR, including modified and unmodified equivalence bounds and varying methods of computing confidence intervals. Using Monte Carlo simulations, we compared these novel tests with traditional methods for evaluating the fit of an SEM model. Our results demonstrate that, in general, ESRMR tests correctly reject poor fitting models and have reasonable power for detecting good-fitting models. Our study also uses an illustrative example with real data to demonstrate how ESRMR may be incorporated into model fit evaluation and included in the

reporting of model fit. Our recommendation is that ESRMR tests be presented in addition to descriptive fit indices for model fit reporting.

8. Fitting Empirically Under-Identified Models: A Two-Factor Example
Dylan Boczar, Eric Loken
Abstract: This poster explores the consequences of empirical under-identification in situations where the model appears to be statistically identified. A CFA with two factors, and two indicators for each factor, is identified as long as the correlation between the factors is not zero. This study examined the estimation of this model as the population factor covariance approached zero. Fit statistics across decreasing simulated factor covariances demonstrate the impact of empirical under-identification. We document issues with model convergence, parameter estimation, and standard errors for data with population covariance near singularity. Even for models that converged, several issues attributable to empirical under-identification emerged: estimates of factor loadings exploded, Heywood cases emerged, and the standard error of the estimated factor covariance rose dramatically. We also explore the role of sample size, noting that large sample sizes are desirable when the population covariance is far from zero, but are problematic closer to zero. We conclude that—although model constraints that lead to unidentified models can be avoided—we cannot always avoid population data-generating mechanisms that lead to poor estimation.

9. Application of Planned Missingness Using 7 Datasets Without a Single Linking Test
Jacqueline M. Caemmerer, Timothy Z. Keith, Matthew R. Reynolds, Eunice Blemahdoo, Natalie Charamut
Abstract: Typically, missingness in planned missingness designs is intentional and under the control of the researcher. All participants complete the same linking test and a subset of other tests. It should be possible, however, to apply the same analytical techniques to incomplete data that were not "planned" to be missing within a single study, given the same missing data assumptions are correct. To explore this possibility, a combination of data from seven samples (n = 3,927 children aged 6-18) were simultaneously analyzed using confirmatory factor analysis and structural equation modeling. In each sample, Pearson Assessments collected data on pairs of intelligence tests or intelligence and achievement tests to evaluate their validity. The seven samples did not share the same linking test, but one or both tests in each sample were given in at least one other sample. This combined dataset was used to examine the applicability of Cattell-Horn-Carroll (CHC) intelligence theory across six intelligence tests. This comprehensive intelligence model was used to predict academic skills across three achievement tests.

10. A Novel Effect Size Measure for Mediation with a Multicategorical Predictor
Zihuan Cao, Heining Cham, Jordan Stiver, Monica Rivera-Mindt
Abstract: There are many limitations for current effect size measures to evaluate mediation when the predictor is nominal with three or more categories. Following the work by Lachowicz, Preacher, and Kelley (2017), a novel effect size measure υ was developed to address the issues. A simulation study was conducted to investigate the performance of its estimators. This simulation study assessed the performance of its estimators, manipulating various factors (group number, sample size, effect sizes, R^2 shrinkage estimators). Results indicate that the Olkin-Pratt extended adjusted R^2 estimator had the lowest bias and smallest MSE for estimating υ. The study concluded with a real data example and guidelines for using the recommended estimator.

11. Sharing is Not Normal: Evaluating the Statistical Methods for Analyzing Multimodal Dictator Game Data

Chen Chang, Sydney Y. Wood, Yue J. He

Abstract: The Dictator Game (DG) is a cross-discipline experimental design used to study altruistic behavior. In the traditional 2-person DG, one person, the dictator, decides how to divide an endowed resource, and the other person, the recipient, must accept what the dictator offers. A fundamental aspect of DGs is that there is no observable benefit to the dictator for sharing their resources nor a cost associated with sharing nothing. Through a systematic review of recently published DG studies ($N = 49$), we found that an overwhelming majority of DG researchers use central tendency comparisons that assume normal distributions. However, due to the zero-inflated multimodal distributions that emerge from the economically rational choice to share nothing and the tendency to converge on two or three theoretically informative values, we argue that the theoretical underpinnings of altruistic sharing lead to a complex data structure which makes the data particularly ill-fitted for methods meant for normally distributed outcomes. To illustrate this problem, we evaluate the statistical and theoretical appropriateness of each type of method found in the literature using four empirical datasets. In addition, we explore mixture modeling methods to compare differences between zero-inflated and multimodal distributions in DG studies.

12. IRT Standard Error Estimation in Item Modelling – A Mathematical Approach

Kamal Chawla

Abstract: Item Response Theory (IRT) models have received greater attention in recent decades. Models for one-way or multidimensional latent features for scoring items, ability estimation methods, and model calibration techniques have been developed. One aspect of item modeling that received far less attention in item response theory is the computation of standard errors when there is variance in item parameters. The current study on the 2-PL model focuses on calculating and comparing the mean standard error estimates in item modeling using a conventional method, bootstrapping, and a newly developed mathematical formula by introducing variance in item parameters.

13. Multilevel Latent Differential Structural Equation Model with Short Time Series and Time-Varying Covariates: A Comparison of Frequentist and Bayesian Estimators

Young Won Cho, Sy-Miin Chow, Christina M. Marini, Lynn M. Martire

Abstract: Continuous-time modeling using differential equations is a promising technique to model change processes with longitudinal data. Among ways to fit this model, the Latent Differential Structural Equation Modeling (LDSEM) approach defines latent derivative variables within a structural equation modeling (SEM) framework, thereby allowing researchers to leverage the advantages of the SEM framework for model building, estimation, inference, and comparison purposes. Still, the performance of multilevel variations of the LDSEM under short time lengths (e.g., 14-time points) remains inconclusive, particularly when coupled processes and time-varying covariates are involved. Additionally, the potentiality of using Bayesian estimation to facilitate the estimation of multilevel LDSEM (M-LDSEM) models with complex and higher-dimensional random effect structures has not been investigated. We present a series of Monte Carlo simulations to evaluate approaches to fitting M-LDSEM under several data configurations. Our findings suggest that the Bayesian approach outperformed the frequentist robust approach. Even with short time series, the Bayesian estimator was able to yield well-

recovered parameter estimation, particularly when higher-order derivative information was used. Specifically, the effects of time-varying covariates were well recovered, and the estimation of the indices of the coupled process was improved.

14. Examining Different Approaches to Treating Zero-Frequency Cells in Polychoric Correlation Estimation

Jeongwon Choi, Hao Wu

Abstract: When examining the correlation among ordered categorical variables, polychoric correlations are in fact more appropriate than Pearson correlations. When the contingency tables of data include zero-frequency cells, the polychoric correlation approaches ±1. This is problematic as it yields a singular or indefinite correlation matrix. Zero-frequency cells emerge when the sample size is insufficient for sampling the cell with a low probability or when a specific combination of variables cannot exist. Although the most commonly employed solution is to add a small value (e.g., 0.5) to zero-frequency cells before the estimation, this method is implemented inconsistently among researchers. There exist disagreements on determining the value to be added, to which cell these values will be added, and whether to maintain the marginal distribution. Therefore, to further previous simulation-based research (Savalei, 2011), this study comprehensively investigates the consequences of these manipulation methods in different dimensions under a variety of conditions by using the Monte Carlo simulation method. This study also provides guidance for future researchers to better estimate polychoric correlations and presents ways to improve computational efficiency for simulation studies by detecting conditions that yield equivalent estimates to avoid redundant computation.

