UCONN

NEAG SCHOOL OF EDUCATION

Modern Modeling Methods Conference

June 24–26th, 2024

University of Connecticut

2024 Modern Modeling Methods Conference

UCONN
NEAG SCHOOL OF EDUCATION

Welcome and thank you for joining us for the 10th Modern Modeling Methods Conference! It's hard to believe that this is our 10th conference! Special thanks to Denny Borsboom, Sacha Epskamp, and all the presenters for making this wonderful program possible! I really 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 from conference services and Anthony Gambino for providing administrative and logistical support for the conference. I couldn't keep this conference going without them!

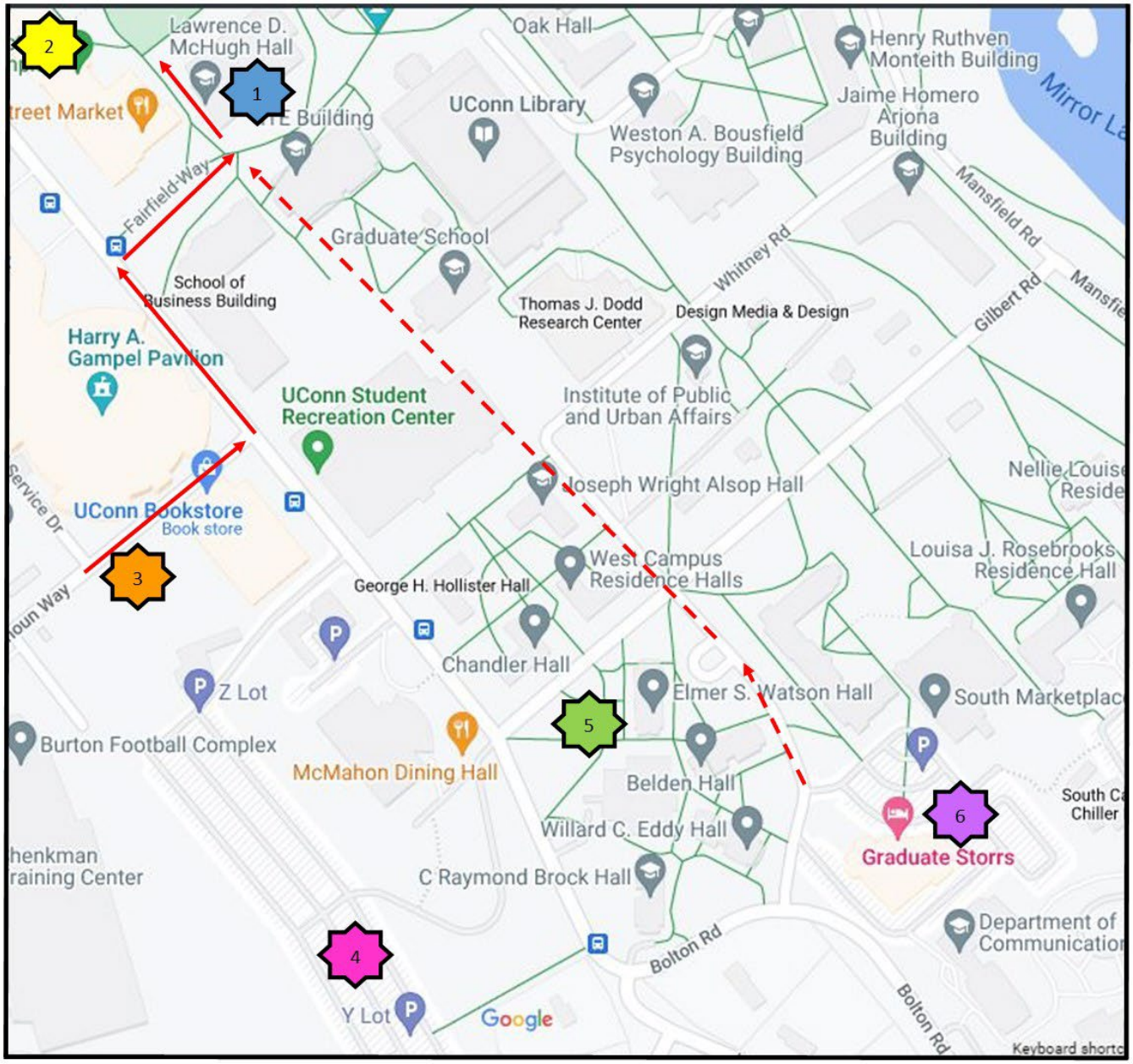
Post-Covid, people travel a bit less than they used to, therefore, after this year, M^3 is moving to a bi-annual conference. The 11th Modern Modeling Methods Conference will be held in late June 2026. Conference proposals will be due January 30, 2026. Be sure to check our website, www.modeling.uconn.edu for additional information about the 2026 conference.

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
University of Connecticut

Modern Modeling Methods Conference Walking Map



1 McHugh Hall

2 Student Union

3 South Garage

4 Y-Lot

5 Watson Hall

6 The Graduate

Modern Modeling Methods – 2024 Preliminary Schedule

Monday, June 24th

Network Psychometrics with psychonetrics in R

Sacha Epskamp

McHugh Hall 102

Continental Breakfast and Registration 8:00 – 9:00 am
McHugh Hall Lobby

Pre-Conference Workshop (McHugh 102) 9:00 am – 5:00 pm
Lunch (on your own) from 12:30-1:30

The freely available psychonetrics package for R provides an encompassing framework for psychometric network modeling, combining typical practices in Structural Equation Modeling (SEM) with undirected network modeling now commonly used in network psychometrics. The psychonetrics package can be used for various types of data (cross-sectional, time-series and panel data), and not only allows researchers to explore relations between observed and latent variables through the use of network models, but also allows researchers to perform confirmatory tests on given network structures and to test for homogeneity in (latent) network structures across groups. This workshop will introduce participants to the psychonetrics package and will teach participants to:

- interpret undirected multivariate network models
- understand differences between within- and between-person effects and cross-sectional and longitudinal data
- install, load and use the psychonetrics package
- use psychonetrics for exploratory network estimation (from cross-sectional, N=1 time-series and panel data)
- use psychonetrics for confirmatory network testing
- use psychonetrics to combine latent variable models with network models
- use psychonetrics for multi-group invariance and homogeneity testing

Familiarity with R and having R and the psychonetrics package installed are recommended for attending the workshop.