15. Implementing an Expanded Mediation Model to Obtain Less Biased Estimates of Mediation Effects Implemented Using SEM

Julian M. Hernandez-Torres, Rafael R. Ramirez

Abstract: We illustrate an expanded mediation model that incorporates baseline data on both the mediator and outcome variable into the classic mediator model (Baron & Kenny, 1986). Using principles derived from instrumental variable estimation, the expanded model permits the identification and estimation of the correlation between residuals of the mediator and outcome variables. This correlation is assumed to be zero in the classical mediation model since it is empirically unidentified. As illustrated by Shrout (2011), if the mediator and outcome variable have moderate stability and are correlated at baseline, the correlation between residuals is bound to be substantial. Failure to incorporate this into the model can result in significant bias in estimates of indirect effects in mediation analysis. The model is illustrated using data from a randomized clinical trial with primary care patients ($N = 179$).

16. Establishing An Optimal Individualized Treatment Rule for Pediatric Anxiety with Longitudinal Modeling for Evaluation

Yifan Hu & Daniel Almirall

Abstract: This research applies longitudinal modeling as the methodological technique for the casual effect estimation of a novel Individualized Treatment Rule (ITR) for treating pediatric anxiety. ITR is a special case of a dynamic treatment regimen that inputs information about a patient and recommends a treatment based on this information. This research contributes to the literature by estimating an optimal ITR, which is said to maximizes a pre-specified outcome Pediatric Anxiety Rating Scale (PARS), to guide which of two common treatments to provide for

children/adolescents with separation anxiety disorder (SAD), generalized anxiety disorder (GAD) and social phobia (SOP): sertraline medication (SRT) or cognitive behavior therapy (CBT). We use data from the Child and Adolescent Anxiety Multimodal Study (CAMS). CAMS is a completed federally-funded, multi-site, randomized placebo-controlled trial, in which 488 children with anxiety disorders were randomized to cognitive-behavior therapy (CBT), sertraline (SRT), their combination (COMB), and pill placebo (PBO). There are four steps to the analysis: (1) Split the data for training (70%) and evaluation (30%) along with transforming and scaling the response PARS. In the training data set: (2) Prune the baseline covariates with patients' demographic information and historical clinical records according to their contribution levels to PARS with a specified variable screening algorithm for subset analysis; (3) Establish an interpretable and parsimonious ITR based on the screened covariates to guide clinicians on deciding personalized treatment plans for patients with pediatric anxiety disorders. Use the evaluation data to (4) Evaluate the effectiveness of ITR versus traditional treatments with only SRT, only CBT, and COMB for pediatric anxiety disorder with causal effect estimation based on comparisons of their clinical outcomes' longitudinal trajectories from Linear Mixed Effect Models. Our initial result is promising for two reasons: (1) Our final ITR is simple with only two most significant covariates, which should be feasible and easy-to-understand for clinicians in a real-world trial; (2) The longitudinal evaluation on ITR shows that it has a non-inferiority pattern compared with the best treatment COMB in the clinical history.

17. Detecting Cohort Effects in Accelerated Longitudinal Designs Using Multilevel Models

Simran Johal, Emilio Ferrer

Abstract: Accelerated longitudinal designs allow researchers to efficiently collect longitudinal data covering a time span much longer than the study duration, but assume that each cohort (a group defined by their age of entry into the study) shares the same longitudinal trajectory. Although previous research has examined the impact of violating this assumption when each cohort is defined by a single age of entry, it is possible that each cohort is defined by a range of ages, such as groups that experience a particular historical event. We examined how including cohort membership in linear and quadratic multilevel models performed in detecting and controlling for cohort effects in this scenario. We assessed performance with a Monte Carlo simulation study, varying the number of cohorts, the overlap between cohorts, the strength of the cohort effect, the number of affected parameters, and the sample size. Our results indicate that models including cohort membership accurately detected cohort effects and returned unbiased parameter estimates. Furthermore, using a proxy variable for cohort membership, based on age at study entry, performed comparably to using true cohort membership, indicating that researchers can control for cohort effects even when cohort membership is unknown.

18. Examining the Heterogeneity in the Effect of Head Start as a Multi-phase Treatment Accounting for Selection Bias and Clustering

Hanna Kim, Jee-Seon Kim

Abstract: When a treatment is available for multiple phases in longitudinal studies, individuals may participate with different patterns over time, which can become an important source of treatment effect heterogeneity. This study proposes to examine the heterogeneity in the effects of the national Head Start program on children's cognitive development using data from the Head Start Impact Study. Children in the 3-year-old cohort could attend Head Start for two years, resulting in four patterns of Head Start attendance at ages 3 to 4. Considering that Head

Start attendance was observed and not manipulated, generalized propensity score methods were applied to account for possible selection bias regarding the longitudinal attendance patterns. Specifically, multinomial logistic regression and generalized boosted modeling were adopted and compared as propensity score models. As substantial variation has been reported among Head Start centers, multilevel models, fixed effects models, and linear regression combined with cluster-robust standard error estimation were investigated as estimation models that address clustering while incorporating estimated propensity scores as model weights. Findings of this research not only provide practical information on Head Start administration strategies, but also illustrate methods for examining a relatively novel source of treatment effect heterogeneity in longitudinal research.

19. Drift Diffusion Modeling of Reaction Times and Accuracies on the Color Shapes Task
Sharon H. E. Kim, Yanling Li, Zita Oravecz
Abstract: The Color Shapes task (CST) is a visual working memory feature binding task that has been shown to be sensitive to cognitive health status and early risk for Alzheimer's Disease and Related Dementia. Assessments of CST capture both the accuracy and reaction time in task performance. In an experimental study, we manipulated different properties of the CST to study whether drift diffusion modeling (DDM) can account for individual differences in observed behaviors and latent decision-making processes. The DDM allows us to quantify theoretically meaningful cognitive parameters that underlie the observed behavioral data on cognitive tasks. Decomposing repeated CST performance measurements with such cognitive process models has the potential to reveal sensitive digital biomarkers of early-stage cognitive decline over daily life. We implemented a single-step Bayesian estimation of experimentally manipulated condition contrasts, latent DDM parameters, and their associations. We show how individual differences in experimental contrasts were related to person-level characteristics (e.g., age) in terms of cognitive (drift rate), non-cognitive (non-decision time), and meta-cognitive (boundary separation) parameters. The results highlight the benefits of using cognitive process models with optimized cognitive tasks to delineate features of cognitive aging from atrophy.

20. Assessing Multiple Imputation to Test Measurement Invariance with Ordinal Items
Hyunjung Lee, Danqi Zhu, Heining Cham, Chansoon (Danielle) Lee
Abstract: In confirmatory factor analysis, mean and variance adjusted weighted least square estimation (WLSMV) is recommended to handle ordinal data. Researchers investigated the use of multiple imputation in factor analysis to handle missing at random items. However, little is known about the multiple imputation inference to SEM with the ordinal items. This study extended previous studies to assess the performance of likelihood ratio test and global fit indices (RMSEA, TLI, CFI) with different pooling procedures with WLSMV in factorial invariance testing with ordinal items. Using Monte Carlo simulation, a two-group three-factor model with three ordinal indicators per factor was examined. The data were missing at random. Various conditions based on missing data rates, sample sizes, distributions of ordinal items, and magnitudes of non-invariance were tested. To examine factorial invariance, the multiple-group approach was used to test configural, metric, and scalar invariance. In each step, imputation-based versions of likelihood ratio test statistics and change of global fit indices were calculated to compare nested invariance models. We compared the results to those from complete cases. This study may provide guidelines about the interpretations of these fit indices in factorial invariance test with WLSMV estimation and multiple imputation.

21. Multi-level Structural Equation Modeling of Factors Influencing Mathematics Achievement

Hyunjung Lee, Chansoon (Danielle) Lee

Abstract: This study aims to examine influential factors on students' mathematics achievement using a multi-level structural equation modeling. The study used the U.S. data of the Trends in International Mathematics and Science Study (TIMSS) in 2019. The data included 1,112 8th grade students nested within 65 teachers, and the average number of students belonging to the teachers was 17. Although the mathematics achievement of students could be affected by various factors of both students and their teachers, there is a limited number of research that investigated the factors of both levels. In addition, most studies applied single-level statistical methods instead of multi-level methods, which ignored the nested data structure. In our analyses, the student-level factors include perceived difficulty, self-confidence, value for math, internet usage, bullying, and school climate, while the teacher-level factors include workload, job satisfaction, teacher characteristics, and teachers' perceived parental involvement. The results indicate that 73% of the total variance in mathematics achievement is explained at the teacher level. Parental involvement at the teacher level and perceived difficulty and value for math at the student level are statistically significant. The paper discusses the details of data characteristics and results, as well as implications.

22. Looking Beyond Fit Indices: A Bifactor Model Best Practice Study

Sijia (Carol) Li, Victoria Savalei

Abstract: The bifactor model has been rediscovered and achieved high popularity among psychologists in recent decades. However, methodologists have shown that bifactor models tend to overfit the data, so that traditional model fit indices are not sufficient for model evaluation. Instead, researchers should follow several recent recommendations for best practices when deciding whether to accept or reject the bifactor model (Bornovalova, Choate, Fatimach, Peterson, & Wiernik, 2020; Rodriguez, Reise, & Haviland, 2016; Watts, Poore, & Walderman, 2019). Most of the studies providing specific advice for evaluating bifactor models are in the context of applying this model to the general "p" (psychopathology) factor; no existing studies focus on reviewing published bifactor model applications across various areas of psychological research. In the current project, we were interested in systematically reviewing empirical applications of bifactor models across all areas of psychology. To limit the scope, we focused on articles published in 2020 from selected journals, for a sample of $N = 106$ articles. For each article, we recorded how the fit of the bifactor model was evaluated and which fit indices were reported, how reliability was evaluated, how model comparison was done, whether a detailed solution was reported, and so on. For articles reporting the detailed solution, we coded whether any of the general factor and group factor loadings were negative, and we computed several reliability coefficients. We present our results on the common pitfalls in the applications of bifactor models, and we provide a summary of best practice guidelines for psychologists to employ in future bifactor model applications.

23. Confidence Interval of Effect Size Measures in Longitudinal Growth Models

Zonggui Li, Ehri Ryu

Abstract: When modeling longitudinal data, multilevel modeling (MLM) is a commonly used framework. We previously developed full-model based effect size measures for longitudinal growth models (LGMs) in the MLM framework using R^2 (R-squared) (Li & Ryu, manuscript

under review). Though previous simulation study has been done to examine the bootstrap confidence intervals (CIs) for effect size in MLM (Lai, 2021), one limitation is that it mainly focused on Cohen's *d* instead of R^2, which has lower and upper bounds. In this study, we extend our previous work to provide CI computation for the developed effect size measures. We aim to examine two different bootstrapping methods (parametric and non-parametric residual based bootstrapping) along with different CI computation methods (normal CI, basic CI, percentile CI, and the bias-corrected and accelerated CI) for R^2, with the transformation in terms of raw R^2, log R^2, and logit R^2 using simulated data in LGMs.

24. Prior Achievement, Future Learning Opportunities, and Student Achievement

Jiachen Liu

Abstract: Studies of achievement growth have help evaluate school efficiency and accountability. However, besides serving as the control variable, student prior achievement was never examined further. This study examined three different roles student prior achievement could play in achievement growth: (1) the symbolization of baseline achievement level, (2) a representative of students' ability to transfer learning opportunities into future achievement, (3) a decisive factor of student opportunity to learn (OTL) in later periods. Simulation datasets used in this study have their origin in the USA dataset of the Second International Mathematics Study (SIMS-USA), which provided the data structure needed for this study. Results of this study indicate that the baseline achievement effect and the decisive effect on learning opportunities both exist. However, the moderating effect of prior achievement on the relationship between student OTL and achievement in the next period is only statistically significant at the classroom level. This lack of evidence for the effects of prior achievement at the individual level could be attributed to the cross-level independence structure of the variance-covariance matrices from which the datasets are simulated.

25. An Evaluation of the Bayesian Hypothesis Test of Mediation Effects via the Bayes Factor

Xiao Liu, Zhiyong Zhang, Lijuang Wang

Abstract: Bayesian approaches have been increasingly used in mediation analysis. Recently, a Bayes factor (BF) based Bayesian hypothesis test for the presence of mediation has been developed. The BF for mediation provides a promising complement to frequentist null hypothesis tests of mediation. For example, the BF can take into account researchers' prior knowledge and quantify data evidence of the absence of mediation. Despite the appealing features, performance of the BF for testing mediation has not been examined. In this study, we examine the sample size requirements for the BF to have a high probability of correctly supporting the presence of mediation while maintaining a low chance of supporting the presence of mediation incorrectly. We also develop easy-to-use BF-based sample size planning tools, including R functions and a web application, to facilitate researchers planning sample sizes for future mediation studies using the mediation BF. Our study provides insights into the performance of the BF for testing the presence of mediation effects and adds to researchers' toolbox for sample size planning of mediation studies.

26. The Role of Cherry Picking in the Replication Crisis

Xinran Liu, Samantha F. Anderson

Abstract: Low replication success rates and high-profile replication failures have threatened to undermine trust in social science research. Questionable research practices (QRPs) have been a contributor to this crisis as they can severely increase the likelihood of false positive results. "Cherry picking" is a type of QRP in which a researcher collects data on multiple versions of a variable but reports only the version with the strongest possible support for their hypothesis. In a Monte Carlo simulation study, we investigated the consequences of cherry picking on factors relevant to replication: original study false positive rates, original study effect size bias, and statistical power in replication studies based on the affected original studies. Results indicated (1) inflated false positive rates, (2) small to moderate effect size bias that could be large when viewed relative to the population effect size, and (3) that (1) and (2) can impact the statistical power of replication studies. Bias increased with the severity of cherry picking, effect size, and sample size, but decreased as the correlation among dependent variables increased. Our results imply that cherry picking contributes to problems with replication. We need to elevate our standards for good research practices and appropriate data analysis procedures.

27. Evaluating the Multiplicity Control Practices of Psychology Researchers

Naomi Martinez Gutierrez, Jordana DeSouza, Udi Alter, Nataly Beribisky, Linda Farmus, Robert Cribbie

Abstract: Whenever researchers test multiple hypotheses, the risk that one or more hypotheses might be falsely supported (i.e., Type I errors) increases with the number of hypotheses evaluated (i.e., the multiplicity problem). Because most studies evaluate multiple hypotheses, the risk that at least one hypothesis is a Type I error appears substantial. However, it is necessary to evaluate how consistently multiplicity control (MC) is applied and the rationale behind its application, to understand the merit of MC in psychological research. We conducted a systematic review of MC practices in 250 articles from 10 high impact journals. In addition to descriptive statistics (e.g., confirmatory/exploratory), we coded: 1) total hypotheses tested; 2) analyses family (e.g., mean comparisons); 3) focus on estimation or significance testing; 4) presence or absence of MC; 5) if applicable, the nature of the MC; and 6) the rationale for the type of MC. The median number of hypotheses tested per article was 79; however, only 8% of all hypotheses were protected by MC. The most popular MC type was familywise, via the Bonferroni method. These results highlight the wide range of situations wherein multiplicity occurs, the inconsistency with which MC is applied, and the lack of rationale for MC decisions.