Bio: Sacha Epskamp is an associate professor at the National University of Singapore, Department of Psychology. Previously, he worked at the University of Amsterdam in the Department of Psychology and the Centre for Urban Mental Health. In addition, Sacha Epskamp is a former research fellow at the Amsterdam Institute for Advanced Studies and has been a visiting researcher at the Complexity Institute of Nanyang Technological University. In 2016, Sacha Epskamp completed his seminal PhD on network psychometrics—estimating network models from psychological datasets and equating these to established psychometric modeling techniques. This dissertation laid the groundworks for the field of Network Psychometrics. He has implemented these methods in several software packages now routinely used in diverse fields of psychological research. Sacha Epskamp teaches multivariate statistics and data science, and his research interests involve (network) psychometrics, meta-science, reproducibility, complexity, time-series modeling, and dynamical systems modeling. Sacha Epskamp has received several awards for his research, including the Leamer-Rosenthal Prize of the Berkeley Initiative for Transparency in the Social Sciences, the dissertation prize of the psychometric society, and the junior scientific award of the Complex Systems Society.

Tuesday, June 25th

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

Welcome and Opening Keynote by Denny Borsboom 8:30 – 10:10 am
McHugh Hall 102

The Nature of the Measurement Game: Psychological Constructs as Complex Systems

Denny Borsboom
8:40 – 10:10 am

In psychology, the relation between observables and theoretical constructs has traditionally been conceptualized in terms of measurement: observables (e.g., symptoms like self-reproach and suicidal ideation) are viewed as noisy measures of a latent psychological construct that acts as a common cause (e.g., major depression). Important psychometric models, such as the Item Response Theory model and the Factor Analysis model, represent this hypothesis in a statistical structure, which allows researchers to evaluate the tenability of their measurement hypothesis by fitting the model to data. In the past decade, I have investigated an alternative way of thinking, in which observables are not indicators of a latent construct, but interact with one another in a complex system; for instance, the symptom of self-reproach may facilitate suicidal ideation quite independently of whether any latent construct of depression exists or not. Such interactions can statistically be represented in a network model, which allows one to translate the abstract theory into a concrete statistical structure. The development of these models has accelerated in the past decade, and they have become popular in various subdomains of psychology. In this talk, I will discuss these models from a psychometric perspective, and evaluate their plausibility as alternatives to traditional measurement models. I argue that network approaches fundamentally change the nature of the measurement game, and that we have only just begun to evaluate the consequences of these changes.

Bio: Denny Borsboom is Director of the Social and Behavioural Data Science Centre and Professor of Psychology of the University of Amsterdam. Summary of activities. His current research focuses on the development of methodologies for psychological research using complex systems theory and network models as well as the development of theory construction methodology. He is the 2014 recipient of the Samuel J. Messick for his work on validity.

10:10 – 10:30 am Break – McHugh Hall Lobby

Concurrent Paper Session 1 Tuesday 10:30 am – 12:00 pm

Session 1A: Disaggregating Level-Specific Effects and Quantifying Explained Variance in Cross-Classified Multilevel Models
Room 201

Paper	Authors
Disaggregating Level-Specific Effects and Quantifying Explained Variance in Cross-Classified Multilevel Models	Jason D. Rights

Session 1B: Modeling Spatial Data
Room 202

Paper	Authors
Modern Spatial Path Analytic Tools to Investigate the Geography of Medical Debt across a US State	Emil Coman Samuel Bruder Corey Grantham
Investigating the Life Expectancies Differences in the US by Comparing Naïve and Spatial Analytic Methods across Census Tracts, Counties, and States	Emil Coman Jason Byers Blair Johnson Sandro Steinbach Peter (Xiang) Chen Stewart Fotheringham

Session 1C: Measurement Invariance and Moderation
Room 205

Symposium	Authors
Invariance: What Does Measurement Invariance Allow Us to Claim?	John Protzko
A Simulation Study of Alignment Structural Equation Modeling in Assessing Measurement Invariance with Bi-Factor Models	Qingzhou Shi Joni M. Lakin Chunhua Cao
Modeling Construct Change Over Time Amidst Potential Changes in Construct Measurement: A Longitudinal Moderated Factor Analysis Approach	Siyuan Marco Chen Daniel J. Bauer

Concurrent Paper Session 1 **Tuesday 10:30 am – 12:00 pm**

Session 1D: Advances in Mixture Modeling
Room 206

Paper	Authors
Bias-Correction and Robustness for the Latent Profile Transition Analysis with Random Intercepts and Auxiliary Variables: Simulation and Empirical Analyses	Hawjeng Chiou Ming-Chi Tseng Pi-Fang Lin
Signals of Uncertainty and Misspecification in Latent Class Analysis	Zachary Collier Joshua Sukumar
Evaluating Bayesian Transition Diagnostic Classification Models for Reporting Within-Year Progress	Jeffrey C. Hoover W. Jake Thompson

---

Lunch **Tuesday 12:00 – 1:10 pm**

Student Union Ballroom
3rd floor, Student Union

Concurrent Paper Session 2 **Tuesday 1:10 – 2:10 pm**

Session 2A: Multilevel R-Squared Effect Size Measures and Bootstrapped Confidence Intervals
Room 201

Paper	Authors
Multilevel R-Squared Effect Size Measures and Bootstrapped Confidence Intervals	Mairead Shaw Jason D. Rights Jessica Kay Flake

Session 2B: Refining Mediation Analysis in Latent Growth Models
Room 202

Paper	Authors
Refining Mediation Analysis in Latent Growth Models with Sensitivity to Omitted Confounders	Davood Tofighi

Session 2C: nMAX: Restoring Caution and Integrity to the Power Analysis Process
Room 205

Paper	Authors
nMAX: Restoring Caution and Integrity to the Power Analysis Process	Greg Hancock Yi Feng

Session 2D: Dealing with Missing Data
Room 206

Paper	Authors
Comparing Alternatives to the Three-Form Planned Missing Data Design	Alexander M. Schoemann E. Whitney Moore Emily M. Meier Kelly L. Reburn Mark C. Bowler
Estimating the Average Treatment Effect in Longitudinal Randomized Controlled Trials with Missing Data: Will It Help to Add a Quadratic Term?	Manshu Yang Lijuan Wang Scott E. Maxwell

---

2:10 – 2:30 pm Break – McHugh Hall Lobby

Concurrent Paper Session 3 **Tuesday 2:30 – 3:30 pm**

Session 3A: Innovations in Longitudinal Analysis
Room 201

Paper	Authors
An Estimation Approach for Time-Varying Effect Models Using Cubic Splines	Jingwei Li Donna Coffman Megan E. Piper
Deriving Models of Change with Interpretable Parameters: Linear Estimation with Nonlinear Inference	Ethan McCormick