28. The Power of Smaller Samples: Developing Age-Stratified Norms From Randomly Selected Sub-Samples

Megan Mulvihill

Abstract: Publishers of educational and psychological tests rely on large, representative samples of research participants for scale and normative score development. However, in recent years there has been an overall decrease in research participation and the cost of recruiting research participants has steadily risen. As a result, test publishers are exploring the use of smaller standardization samples during the test development process. Using explanatory IRT models and psychological test data from a nationally representative sample of participants, this study produces age-stratified normative score distributions from randomly selected sub-samples made up of 80%, 60%, 40%, 20%, and 10% of the original sample. The resulting score distributions,

model fit, and statistical power are discussed for each sample size and compared to the original score distributions of the full sample. The results of this project should help test developers select sample sizes that reduce overall recruitment costs while maintaining psychometric integrity of published assessments.

29. Is There a Time to Be Discrete? Comparisons of Difference Score and Autoregressive Models

Ascher Munion

Abstract: Both autoregressive and difference score models have a long history of utilization within time-series modeling. These models have been used to characterize stability effects as well as bi-directional associations in dyadic analyses, such as in the cross-lagged panel model and actor-partner interdependence model. Historically, there has been contention as to when autoregressive or difference scores will produce more bias. However, when both the autoregressive and difference score models are contextualized within the characterization of temporal patterns implied by Dynamical Systems Theory (DST), we see that both can characterize the same temporal patterns. First, we demonstrate that both models describe common first-order dynamics (e.g., attractors and repellers). Then, using a series of simulations, we produce commonly hypothesized temporal patterns and data properties (e.g., single-level and multilevel data, long and short time series, measurement and transient error, and unidirectional and bidirectional dyadic associations). Finally, we demonstrate that a variety of common modeling applications of autoregressive and change score models produce equivalent results as indexed by coefficients that are transformations of each other, and identical standard errors.

30. Within-Person Analysis of Developmental Cascades of Externalizing Problems, Academic Competence, and Internalizing Problems from Kindergarten to Fifth Grade: Evidence from Population-Based Sample of US Elementary School Students

Yoonkyung Oh, Paul Morgan, Gabriella Keller

Abstract: Externalizing (e.g., aggression) and internalizing problems (e.g., anxiety) are increasingly prevalent among children. Although classified as distinct domains of behavior problems, externalizing and internalizing problems often co-occur. Academic competence has been proposed as a key mechanism through which one domain of behavior problems could lead to the other domain. We tested this developmental cascade model involving academic competence as a mediator linking externalizing and internalizing problems across the elementary school grades. We analyzed the data drawn from the ECLS-K: 2011, a longitudinal study of a nationally representative sample of U.S. kindergarteners followed through Grade 5. We used teacher-reported externalizing and internalizing problems and direct assessments of reading and math achievement that were measured annually from kindergarten through Grade 5. We used Latent Curve Models with structured residuals (LCM-SR) to evaluate developmental links among externalizing problems, academic achievement, and internalizing problems at the within-person level, after partialling out the systematic between-person components. We found evidence for a cascade from externalizing to internalizing problems via reading difficulties in the earliest grades. However, we found no evidence for a mediating role of math competence in developmental cascades between externalizing and internalizing problems. Instead, we found evidence for developmental cascades where lack of math competence promotes internalizing problems in a subsequent grade, which in turn lead to poorer math outcomes in later grades.

31. Bayesian Factor Analysis with Regularizing Priors on Threshold Parameters

R. Noah Padgett

Abstract: Discrete-ordered response options are commonly utilized in social science, and Bayesian factor analysis provides a flexible approach to modeling multivariate survey data. However, when some response categories are infrequently endorsed, a computational issue arises due to insufficient information to update the prior. This issue is multiplied in multiple-group measurement models when one group has no responses to an option. In this study, we evaluated the use of a joint prior on item threshold parameters to overcome the limitation of setting a prior on individual threshold parameters. A measurement invariance analysis is illustrated when groups have items with a different number of response options with endorsement, and the effects of threshold prior specification are shown. The results of a Monte Carlo simulation study evaluating the parameter recovery of the prior specifications is summarized.

32. An Evaluation of Methods for Handling Missing Data in Randomized Controlled Trials with Omitted Moderation Effects

Elizabeth Pauley, Manshu Yang

Abstract: Randomized Control Trials (RCTs) are widely used in behavioral and health-related studies to compare the effectiveness of intervention strategies; however, missing data in RCTs are almost inevitable. In many RCT studies, the key focus is to examine the average treatment effect (ATE) within an entire population. Heterogenous treatment effects, often related to moderation effects of baseline personal attributes, do not typically get included in analyses. To handle missing data in RCTs, multiple imputation (MI) or inverse probability weighting (IPW) could be used. MI, although often preferred over IPW, may lead to biased ATE results when the probability of missingness depends on a moderator and the moderation effect is omitted from analyses. In contrast, IPW may produce less biased but imprecise results when the sample size is small. This study aims to evaluate the performance of MI via joint modeling, MI via chained equations, and IPW in estimating ATE in RCTS with missing data and omitted moderation effects. A Monte Carlo simulation study is conducted to compare methods under various scenarios. It is expected that IPW and MI via chained equations will outperform MI via joint modeling given a larger sample size and a small number of strong moderators.

33. Anchor Item Requirements for the Free Baseline DIF Approach in Thurstonian-IRT

Jake Plantz, Jessica Kay Flake

Abstract: The forced-choice (FC) response format is an approach for limiting response biases in survey- research. FC blocks consist of two to four items which the respondent orders from most to least like them. The item responses in each block are dependent upon each other violating the local independence assumption in item response theory (IRT). Thurstonian-IRT incorporates these dependencies into estimation, facilitating the use and scoring of FC tests. However, only one method has been developed to test differential item functioning (DIF) in Thurstonian IRT models: the free-baseline approach (Lee & Smith, 2022). This study has two goals: to apply the free baseline method to a real high stakes test and then conduct a simulation study to assess the limitations and assumptions of this new method. Preliminary results using the free baseline method exhibit high rates of nonconvergence and suggest that identifying the appropriate number of anchor blocks in real data is a challenge. Thus our simulation study focuses on how many anchor blocks are necessary and ways of identifying them. We extend previous work on the method by simulating realistic testing data with various levels of anchor blocks. The goal of the

study is to determine when/if the free baseline method is appropriate for use with real testing data and to recommend best steps for using the method.

34. Investigating the Performance of the Bi-Factor AESEM with Dichotomous Indicators Under Conditions of Measurement Non-Invariance

Qingzhou Shi, Joni Lakin, Chunhua Cao, Stefanie Wind

Abstract: The alignment method was developed by Asparouhov and Muthén (2014) and extensively researched as an effective alternative to the multi-group CFA (MGCFA) for estimating the means and variances of group-specific factors under the assumption of approximate measurement invariance. The alignment method has recently extended to the general structural equation models (SEMs), allowing for its application to models with covariates and cross-loadings, as well as bi-factor models. This Monte Carlo simulation study investigates the performance of the alignment exploratory SEM (AESEM) with bi-factor models and dichotomous indicators under various conditions based on group size, the magnitude of non-invariance, item and group non-invariance rates, number of covariates, and the proportion of indicators with cross-loadings in each factor. Results examine whether the AESEM adequately recovers parameter estimates when there are small, moderate, and large magnitudes/proportions of non-invariance in item parameters and groups. Based on these findings, recommendations for implementing the AESEM are also included.

35. Deep Learning Imputation for Unbalanced and Incomplete Likert-type Items

Olushola Soyoye, Kamal Chawla, Zachary Collier, Minji Kong, Yasser Payne, Ann Aviles

Abstract: Unbalanced Likert-type items are skewed-scaled with either no neutral response option or an uneven number of possible favorable and unfavorable responses. Modern missing data methods may be problematic when respondents do not answer unbalanced items because they assume multivariate normality. Alternatively, list-wise deletion and mean imputation assume that data are missing completely at random, which is often unlikely in surveys and rating scales. This article explores the potential of implementing a scalable deep learning-based imputation method. Additionally, we provide access to deep learning-based imputation to a broader group of researchers without requiring advanced machine learning training. We apply the methodology to the Wilmington Street Participatory Action Research (PAR) Health Project.

36. Closing the Loop Between Educational Data Mining and Educational Measurement and Evaluation

Zachary Collier, Joshua Sukumar, Roghayeh Barmaki

Abstract: This poster introduces researchers in the science concerned with developing and studying research methods, measurement, and evaluation (RMME) to the educational data mining (EDM) community. It assumes familiarity with traditional priorities of statistical analyses, such as accurately estimating model parameters and inferences from those models. Instead, we focus on data mining's adoption of statistics and machine learning to produce cutting-edge methods in educational contexts. It answers three questions: (1) What are the primary interests of EDM and RMME researchers? (2) What is their discipline-specific vocabulary? and (3) How is data mining applied differently from traditional statistical approaches in educational research?

37. Addressing Potential Model Misspecification: A Novel Application of the Latent-Curve Model with Structured Residuals to Well-Being in Justice-Involved Youth

Jennifer M. Traver

Abstract: Many hypotheses regarding well-being are inherently related to within-person changes, such as understanding mechanisms that increase an individual's level of well-being. Although recent work in this area has used methods well suited to disentangling within- and between-person effects (e.g., Vaughan et al., 2021), this often comes at the price of introducing model misspecification. For example, it is known that there are bidirectional relationships between an individual and their parents or peers, yet reciprocal effects are rarely considered while simultaneously assessing within-person effects. This may result in biased parameter estimates. To address this concern, the current study used latent-curve models with structured residuals (LCM-SR; Curran et al., 2014) to simultaneously model within-person effects and reciprocal relations between parental or peer warmth and well-being in justice-involved youth. Participants (N = 1,216) were male offenders aged 13-17, and measures were collected annually across six years. Results indicate that when parental or peer warmth was higher than expected given an individual's trajectory, their subsequent well-being was higher than expected. Further, a reciprocal relationship between these constructs was confirmed. This novel application of the LCM-SR demonstrates the potential utility of modeling within-person effects and reciprocal relationships simultaneously, which has many practical applications to resilience research.

38. Bayesian Workflow for Model Building in Educational Research

M. Shane Tutwiler

Abstract: Educational data often contain a large degree of measurement error, missingness, and complexity due to the clustering of students. Traditional statistical inference via frequentist models and null hypothesis significance testing, while computationally efficient and relatively easy to interpret, fail to address these issues directly and often mask the degree of uncertainty in findings by focusing on point estimates and p-values. In this poster, we demonstrate a proposed workflow for modeling and communicating relationships in educational data via Bayesian multilevel modeling. Benefits include the ability to encode domain knowledge via informative priors, using said priors to regularize estimates to prevent the over-fitting of models to data, inherent mechanisms to impute missing data, and the capacity to handle measurement error in the outcome and predictors.

39. Patterns of Current Tobacco Use among Adults in US and Israel and the Correlates

Yan Wang, Zongshuan Duan, Yuxian Cui, Carla J. Berg

Abstract: Globally, tobacco products have diversified dramatically, but limited research has examined use patterns, which are critical to understand tobacco industry marketing and collective population impact. We conducted latent class analysis of 2021 survey data among US (n = 1,128) and Israeli adults (n = 1,094) assessing past-month use of various tobacco products (i.e., cigarettes, e-cigarettes, heated tobacco products, hookah, cigars, pipe, smokeless) and examined class correlates (sociodemographics, diffusion of innovation profile, i.e., innovator/early adopter vs. later adopter). Three classes were identified: low-use (infrequent/no use, n = 1673, 75%), traditional-use (of cigarettes, e-cigarettes, and cigars, n = 401, 18%), and high-use (of various products, n = 147, 6.6%). In multivariable multinomial regression (ref: low-use), living in US and being male correlated with traditional-use and high-use, being married correlated with traditional-use, and being sexual minority correlated with high-use. Being an

innovator/early adopter correlated with traditional-use and high-use (aOR = 1.12, 95% CI: 1.08-1.15, $p = 0.046$, aOR =1.19, 95% CI: 1.13-1.25, $p < 0.001$); being a later adopter negatively correlated with high-use (vs. low-use; aOR = 0.95, 95% CI: 0.92-0.98, $p = 0.002$).

40. Derivative Estimation to Quantify Pain Variability in Chronic Pain Patients

Mirinda Whitaker, Pascal R. Deboeck, Akiko Okifuji

Abstract: Pain is a highly variable and dynamic experience. Yet, most approaches in pain research treat pain as if it were static, with any intraindividual variability averaged out, or treated as noise/error. This is consequential because across multiple chronic pain conditions, higher pain variability has been associated with a variety of negative outcomes. In the current analysis, we used data from an experiment where participants ($N = 54$; 30 chronic-pain patients and 24 pain-free controls) continuously rated the intensity of a painful thermal stimulus for 30 seconds. We estimated first and second derivatives to examine change variance, in addition to observed score variance in pain scores. Overall, chronic pain patients had higher variability in their pain ratings (SD of 0th: $\beta = .28$, $p = .04$; SD of 1st: $\beta = .27$, $p = .05$; SD of 2nd: $\beta = .17$, $p = .23$). The current application suggests one way of quantifying pain variability and highlights the need for further application and development of statistical methods to quantify pain variability.

41. Item Quality for Cognitive Assessments in Low-to-Middle Income Countries: Evidence from the Ethiopia Young Lives Data.

Winifred Graham Wilberforce, Ann A. O'Connell

Abstract: This study reports on an analysis of item quality for the Math and English assessment of the Young Lives survey in Ethiopia. The primary goal of the Young Lives International Study is to develop understanding of the contribution of educational experiences in relation to the causes and consequences of child poverty. Rasch Measurement approaches are used to explore the cultural and gender relevance of items on the cognitive assessments. Results of this study show that item targeting can be improved by introducing more difficult items that target highly proficient students. There is also evidence of location (urban versus rural) differential item functioning (DIF), suggesting that some items may be more difficult for students in rural schools than those in urban schools with similar ability levels. These biases are observed in both Math and English test items but are more prevalent in the English items.

42. Does Self-Motivation Affect Academic Performance among Middle Schoolers in Ethiopia? A Partial Latent Mediation Analysis.