Session 3B: Machine Learning and Modeling
Room 202

Paper	Authors
Machine Learning Structural Equation Modeling and Falsificatory Data Analysis	Michael Truong Ji Yeh Choi
Unsupervised Survey Bot Detection: In Search of High Classification Accuracy	Carl F. Falk Amaris Huang Michael J. Ilagan

Session 3C: Multilevel Modeling in Stata: A Teaching Demonstration
Room 205

Paper	Authors
Multilevel Modeling in Stata: A Teaching Demonstration	Meghan Cain

Session 3D: Understanding Composite-Based Methods via Regression Component Analysis
Room 206

Symposium	Authors
Understanding Composite-Based Methods via Regression Component Analysis	Edward Rigdon

Session 3E: Issues in Factor Analysis
Room 301

Paper	Authors
How Many Factors? Comparing Factor Retention Criteria in Exploratory Factor Analysis	Briana Oshiro D. Betsy McCoach Jessica Kay Flake
Fitting CFA Models with a Mixture of Continuous and Categorical Observed Variables	Christine DiStefano Dexin Shi Guyin Zhang

3:30 – 3:50 pm Break

Concurrent Paper Session 4 Tuesday 3:50 – 4:50 pm

Session 4A: A Framework for Modeling Dyadic Discrepancy
Room 201

Paper	Authors
A Framework for Modeling Dyadic Discrepancy	Robert E. Wickham Kathryn S. Macia

Session 4B: Graphical Modeling
Room 202

Paper	Authors
Application of Gaussian Graphical Models to Visualization and Prediction of Assessment Outcomes	James J. Thompson

Session 4C: Factor Models for Dynamics
Room 205

Paper	Authors
Beyond Cross-sectional Factor Analysis: Dynamic Invariance of Items	Pascal Deboeck Jonathan E. Butner Ascher K. Munion Brian R.W. Baucom R. Chris Fraley Omri Gillath
Bayesian Estimation of Factor Models characterizing Dynamics	Ascher Munion Pascal Deboeck Jonathan Butner

Concurrent Paper Session 4 — Tuesday 3:50 – 4:50 pm

Session 4D: Advances in Modeling for Causal Inference
Room 206

Paper	Authors
Exploring Model-Based Causes for Effect-Size Shrinkage in Educational Research	M. Shane Tutwiler Michael Carlozzi Zoe Kao
Evaluating the Impact of Analytic Approaches in a Multilevel Regression Discontinuity Application	Jason Schoeneberger Christopher Rhoads Faeze Safari

Session 4E: Tools for Teaching Multilevel Modeling
Room 301

Paper	Authors
Using R to enhance shared understanding of linear mixed effect models with and without data across disciplines	Katherine Zavez Ofer Harel
A Communication-Focused Approach to Building Path Diagrams for Multilevel Models	Jeffrey M. Girard

Poster Session and Reception — Tuesday 5:00-7:00pm

Poster Session and Reception
Student Union Ballroom
5:00 – 7:00 pm
Third Floor, Student Union

Come join us for a lively reception with over 40 posters, appetizers, and drinks.

Note to Poster presenters: You may drop off your posters in the Student Union Ballroom (Third Floor, Student Union) between 4:00-5:00 pm on Tuesday. Presenters' poster board numbers are listed on pages 12-14 of the program.

Posters

1. Incorporating Think-Aloud Interviews into Instrument Development: An Applied Example
Lindsay J. Alley, Cole Johnson, Kristy A. Robinson, Jessica K. Flake

2. Predictors of Cause Specific Survival and Prognosis of Pharyngeal Squamous Cell Carcinoma Using Type I Generalized Half Logistic Distribution: Comparison with Cox Regression Model and Random Forest
Phillip Awodutire, Michael Kattan, Oladimeji Akadiri

3. Assessing Genetic Algorithms for Variable Selection in Predictive Modeling Based on Classification: Comparing Loss Functions and Internal Models through a Simulation Study
Catherine Bain, Dingjing Shi

4. Dynamic Modeling of Physiological Reactions: A Markov Chain Approach to Affect Dynamics
Francesca Borghesi, Pietro Cipresso

5. Missing Data Methods for Functional Causal Mediation Analysis with Applications to Smoking Cessation
Cody Campen, Donna Coffman

6. Investigating the Development of Academic Motivations in Adolescents Living in Poverty: A Latent Growth Model Analysis
Elisa Cavicchiolo, Sara Manganelli, Tommaso Palombi, Fabio Lucidi, Fabio Alivernini

7. A Project-Based Instruction Approach to Improving Student Lunar Phases Learning Outcomes: A Quantitative Inquiry
Merryn Cole, Hongwei Yang, Jennifer Anne Wilhelm

8. The Impact of Group Size Ratio and Model Size on the Sensitivity of Fit Measures in Measurement Invariance Testing: A Monte Carlo Simulation Study
Ruiqin Gao, Christine DiStefano, Jin Liu, Ning Jiang, Jiali Zheng

9. Airbag Moderation: A New Hypothesis and Range of Statistical Methods that Capture Many Hitherto Neglected Types of Process
James Hall, Lars-Erik Malmberg, Ariel Lindorff, Nicole Baumann, Pamela Sammons

10. Selection Effects for Inequity in Education: Identifying and Evaluating a New Type of Educational Effect via Application of Airbag Moderation
James Hall, Lars-Erik Malmberg, Gregory Palardy

11. Artifact Corrections for Effect Sizes: Seeing Reality for What It Is
Matthew B. Jané, Edward Kroc, Brenton M. Wiernik, Frederick L. Oswald, Blair T. Johnson

12. What's the Issue with Cut-Scores? A QuantCrit Perspective
Madeline Klotz, Lori Skibbe, Ryan Bowles

13. Moderator Under- and Overspecification in Multiple Regression Analysis
Noah Koehler, Wen Luo, Olukayode Apata

14. Exploring Estimates of Multilevel Reliability for School Based Behavioral Measures
Katie Lane, D. Betsy McCoach

15. Improving the Accuracy in Distinguishing between Second-Order and Bifactor Models Using Penalized Structural Equation Modeling
Hyeryung Lee, Xi Xu

16. Bayesian Analysis of Ordinal Response Variables in Educational Research: A Comparison of Different Priors
Xing Liu

17. Unpacking Subgroup Differences in Treatment Effects: A Causal Decomposition Approach for Mediated Moderation Analysis
Xiao Liu