Winifred Graham Wilberforce, Krisann Stephany

Abstract: Childhood poverty levels affects academic outcomes in the classroom. Existing literature suggests several factors that might explain this association and self-motivation is one. Most of the research in this area is conducted in High-Income countries such as the U.S and U.K. but these important relationships are not established in low-to-middle income countries like Ethiopia. The current study explored self-motivation as a potential mediator between a child's poverty level (hours of work on the farm, chores at home and work for pay on school day) and their academic outcomes (Math and English). We were also interested in how the gender of the child was affected by this mediating effect on academic achievement. A partial latent mediation analysis was used to model the relationships using the Wave 1 and 2 data from the International Young Lives survey for Ethiopia. It was completed by a total of $N = 12,182$ children from 63 schools. We found that self-motivation mediates the effect for students with lower poverty than

for those with higher poverty levels as well as the effects for female students than male students. These findings suggest that self-motivation partly explains the association between poverty, gender, and academic outcomes for middle school children in Ethiopia.

43. Exploring Individual Differences in Continuous Time Dynamics of Loving Feelings in Daily Life

Lindy Williams, Sharon H. E. Kim, Zita Oravecz

Abstract: It is hypothesized that feelings of love, both expressed and felt, change dynamically over time, much like affective feelings. We explored the links between momentary states of felt love and expressed love with intensive longitudinal data collected in daily life settings. The associations between feelings of love and expressions of love were analyzed using a multilevel Bayesian Ornstein-Uhlenbeck (OU) model. With this approach, we can model changes in felt and expressed love in a continuous time framework. The approach allowed us to study whether increases in expressed love were followed by increases in felt love reported and vice versa. In addition, we explored sources of individual differences (i.e., person-level predictors such as gender, relationship status etc.) for the return to baseline for felt and expressed love and the inertia back to the attractor state, as well as intra-individual variability and contemporaneous correlations. We demonstrate how the Bayesian framework allows for all analyses to be done simultaneously, with the latent OU parameters estimated and regressed on the person-level predictors in one-step.

44. Detrending Multi-Subject, Short Time Series Data for Vector Autoregressive and Dynamic Structural Equation Models

Xiaoyue Xiong, Sy-Miin Chow, Yanling Li

Abstract: In time series analysis, trends capture systematic variations that unfold over slower time scales than other more nuanced, "momentary" patterns of intraindividual variability to be represented by standard time series models such as autoregressive models. Trends can take various forms, including monotonic linear and nonlinear (e.g., sigmoid and logistic) and non-monotonic change functions. Removing trends, a process known as detrending, is critical to any time series analysis, the omission of which would lead to violations of the stationarity and related assumptions, and in turn, biases in parameter estimates. Even though many detrending methods have been proposed in the time series context involving individual-level, the tenability, strengths, and limitations of these methods when extended to multi-subject time series data with relatively few measurement occasions have rarely been studied. In the current study, we evaluate the impact of several single- and two-stage detrending methods on results from fitting vector autoregressive (VAR) and the related Dynamic Structural Equation Modeling (DSEM) models. The results of this study will provide insights into the impact of detrending on the estimation of VAR and DSEM models, and help researchers make informed decisions concerning the appropriate detrending strategies for their data.

45. Impact of Missing Days in Daily Diary Studies: A Simulation Study

Mustafa Yildiz, Maria Yefimova, Christopher D. Maxwell, Tami Sullivan, Carolyn E. Pickering

Abstract: Missing data studies in the past are dominated by cross-sectional data structures. In diary studies, in addition to the missing data to a particular item, an entire day could be missing. This simulation study investigated daily diary datasets with missing days via a simulation study in which a binary outcome variable was predicted using generalized linear mixed models. The

following aspects of diary studies were manipulated: number of individuals (30 to 400), number of diaries (7 to 90), proportion of missing data (from 0.05 to 0.50), effect size (small or large), and the nature of missing diaries (random, model based, and covariate predicted). In this fully crossed simulation design, 40 simulation replications were run for each simulation condition. The simulation study was evaluated for the following quantities: root-mean squared error and bias for the intra-class correlation coefficient and the fixed effect parameters. In addition, the convergence and the power were investigated. Results indicated that convergence was always a problem when a small number of days (7 days) was used even with the largest sample sizes. In RMSE and bias of ICC conditions, smaller effect size conditions with fewer number of days had the worst performance. Underestimation of ICC may be to 0.25, depending on proportion of missing days.

46. The Effect of Model Size on the Root Mean Square Error of Approximation (RMSEA): The Nonnormal Case

Yunhang Yin, Dexin Shi, Amanda J. Farichild

Abstract: This study aimed to understand the effect of model size on the root mean square error of approximation (RMSEA) under nonnormal data. We considered three methods for computing the sample RMSEA and the associated confidence intervals (CIs; i.e., the normal theory method, the BSL method, and the Lai method). The performance of the three methods was compared across various model sizes, sample sizes, levels of misspecification, and levels of nonnormality. Results indicated that the normal theory RMSEA should not be used under nonnormal data unless the model size is very small. In the presence of nonnormal data, researchers should consider using either the BSL or the Lai method to estimate RMSEA and its CIs. The Lai method is recommended when very large models are fit under nonnormal data.

47. Using RMSEA Associated with the Chi-Square Difference Test to Compare Bifactor and Hierarchical Factor Models under Minor Misspecification

Paradox Zhou, Victoria Savalei

Abstract: The bifactor model (BFM) has been widely used in psychology. Researchers commonly use fit indices to compare BFM with its nested models, such as the higher-order factor model (HFM). However, studies have shown that fit indices tend to indicate a better fit of the BFM relative to the HFM, even when HFM is the data-generating model (Greene et al., 2019; Morgan et al., 2015; Murray & Johnson, 2013). The superior model fit of the BFM has been described as a fit index "bias" rather than an indication of model correctness. However, the dominant approach of comparing fit indices across models is not ideal for nested models. Because the less restricted model will always fit better, an index that focuses on the degree of deterioration in fit is preferred. Focusing on the root mean square error of approximation (RMSEA), we argue that using the difference between model RMSEAs (i.e., computing ΔRMSEA or simply eyeballing the RMSEA values across models) is a problematic approach. Instead, we advocate the fit index RMSEA_D (Savalei, Brace, & Fouladi, 2022), an RMSEA associated with the chi-square difference test, which captures the deterioration of fit per added degree of freedom in the constrained model (HFM) relative to the original model (BFM) on the familiar RMSEA metric. We report on the results of a large-scale simulation study that systematically investigated the performance of the RMSEA_D when the true model is HFM, either correctly specified or containing varying degrees of misspecification. We report under which conditions and degrees of underlying misspecification RMSEA_D were able to retain the

otherwise correct constraints in the HFM and which levels of misspecification were too great that led to the rejection of HFM. We also compute and evaluate the performance of a confidence interval for RMSEA_D.

48. Compliance and Outcomes in the Toddler Obesity Prevention Study (TOPS), a Behavioral Intervention to Improve Physical Activity

Shijun Zhu, Yan Wang, Erika Friedmann, Maureen Black

Abstract: Intervention efficacy is thought to be dependent on attendance. In many trials, participants' attendance varies and is unreported. We examined factors related to attendance and the relationship between attendance compliance and the effect of a behavioral intervention on promoting physical activity among mother-toddler dyads. 277 mother-toddler dyads were randomized into three arms [Mom-TOPS (maternal lifestyle), Tot-TOPS (responsive parenting), and Safe-TOPS (attention control)] in a four-week, eight-session RCT. Poisson regression assessed sample characteristics associated with attendance. Complier average causal effects from a growth mixture modeling framework assessed intervention effects on moderate-to-vigorous physical activity (MVPA) and body mass index, separately, with compliance defined as attendance > 60% (≥ 5) sessions. Mean age was 27.3 (SD = 6.2) years for mothers and 20.1 (SD = 5.5) months for toddlers. Higher maternal education, older age, being married, fewer depressive symptoms, and lower BMI were related to greater attendance in Tot-TOPS; higher education was related to greater attendance in Mom-TOPS. Among compliers, intervention increased MVPA (mothers: $b = 11.36$, SE = 4.89, $p = 0.020$, toddlers: $b = 67.03$, SE = 18.25, $p < 0.001$ in Mom-TOPS; and toddlers: $b = 77.81$, SE = 19.73, $p < .001$ in Tot-TOPS). Intervention attendance is critical to promote MVPA among toddlers and their mothers. Future trials need to incorporate strategies to promote attendance in behavioral interventions.