18. Multilevel Mediation with Unmeasured Cluster-Level Confounders: Evaluating Propensity Score Models
Cameron McCann, Xiao Liu

19. Addressing the Effects of Home Resource Variables on Achievement in International Large Scale Assessment: Are Latent Variables the Right Approach?
Lionel Meng, Daniel Bolt

20. Assessing Differential Item Functioning of the PISA 2018 Academic Resilience Scale
Valerie Ofori Aboah, Latif Kadir, Ann O'Connell

21. Exploring Fit Indices Using Many-Facet Rasch Analysis Model for Validating Economics Test Items
Daniel O. Oyeniran, Enoch O. Olayori, Mopelola F. Oyeniran

22. Comparing Model Complexity of Item Response Theory Models with Randomly Generated Data
Kirsten Reyna, Eric Loken

23. LMMCov: An Interactive Research Tool for Efficiently Selecting Covariance Structures in Linear Mixed Models Using Insights from Time Series Analysis
Perseverence Savieri, Kurt Barbé, Lara Stas

24. Contributions of Evolution Knowledge, Conflict, and Religiosity on Longitudinal Patterns of Evolution Acceptance
Gena Sbeglia, Ross Nehm

25. An Illustration of Advanced Intraclass Correlations for Inter-Rater Reliability
Aaron Matthew Simmons, Jeffrey Girard

26. Measurement Invariance of Attachment and Perceived Stress across Six Intersectional Race/Ethnicity and Gender Identities in Emerging Adults
Jodi Sutherland Charvis, Liam Rozum, Sydney Iacoi, Chrystal Vergara Lopez, Hector Lopez Vergara

27. Planting Decision Trees: Human-Friendly Interpretation of Monte Carlo Simulations, Multiverse Analyses and Multivariate Posterior Distributions
Michael S. Truong, Gabriel Crone, Udi Alter, Ji Yeh Choi

28. Adjusting for Measurement Error in Learning Sciences Data via Bayesian Multilevel Modeling
M. Shane Tutwiler, Alana D. Newell

29. Response Time Modeling for PISA 2018 Math Items Using the Joint Model
Claudia Ventura, Eric Loken

30. Applying a Stepwise MIMIC Modeling Approach to Assess Measurement Invariance in a Latent Class Analysis of Doctoral Socialization
Boshi Wang, Katherine Masyn

31. Examination of Methods of Bayesian Hypothesis Testing for Direct Replication Studies in a Meta-Analytic Framework
Naike Wang

32. Underpowered Studies and Overrepresented Significant Findings in Educational Psychology: A Comprehensive Examination of Empirical Evidence
Naike Wang

33. When to Standardize Your Data? A Comparison of Percentile Bootstrap Confidence Interval for Standardized Indirect Effects
Yiwei Wang, Amanda Kay Montoya

34. Modeling the Dynamics of Pain and Emotion Interactions in Fibromyalgia Using the Valved Bayesian Reservoir Model
Mirinda Whitaker, Akiko Okifuji, Jeanine Stefanucci, Pascal Deboeck

35. A Methodological Review of Intersectionality in Differential Item Functioning, Issues, and Challenges
Winifred G. Wilberforce, Ann A. O'Connell

36. Assessing Psychometric Properties of the Japanese Version of the Perceptions of Inclusion Questionnaire (PIQ) in Student and Teacher Samples
Akie Yada, Susanne Schwab, Kanako Korenaga, Carmen L. A. Zurbriggen

37. Item Sensitivity Matters for Subtle Changes of Emerging Skills
Eleanor Fang Yan, Ryan Bowles, Gary E. Bingham, Xiao Zhang, and Hope Gerde

38. Assessing Model Fit Indices in Ordinal Factor Analysis Models: ML vs. ULS
Guyin Zhang, Dexin Shi, Amanda Fairchild

39. Comparing the Psychometric Performance of Likert Scales and Sliders
Guyin Zhang, Dexin Shi, Amanda Fairchild

40. A Feature Selection Approach to Improve Subgrouping Accuracy in Multivariate Dynamic Processes
Di Jody Zhou, Sebastian Castro-Alvarez, Siwei Liu

Wednesday, June 26th

Continental Breakfast
8:30 – 9:00 am
McHugh Hall Lobby

Concurrent Paper Session 5 Wednesday 9:00 – 10:00 am

Session 5A: Dealing with Daily Data
Room 201

Paper	Authors
How We Cycle: A Tutorial on Combining Day-to-Day Dynamics with Day-of-Week Effects and Weekly Dynamics	Mohammadhossein (Manuel) Haqiqatkhah Ellen L. Hamaker
Generalizability Theory Applied to Daily Relationship Quality: Substantive and Statistical Directions	Madison Shea Smith Susan C. South

Session 5B: Variable Selection in Building Generalized Linear Mixed Models
Room 202

Paper	Authors
A New Algorithm for Variable Selection in Building Generalized Linear Mixed Models	Yutian T. Thompson Yaqi Li Hairong Song David E. Bard
A New Method for Variable Selection in Building GLMMs with Incomplete Data	Yutian T. Thompson Yaqi Li David E. Bard

Session 5C: Multilevel R-Squared Measures in Stata
Room 205

Paper	Authors
Using the mlmr2 Package to Estimate Multilevel R-Squared Measures in Stata	Anthony J. Gambino

Session 5D: Disentangling Person-Dependent and Item-Dependent Causal Effects
Room 301

Paper	Authors
Disentangling Person-Dependent and Item-Dependent Causal Effects: Applications of Item Response Theory to the Estimation of Treatment Effect Heterogeneity	Joshua Gilbert Luke Miratrix Mridul Joshi Benjamin Domingue

10:00 – 10:20 am Break

Concurrent Paper Session 6 Wednesday 10:20 – 11:50 am

Session 6A: Bayesian Models
Room 201

Paper	Authors
Comparing the Accuracy of Three Predictive Information Criteria for Bayesian Linear Multilevel Model Selection	Sean Devine Carl F. Falk Ken A. Fujimoto
Modeling Misspecification as a Parameter in Bayesian Structural Equation Models	James Uanhoro
Bayesian Semiparametric Item Response Theory Models: A Methodological Illustration	Meng Qiu Sally Paganin

Session 6B: Novel Applications of Latent Variable Modeling
Room 202

Paper	Authors
Decomposing the Effects of Suffering on Depression Using a Reparameterized SEM and Penalized Maximum Likelihood	Noah Padgett Richard Cowden Tyler J. VanderWeele
Intensive Longitudinal Modeling of Big Social Media Data	Jeffrey M. Girard
Validation of the Evidence-Based Practices Attitudes Scale (EBPAS) using Dynamic Fit Index Cutoffs	Julian M. Hernandez-Torres Natalia Giraldo-Santiago Daniel McNeish

Session 6C: Reproducibility, Replicability, and Registration of Simulation Studies
Room 205

Paper	Authors
Why Do Reproducibility and Replicability of Simulation Studies Matter?	Jessica Kay Flake
Replicating Simulation Research: A Case Study	Tristan D. Tibbe
Registered Reports for Simulation Studies	Amanda Kay Montoya
How to Register Your Simulation Study: Our Lessons Learned	Lindsay Alley Mairead Shaw

Concurrent Paper Session 6 Wednesday 10:20 – 11:50 am

Session 6D: Missing Data: Problems and Solutions
Room 301

Paper	Authors
Evaluating the Effect of Change on Change in Cross-Domain Latent Growth Curve Analysis with Missing Data	Parisa Rafiee Manshu Yang
A Solution for Including Auxiliary Variables with Categorical Dependent Variable Estimation in SEM	Jason Newsom Mallory R. Kroeck Brian T. Keller Nicholas A. Smith
Dummy Variable Adjustment Technique: An Alternative to Maximum Likelihood and Multiple Imputation	Roula Aldib Lee Branum-Martin

---

Lunch Wednesday 12:00 – 1:00 pm

Student Union

Please be sure to return your lunch card!

Please note: Dining cards can be used *only* on Wednesday, June 26th 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.

Concurrent Paper Session 7 — Wednesday 1:00 – 2:30 pm

Session 7A: Modeling Individual Differences
Room 201

Paper	Authors
Measurement Invariance of the Big Five across Socioeconomic Background: Multigroup Confirmatory Factor Analysis and Alignment Optimization	Emilija Meier-Faust Sandra Bohmann
Applying SEM Based Person-Fit to the Wechsler Adult Intelligence Scale IV Demonstrate the Validity of Measurement at the Individual Level	Jared Block Steven Reise

Session 7B: Network Applications
Room 202

Paper	Authors
Exploring the Emotional Well-Being of Young Adults through Network Psychometrics	Daniel Hernández-Torrano
Exploring the Dynamics of Motivation in Physical Activity among Older Adults Through Panel Network Approach	Tommaso Palombi Denny Borsboom René Freichel Elisa Cavicchiolo Fabio Lucidi Fabio Alivernini
Estimating the Group Differences of Longitudinal Network Analysis: An Example of Eating Disorder Psychopathology	Jihong Zhang Jinbo He

Session 7C: Advances in Multilevel Mixture Modeling
Room 205

Paper	Authors
Examining the Effect of Nested Data on Class Enumeration and Model Fit in Latent Profile Analysis	Angela Starrett Katherine Masyn
Application of a Multilevel Latent Class Analysis with Cross-Classified Data	Audrey Leroux Katherine Masyn
Moderated Nonlinear Mixture Analysis for Longitudinal Invariance Testing in Latent Transition Analyses	Katherine Masyn Boshi Wang

Concurrent Paper Session 7 **Wednesday 1:00 – 2:30 pm**

Session 7D: Structural Equation Modeling Methods
Room 301

Paper	Authors
Comparing Approaches to Examine Multiple Binary Moderators in Latent Variable Models	Kaylee Litson Amanda Kay Montoya Yiwei Wang
Quantile Structural Equation Modeling: Testing a Novel Distance Based Approach	Jeffrey Shero Zhixin Zhu Jessica Logan

---

2:30 – 3:00 pm Break – Dairy Bar Ice Cream in McHugh Lobby

Concurrent Paper Session 8 **Wednesday 3:00 – 4:00 pm**

Session 8A: Innovations in Mixed Modeling
Room 201

Paper	Authors
Model Selection of GLMMs in the Analysis of Count Data in SCEDs: A Monte Carlo Simulation	Haoran Li
Latent Class Clustering of Random Coefficient Estimates Obtained from a Multilevel Analysis	Jay Magidson Jeroen Vermunt

Session 8B: Innovations in Structural Equation Modeling
Room 202

Symposium	Authors
Introducing the Deleted One Covariance Residual Measure to the Structural Equation Modeling	Fathima Jaffari Jennifer Koran
Ruling Out Latent Time Varying Confounders in Two-Variable Multi-Wave Studies	David Kenny D. Betsy McCoach

Session 8C: Restructuring Basic Statistical Curricula: Mixing Older Analytic Methods with Modern Software Tools in Psychological Research
Room 205

Paper	Authors
Restructuring Basic Statistical Curricula: Mixing Older Analytic Methods with Modern Software Tools in Psychological Research	Emil Coman James Jaccard Sabrina Uva Ana-Maria Cazan

Session 8D: InclusiMetrics
Room 301

Paper	Authors
Advancing Research on Methodology: A Panel Discussion on the Creation of a Minority-Centered Methodological Conference, InclusiMetrics	Marcus Harris Zachary Collier

Paper Abstracts

Concurrent Paper Session 1 — Tuesday 10:30 – 12:00 pm

Session 1A: Disaggregating Level-Specific Effects and Quantifying Explained Variance in Cross-Classified Multilevel Models (Room 201)

Disaggregating Level-Specific Effects and Quantifying Explained Variance in Cross-Classified Multilevel Models

Jason D. Rights

In many areas of research, data often have a cross-classified structure, whereby observations are nested within multiple types of non-hierarchical clusters. In this talk, I first discuss ways that, in cross-classified multilevel models, slopes of lower-level predictors can implicitly reflect an ambiguous blend of multiple level-specific effects. This general concept is well-established in hierarchical multilevel models, but is almost always ignored in cross-classified models. I clarify the specific consequences of conflating these effects in cross-classified models, with consideration of both the fixed portion and random portion of the model. I further discuss alternative ways that researchers can disaggregate these effects, including approaches that involve fully cluster-mean-centering, partially cluster-mean-centering, and contextual effect modeling, each of which provides a unique interpretation of model parameters. With this understanding of level-specific effects, I then show how it can be used to inform a full decomposition of outcome variance for cross-classified models that can, in turn, yield a framework of R-squared measures to quantify explained variance. I discuss this suite of R-squared measures in juxtaposition to pre-existing measures and show that the full decomposition provides more complete and informative results.