49. Integrating Fuzziness into Vector Autoregressive (VAR)-Based Clustering

Jonathan Park, Sy-Miin Chow, Zachary Fisher, Peter C. M. Molenaar

Abstract: The past decade has seen a plethora of methods seeking to bridge the gap between idiographic and nomothetic inference. These methods, dubbed idio-thetic approaches, operate by identifying group-level commonalities in person-specific dynamics and largely make use of community detection approaches from the network science literature. These idio-thetic methods can help to reduce bias by placing subjects into subgroups which align more closely to their individual dynamics while still yielding valuable group-level models for generalizability and inference. However, many (if not all) of these approaches assume subjects are discretely packaged into one group or another (A or B). This assumption may not be tenable as subjects may belong to multiple communities, exist between them, or belong to none altogether. A failure to account for these *fuzzy* subjects may result in bias to the group- and subgroup-level models, and mask important details regarding individuals who manifest features associated with more than one subgroup. Broadly, fuzzy community detection assigns weights that quantify the degree to which a subject belongs to any k-clusters. Thus, fuzzy approaches allow researchers to quantify the strength of, or confidence in, each individual's group membership, and identify ambiguous cases for individual modeling. To address these points, we integrate fuzzy community detection into vector autoregressive (VAR)-based clustering, and present results from a Monte Carlo simulation to elucidate the strengths of the proposed approach relative to conventional hard clustering methods. Further implications and guidelines for using fuzzy clustering in VAR-based clustering in empirical contexts will be discussed.

2024 Modern Modeling Methods Conference: Call for Proposals

The Modern Modeling Methods (M^3) conference is an interdisciplinary conference designed to showcase the latest modeling methods and to present research related to these methodologies. The 10th M^3 conference will be held June 25-26, 2024 at the University of Connecticut, preceded by an all-day pre-conference workshop on June 24, 2024. Confirmed keynote speakers for the 2024 conference include Denny Borsboom.

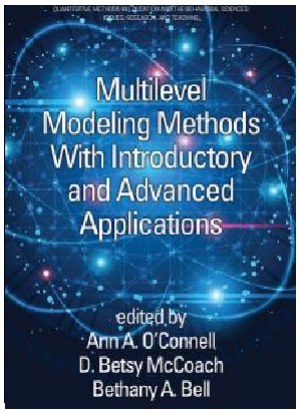
Submissions for the 2024 conference are due 1/29/24. We welcome both methodological research papers and papers that illustrate novel modeling applications of modeling. Papers related to latent variable modeling, multilevel modeling, mixture modeling, longitudinal modeling, dynamic modeling, and psychometric modeling are especially encouraged. Given the interdisciplinary focus of the conference, it is completely acceptable to present papers that have been published or presented elsewhere. Presenters may select the length of the session that they prefer: 30 minutes, 60 minutes, or 90 minutes. We also welcome proposals for multi-paper symposia on thematically grouped topics. Generally, symposia sessions are 90 minutes in length. We are also soliciting proposals for the poster session. Students are also encouraged to submit proposals, especially for the poster session.

Conference proposals for the Modern Modeling Methods conference may fall into one (or more) of four categories: Methodological Innovation, Methodological Application, Methodological Illustration, or Methodological Evaluation. Methodological Innovation proposals introduce a new technique. Methodological Evaluation proposals present the results of empirical research evaluating a methodology. Most often, these will involve simulation studies. Methodological Application proposals present the methods and results of a real research study in which the technique was used. Methodological Illustration proposals provide a pedagogical illustration of when and how to use the technique; these papers are designed to help the audience be able to implement the technique themselves.

There are three different types of presentations: Paper sessions (in which authors submit a single paper), Symposia (in which a group of authors submit a set of related talks/papers), and posters. All papers should include a 150-200 word abstract that will appear in the conference program. Methodological Research paper proposals should be no longer than 1000 words and should include purpose, background, methods, results, discussion, and significance. Methodological Illustration paper proposals should be no longer than 1,000 words and should include a description of the methodology to be illustrated as well as an outline of the paper/talk. Proposals for symposia should include titles, authors, an abstract for the symposium, and brief descriptions/abstracts for all the paper presentations within the symposium. Symposium proposals may be longer than 1000 words if needed, but they should be less than 2000 words. Proposals for the poster session need only submit an abstract: the 1000-word proposal is ***not*** required for poster session proposals.

Proposals for the 2024 conference are due January 29, 2024. Notifications of presentation status will be emailed by February 20th, 2024. For more information about the 2024 Modern Modeling Methods conference and/or to submit a proposal, please visit http://www.modeling.uconn.edu/ .

www.infoagepub.com



Multilevel Modeling Methods with Introductory and Advanced Applications

Edited by **Ann A. O'Connell,** *The Ohio State University*
D. Betsy McCoach, *University of Connecticut*
and **Bethany A. Bell,** *University of Virginia*

A volume in **Quantitative Methods in Education and the Behavioral Sciences: Issues, Research, and Teaching**
Series Editor Jeffrey R. Harring, University of Maryland

Multilevel Modeling Methods with Introductory and Advanced Applications provides a cogent and comprehensive introduction to the area of multilevel modeling for methodological and applied researchers as well as advanced graduate students. The book is designed to be able to serve as a textbook for a one or two semester course in multilevel modeling. The topics of the seventeen chapters range from basic to advanced, yet each chapter is designed to be able to stand alone as an instructional unit on its respective topic, with an emphasis on application and interpretation.

In addition to covering foundational topics on the use of multilevel models for organizational and longitudinal research, the book includes chapters on more advanced extensions and applications, such as cross-classified random effects models, non-linear growth models, mixed effects location scale models, logistic, ordinal, and Poisson models, and multilevel mediation. In addition, the volume includes chapters addressing some of the most important design and analytic issues including missing data, power analyses, causal inference, model fit, and measurement issues. Finally, the volume includes chapters addressing special topics such as using large-scale complex sample datasets, and reporting the results of multilevel designs.

Publication Date: 2022

ISBNs:

Paperback: 978-1-64802-871-7
Hardcover: 978-1-64802-872-4
E-Book: 978-1-64802-873-1

Paperback: $ 65.99
Hardcover: $ 95.99

Trim Size: 6.14 X 9.21

Page Count: 644

Subject: Education, Statistics, Mathematics, Longitudinal Research, Multilevel Modeling

BIC Code: PBW

BISAC Codes:

EDU029010
EDU030000
EDU011000

Each chapter contains a section called *Try This!*, which poses a structured data problem for the reader. We have linked our book to a website (http://modeling.uconn.edu) containing data for the *Try This!* section, creating an opportunity for readers to learn by doing. The inclusion of the *Try This!* problems, data, and sample code eases the burden for instructors, who must continually search for class examples and homework problems. In addition, each chapter provides recommendations for additional methodological and applied readings.

CONTENTS: Acknowledgments. Introduction to Multilevel Modeling Methods: Pedagogy and Context, *Ann A. O'Connell, D. Betsy McCoach, and Bethany A. Bell.* **SECTION I: ORGANIZATIONAL DATA.** Introduction to Multilevel Models for Organizational Data, *Bethany A. Bell and Jason A. Schoeneberger.* Evaluation of Model Fit and Adequacy, *D. Betsy McCoach, Sarah D. Newton, Anthony J. Gambino.* Causal Inference in Multilevel Settings, *Chris Rhoads and Eva Yujia Li.* Statistical Power for Linear Multilevel Models, *Jessaca Spybrook, Benjamin M. Kelcey, and Nianbo Dong.* Cross-Classified Random-Effects Models, *Audrey J. Leroux and S. Natasha Beretvas.* Multilevel Logistic and Ordinal Models, *Ann A. O'Connell, Meng-Ting Lo, Jessica Goldstein, H. Jane Rogers, and C.-Y. Joanne Peng.* Single and Multilevel Models for Counts, *Ann A. O'Connell, Nivedita Bhaktha, and Jing Zhang.* **SECTION II: LONGITUDINAL DATA.** Individual Growth Curve Models for Longitudinal Data, *D. Betsy McCoach, Bethany A. Bell, and Aarti P. Bellara.* Modeling Nonlinear Longitudinal Change With Mixed Effects Models, *Jeffrey R. Harring and Shelley A. Blozis.* Within-Subject Residual Variance–Covariance Structures in Longitudinal Data Analysis, *Minjung Kim, Hsien-Yuan Hsu, and Oi-man Kwok.* Modeling Variation in Intensive Longitudinal Data, *Donald Hedeker and Robin J. Mermelstein.* **SECTION III: DESIGN AND SPECIAL ISSUES.** Using Large-Scale Complex Sample Datasets in Multilevel Modeling, *Laura M. Stapleton and Scott L. Thomas.* Common Measurement Issues in a Multilevel Framework, *Brian F. French, W. Holmes Finch, and Thao Vo.* Missing Data Handling for Multilevel Data, *Craig K. Enders and Timothy Hayes.* Multilevel Mediation Analysis, *Nicholas J. Rockwood and Andrew F. Hayes.* Reporting Results of Multilevel Designs, *John M Ferron, Yan Wang, Zhiyao Yi, Yue Yin, Eunsook Kim, and Robert F. Dedrick.*

https://www.infoagepub.com/series/Quantitative-Methods-in-Education-and-the-Behavioral-Sciences

THE FEDERATION OF EDUCATION HUBS

Interested in ***Open Science*** and ***educational research***?

Join the New Open Science Listserv for Educational Researchers- [OpenSciEdu-L]

There is a new community space [OpenSciEdu-L] reserved for sharing information and engaging in discussions about Open Science practices with a focus on promoting Open Science, especially in the field of Educational Research. This listserv was created by members of the *Federation* of education hubs. The goal of the Federation is to "Disseminate information about open science across communities and to accelerate adoption of open scholarship in education research through community and connection." Instructions for joining the [OpenSciEdu-L] listserv appear below.

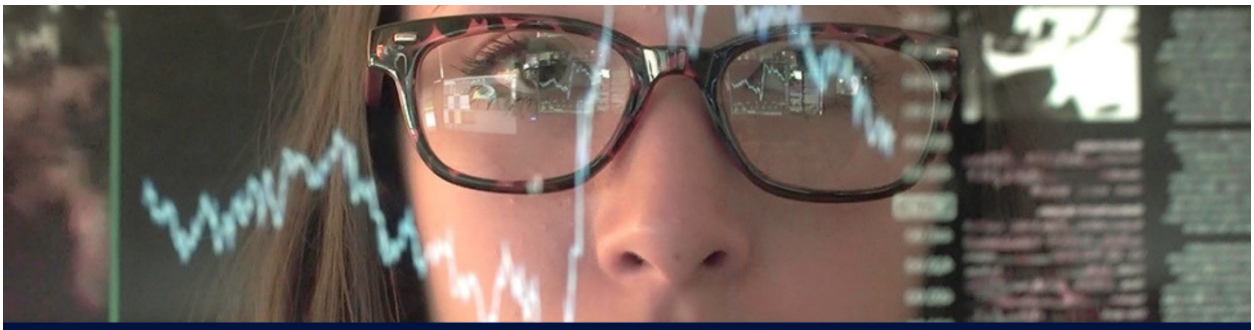
Directions to join OPENSCIEDU-L Listserv:

1. Send an email to **listserv@listserv.uconn.edu**
2. **Make sure to leave the subject line blank**
3. Then type in the body of the email: **SUBSCRIBE OPENSCIEDU-L First Name Last Name**

After joining, to start a conversation on this listserv, simply email OpenSciEdu-L@listserv.uconn.edu

If you have any questions about the listserv, email Betsy McCoach (betsy@uconn.edu). If you have additional questions about the Federation, or if you would like more information about how you can get involved, please contact Matt Makel (makel@jhu.edu).





Enhance Your Quantitative Research & Evaluation Skills in UConn's

Research Methods, Measurement, & Evaluation (RMME) Programs

Prepare to succeed in research, measurement, and evaluation practice at your school, agency, or organization. The full suite of RMME programs offers advanced educational solutions and methods training for educators; district, higher education, and nonprofit administrators; institutional researchers; evaluators; social and behavioral science researchers; analysts; and more!

Find the best program for your needs:

- **Graduate Certificate in Program Evaluation (100% Online OR Campus-based Instruction Options Available**) — Builds and refines skills needed for competent, professional program evaluation practice
- **Master of Arts Degree (100% Online OR Campus-based Instruction Options Available)** — Facilitates development of expertise in quantitative research methodology, psychometrics, measurement, and program evaluation
- **Doctor of Philosophy Degree (Campus-based Instruction)** — Integrates theory and practice to promote the scientific use of quantitative research methodology, cutting-edge data analytic techniques, educational measurement/psychometrics, and evaluation methods within education and the social/behavioral sciences

Learn more about RMME:

UCONN
NEAG SCHOOL OF EDUCATION
RESEARCH METHODS, MEASUREMENT, & EVALUATION

rmme.education.uconn.edu

Statistical Innovations

Thinking outside the brackets!™

Latent GOLD®: #1 program for latent class and mixture modeling

It has both a point and click menu (GUI), and a powerful, flexible syntax language to estimate latent class, latent trait and extended models based on variables of different scale types including continuous, ordinal, nominal, and count.

Recently updated LG 6.0 features include:

Latent GOLD®

Finding gold in your data™

co-developed with Jeroen K Vermunt

- Expanded Stepwise LCA
  - Accounting for differential item functioning
  - Bakk-Kuha two-step approach
  - Latent Markov model in Step3 point-and-click
- Expanded SEM and Factor Analysis
  - Fast EM algorithm for MG-FA with many groups
  - Elegant Factor rotation for MG-EFA
  - Mixture SEM modeling, including stepwise approach
- Latent Class Tree Models
- Parametric and Non-parametric Survival Models
- New output, such as Marginal Effects, Scoring Equations, VLMR Statistics, and User-defined Functions of Parameters
- New Features Simplifying MC Simulations
- Latent Markov Transition Models*

*Accommodates over 100 time points!

View the Latent GOLD® video tutorials:

http://www.statisticalinnovations.com/video-tutorials/
LatentGOLD Course

Download the demo program with hundreds of examples today!

Special Academic and Student rates are available*