Session 1B: Modeling Spatial Data (Room 202)

Modern Spatial Path Analytic Tools to Investigate the Geography of Medical Debt across a US state

Emil Coman, Samuel Bruder, Corey Grantham

We illustrate the analysis of individual resident data using spatial analytics methods, and a small claims medical debt dataset from a US state (CT). To what extent are racial/ethnic minority communities (census tracts with more non-White residents) differentially impacted by medical debt? Do these differences shrink when considering the income differences between communities? To describe the problem in CT, we present the geography of medical small claims in CT, and the populations of residents most affected by it. We use geographic/spatial ("econometric") models aimed at answering whether racial/ethnic minorities are more affected. We describe the "geography" of small claims medical debt in CT across census tracts (N = 820) and state senate (upper house, n = 36) levels, in terms of number of court claims per 10,000 residents. We first describe the issue of neighbors influencing each case's datapoint (through "contagion" or "interference"), and the practical solution of turning naïve analyses into proper spatial analyses, by controlling in any model for the spatial lag (average of neighbors' values of the outcome), the so-called spatial lag model. Census tracts with more non-White minority residents experience higher rates of small claims medical debt.

Investigating the Life Expectancies Differences in the US by Comparing Naïve and Spatial Analytic Methods across Census Tracts, Counties, and States

Emil Coman, Jason Byers, Blair Johnson, Sandro Steinbach, Peter (Xiang) Chen, Stewart Fotheringham

We investigate a simple research inquiry about the relation between life expectancy and racial/ethnic composition: RQ. Do regions in the US with more non-White residents experience different life expectancies, and if so by how much? We provide a range of answers by analyzing a comprehensive US dataset at 3 levels: census tracts, counties, and states (contiguous). We provide answers by: (i)

Zooming in and out into different geographical levels, and (ii) Dialing down and up the range of the statistical tools. We examine the different variabilities, co-variabilities, and spatial effects between: %Non-White, Life Expectancy and Income, at 3 levels: census tracts, counties, and US states. We report each indicator's Moran I, the non-independence parameter. All naïve/a-spatial parameters are inflated by spatial non-independence (the spatial structure). When zooming out to larger regions, the patterns of relations and effects change, but not in a consistent manner, i.e. not necessarily by parameters decreasing in size. The processes that appear to be at work differ by geography. There are direct, indirect (through income) and total effects of a region having more racial/ethnic minority residents on its aggregate (lower) life expectancy, seen across census tracts and across counties, but not when analyzing across states.

Session 1C: Measurement Invariance and Moderation (Room 205)

Invariance: What Does Measurement Invariance Allow Us to Claim?
John Protzko

Measurement involves numerous theoretical and empirical steps—ensuring our measures are operating the same in different groups is one step. Measurement Invariance occurs when the factor loadings and item intercepts or thresholds of a scale operate similarly for people at the same level of the latent variable in different groups. This is commonly assumed to mean the scale is measuring the same thing in those groups. Here we test the assumption of extending measurement invariance to mean common measurement by randomly assigning American adults (N=1500) to fill out scales assessing a coherent factor (search for meaning in life) or a nonsense factor measuring nothing. We find a nonsense scale with items measuring nothing shows strong measurement invariance with the original scale, is reliable, and covaries with other constructs. We show measurement invariance can occur without measurement. Thus, we cannot infer that measurement invariance means one is measuring the same thing, it may be a necessary but not a sufficient condition.

A Simulation Study of Alignment Structure Equation Modeling in Assessing Measurement Invariance with Bi-Factor Models
Qingzhou Shi, Joni M. Lakin, Chunhua Cao

This study explores the performance of Alignment Optimization within Structural Equation Modeling (ASEM) in bi-factor models under varied conditions using the Monte Carlo simulation. It focuses on different group sample sizes, non-invariance magnitudes, indicator non-invariance proportions, and cross-loading proportions and magnitudes. The model, reflecting a common structure in cognitive assessments, includes one general and three specific factors, each specific factor influencing 15 dichotomous items. Simulations employed theta parameterization with WLS estimator and Fixed alignment, assessing performance through absolute bias and 95% coverage rate. The findings indicate that smaller group sizes significantly affect mean coverage rate and bias in factor means and variances. Increased non-invariance magnitudes and the presence of cross-loadings also elevate mean bias. These trends extend to factor loadings and item thresholds, with smaller sizes leading to higher mean bias. Regression analyses reveal that smaller group sizes, higher non-invariance proportions, and the presence of cross-loadings significantly increase bias and decrease the 95% coverage rate for factor means and variances. Larger group sizes, higher non-invariance levels, and the presence of cross-loadings had mixed effects on the coverage rates for item loadings and thresholds, indicating a more complex relationship between these conditions and the performance of ASEM.

Modeling Construct Change Over Time Amidst Potential Changes in Construct Measurement: A Longitudinal Moderated Factor Analysis Approach
Siyuan Marco Chen, Daniel J. Bauer

In analyzing longitudinal data, growth curve models are often fit to a repeated measure constructed as a sum or mean of scale items, making an implicit assumption of constancy of measurement. This practice risks confounding actual construct change with changes in measurement (i.e.,

differential item functioning; DIF), threatening the validity of conclusions. An improved method that avoids such confounding is the second-order growth curve model (SGC). It specifies a measurement model at each occasion of measurement. The applicability of SGC is hindered by key limitations: (1) SGC treats time as continuous when modeling construct growth but as discrete when modeling measurement, reducing interpretability and parsimony; (2) the evaluation of DIF becomes increasingly error-prone given multiple timepoints and groups; (3) DIF associated with continuous covariates is difficult to incorporate. Drawing on moderated nonlinear factor analysis (MNLFA), we propose an alternative approach that provides a parsimonious framework for including many timepoints and DIF from different types of covariates. We implement this model through Bayesian estimation, allowing for incorporation of regularizing priors to facilitate efficient evaluation of DIF. We demonstrate a two-step workflow of measurement evaluation and growth modeling, with an empirical example examining changes in adolescent delinquency over time.

Session 1D: Advances in Mixture Modeling (Room 206)

Bias-Correction and Robustness for the Latent Profile Transition Analysis with Random Intercepts and Auxiliary Variables: Simulation and Empirical Analyses

Hawjeng Chiou, Ming-Chi Tseng, Pi-Fang Lin

Latent profile transition analysis with random intercept (RI-LPTA) can estimate the individual differences across times. By introducing a random intercept factor to represent latent class transitions within individuals at the within-subject level, the latent profile model is able to simultaneously capture variation across participants at the between-subject level (Tseng, 2022). As a result, the random intercept as well as the latent class membership could be further explained by covariates to predict the distal outcomes for identifying the antecedents and consequences in more complex, structural models. However, adding the covariates or distal variables (auxiliary variables) into model may cause an undesirable shift of the latent class variable. Estimates of the structural parameters may also be distorted by ignoring the classification error that is introduced in the classification step. To overcome these issues, this paper introduces the bias-adjusted three-step approach (Bolck, Croon, & Hagenaars, 2004) into RI-LPTA with auxiliary variables. Simulation research suggests that the BCH's correction on class membership impacts the estimation of structural parameters. To implement the BCH's correction, some restrictions and constraints need to be imposed in the Mplus syntax. Researchers constructing empirical longitudinal models are advised to apply the bias correction approach for auxiliary variables with RI-LPTA. Empirical analyses based on the KIT (Kids in Taiwan: National Longitudinal Study of Child Development & Care) also demonstrate the application of RILPTA with auxiliary variables corrected for bias of categorization error.

Signals of Uncertainty and Misspecification in Latent Class Analysis

Zachary Collier, Joshua Sukumar

Modification indices play a crucial role in improving model fit by suggesting specific parameter modifications. However, they are not applicable to latent class analysis (LCA) because the estimation and fit assessment methods are different from traditional structural equation models (SEMs). To address this gap, our approach introduces the use of gradient descent to identify sensitive parameters, bolstering the reliability of study conclusions. Through iterative adjustments guided by the objective function, gradient descent aligns model predictions with observed data. Consequently, updated parameter estimates highlight potential inadequacies in the original model's capture of true relationships. This finding prompts further investigation, potential refinement, and consideration of alternative specifications to mitigate misspecification or omitted variable bias. By adopting gradient descent, researchers gain a powerful tool for enhancing the robustness and validity of findings in LCA scenarios.

Evaluating Bayesian Transition Diagnostic Classification Models for Reporting Within-Year Progress

Jeffrey C. Hoover, W. Jake Thompson

Transition diagnostic classification models (TDCMs) provide feedback on students' mastery of attributes over time. When assessment results are scaled using a log-linear cognitive diagnosis model (LCDM), it may be practical to report feedback on within-year progress using LCDM-based approaches, since the LCDM would be consistent with models used for reporting students' performance. We evaluated the classification accuracy of a Bayesian TDCM and two LCDM-based approaches for measuring within-year progress using data simulated based on data from an operational assessment program. We simulated data using calibrated parameter estimates from the operational assessment program. The item responses were simulated to be consistent with the operational assessment program's administration guidelines. We then compared the classification accuracy of the three approaches. The TDCM had higher mean classification accuracy than the LCDM-based approaches, meaning the Bayesian TDCMs best classified students' attribute mastery transitions from the fall to the spring. The higher classification accuracy suggests the TDCM provides more accurate feedback to students, parents, and teachers regarding attribute mastery changes throughout the school year, even though the assessments are scaled with LCDMs.

Concurrent Paper Session 2 — Tuesday 1:10 – 2:10 pm

Session 2A: Multilevel R-Squared Effect Size Measures and Bootstrapped Confidence Intervals (Room 201)

Multilevel R-squared Effect Size Measures and Bootstrapped Confidence Intervals

Mairead Shaw, Jason D. Rights, Jessica Kay Flake

Multilevel models (MLMs) are widely used to analyze clustered data structures such as people nested within groups or trials nested within people. Effect sizes are necessary for contextualizing results from statistical models and are often required by journals and funders. Rights and Sterba (2019) developed a comprehensive approach for R-squared effect size measures in MLMs. We developed an R package, r2mlm (Shaw et al., 2023), for calculating these measures and recently updated the package to include bootstrapped confidence intervals. In this illustration we will: (1) Define and understand effect sizes for MLMs; (2) Estimate and interpret the effect sizes using R; (3) Detail bootstrapping and confidence interval calculation methods for multilevel R-squared; and (4) Compare the coverage of different bootstrapping/confidence interval methods. The talk involves both lecture and live demonstration in R, so participants are encouraged to bring their laptops and follow along.

Session 2B: Refining Mediation Analysis in Latent Growth Models (Room 202)

Refining Mediation Analysis in Latent Growth Models with Sensitivity to Omitted Confounders

Davood Tofighi

Mediation analysis, fundamental for elucidating indirect pathways between variables, hinges on the no omitted confounders assumption, which presumes the non-existence of unmeasured variables influencing the antecedent, mediator, and outcome. This assumption is particularly challenging to satisfy in observational studies, leading to potential biases. Addressing this, we introduce a novel extension to sensitivity analysis for non-randomized latent growth curve models (LGCMs), termed extended correlated augmented mediation sensitivity analysis (ECAMSA). ECAMSA advances the assessment of mediation effects' robustness against omitted confounders by quantifying the potential impact of unobserved variables on the mediation process. Our approach enhances the reliability of conclusions drawn from LGCMs in nonexperimental settings by systematically evaluating how mediation effects may vary under the influence of omitted confounders. By integrating considerations of unmeasured variables' effects,

ECAMSA provides a critical tool for researchers to discern the true nature of mediated relationships, ensuring a deeper and more accurate understanding of the dynamics at play. This methodological advancement represents a significant contribution to the field, allowing for more nuanced and confident interpretations of mediation analysis outcomes in the face of inherent observational study limitations.

Session 2C: nMAX: Restoring Caution and Integrity to the Power Analysis Process (Room 205)

nMAX: Restoring Caution and Integrity to the Power Analysis Process
Greg Hancock, Yi Feng

In a time when the alarms of research replicability are sounding louder than ever, mapping out studies with statistical and inferential integrity is of paramount importance. That said, as methodologically-oriented researchers cannot help but to have observed, in an increasingly competitive environment with financial belts ever-tightening, the process of a priori power analysis has devolved from its intended origins as a planful insurance policy to a collection of yard sale hagglings whose goal is slinking away with the lowest acceptable bid. But even the most sincere attempt at sample size planning is fraught with the fundamental challenge of setting values for not only a model's focal parameters for which statistical tests are planned, but also each of the model's peripheral parameters as well. The current work presents new methods that attempt (1) to restore appropriate conservatism and robustness to the sample size planning process, and (2) to greatly simplify that process. Derivations and suggestions for practice are presented using the framework of measured variable structural models as they subsume many of the types of models (e.g., multiple regression, ANOVA) for which sample size planning is of interest.

Session 2D: Dealing with Missing Data (Room 206)

Comparing Alternatives to the Three-Form Planned Missing Data Design
Alexander M. Schoemann, E. Whitney Moore, Emily M. Meier, Kelly L. Reburn, Mark C. Bowler

Planned missing data designs, particularly the three-forms design, are a promising method for survey research. These methods allow researchers to shorten a survey while assessing the same number of items, resulting in lower participant fatigue or to increase in the number of items participants can complete in a short amount of time. In this presentation we compare a three forms design to two alternate designs, a complete data design with no planned missingness and a design with random planned missing. In the three-forms design, the items on a survey are broken up into four sets: X, A, B, and C. All participants complete the items in the X set and participants are randomly assigned to complete items in two of the remaining three sets resulting in three different forms of the survey: XAB, XAC, and XBC. In a complete data design, all participants are assigned to complete all questions. A random planned missing data design is where participants are assigned to receive a random subset of the total items. We compare these designs using real world data and Monte Carlo simulations.

Estimating the Average Treatment Effect in Longitudinal Randomized Controlled Trials with Missing Data: Will It Help to Add a Quadratic Term?
Manshu Yang, Lijuan Wang, Scott E. Maxwell

Longitudinal randomized controlled trials (LRCTs) have been commonly used in behavioral and health-related studies to evaluate the effectiveness of interventions. Outcomes in LRCTs may follow either straight-line or curvilinear change trajectories over time, and missing data are almost inevitable in such trials. The current study aims to investigate (1) whether the estimate of average treatment effect (ATE) would be biased if a straight-line growth model is fit to LRCT data with quadratic growth and missing-completely-at-random (MCAR) or missing-at-random (MAR) data, and (2) whether adding a quadratic term to a straight-line growth model would improve the ATE estimation and inference. Four models were compared via a simulation study, including the straight-line growth (SLG) model and three types of quadratic growth models with arm-invariant and fixed quadratic effect (QG-AIF), arm-specific and fixed quadratic effects (QG-ASF), or arm-specific and random quadratic effects (QG-ASR). Results

suggest that fitting an SLG model to quadratic growth data often yielded severe biases in ATE estimates, even if data were MCAR or MAR. Given four or more waves of longitudinal data, the QG-ASR model outperformed the other methods; for three-wave data, the QG-ASR model was not applicable and the QG-ASF model performed well.

Concurrent Paper Session 3 — Tuesday 2:30 – 3:30 pm

Session 3A: Innovations in Longitudinal Analysis (Room 201)

An Estimation Approach for Time-Varying Effect Models Using Cubic Splines
Jingwei Li, Donna Coffman, Megan E. Piper

Traditional mediation analysis typically examines the relations among an intervention, a time-invariant mediator, and a time-invariant outcome variable. Although there may be a total effect of the intervention on the outcome, there is a need to understand the process by which the intervention affects the outcome. This indirect effect is frequently assumed to be time-invariant. With improvements in data collection technology, it is possible to obtain repeated assessments over time resulting in intensive longitudinal data. This calls for an extension of traditional mediation analysis to incorporate time-varying variables as well as time-varying effects. We focus on estimation and inference for the time-varying mediation model, which allows mediation effects to vary as a function of time via cubic spline interpolation. Two simulation models and a smoker's health research are studied to compare this method with local smoothing. More accurate results are obtained for the cubic spline interpolation when there are fewer time points.

Deriving Models of Change with Interpretable Parameters: Linear Estimation with Nonlinear Inference
Ethan McCormick

In the modeling of change over time, there are often disconnects between substantive developmental theories and longitudinal statistical models. That is, theory is understood and advanced in terms of meaningful developmental quantities (e.g., peaks, inflections, timing, and tempo) while common polynomial models estimate effects of powered terms of time in a linear, additive form. These linear parameterizations have many advantages, especially their computational efficiency, but the quantities estimated in these models are often difficult to directly connect to theoretical ideas of change over time. To bridge the gap between estimation and theory development, I propose a series of approaches for linear estimation with nonlinear inference (LENI), where the stable, easily-estimated linear model results are converted through transformation functions into nonlinear estimates which align better with theoretical quantities of interest. I first derive these interpretable nonlinear parameters, then demonstrate transformations of linear model results—including fixed and random effects and conditional effects of covariates—into the results of a nonlinear alternative model. I summarize a linearized structural equation model approach which can be flexibly applied to model any known nonlinear target function into a linearly-estimable model. I also discuss recommendations for researchers and directions for future work.

Session 3B: Machine Learning and Modeling (Room 202)

Machine Learning Structural Equation Modeling and Falsificatory Data Analysis
Michael Truong, Ji Yeh Choi

The synthesis of Machine Learning (ML) and Structural Equation Modeling (SEM) methods is a rapidly developing field of quantitative methods. ML leverages high dimensional datasets for powerful predictive modeling, but may suffer from poor interpretability. In contrast, SEM facilitates the interpretable analysis of the relationships between theoretical constructs and for testing how well the data fits theory, but it may be inappropriate for high dimensional datasets and where theory is immature.

Hence, the goal of ML-SEM research has been to synergize the strengths of ML and SEM, while offsetting their respective weaknesses. However, a persistent critique of ML-SEM is the concern of over-fitting. In this methodological innovation talk, we aim to side-step this critique by proposing the use of ML-SEM for *Falsificatory Data Analysis* (FDA). To do this, we propose a new ML-SEM technique called Integrated-Generalised Structured Component Analysis Trees (IGSCA trees). We introduce FDA as: (1) a precipitation of recent trends; and (2) as a potentially helpful mindset for guiding the development of theory, data collection and data analysis. Finally, we show how the FDA mindset can guide the use of IGSCA-trees towards testing auxiliary hypotheses, such as data quality or measurement validity in both simulated and real data.

Unsupervised Survey Bot Detection: In Search of High Classification Accuracy
Carl F. Falk, Amaris Huang, Michael J. Ilagan

In this research, we study a recently developed algorithm to detect survey bots. The algorithm requires neither a measurement model nor a sample of known humans and bots; thus, it is model agnostic and unsupervised. It involves a permutation test under the assumption that Likert-type items are exchangeable for bots, but not humans. Although the algorithm maintains a nominal 95% sensitivity for detecting bots, its classification accuracy may depend on other inventory-specific or demographic factors. Generating hypothetical human responses from a well-known item response theory (IRT) model, we use simulations to understand how classification accuracy is affected by IRT-based item properties, the number of items, the number of latent factors, and factor correlations. In an additional study, we simulate bots to contaminate real human data from 36 publicly available datasets to understand the algorithm's classification accuracy under a variety of real measurement instruments. We identify conditions under which classification accuracy is around 95% or above, but also conditions under which accuracy is quite low. In brief, performance is better with more items, more categories per item, and a variety in the IRT difficulty or means of the survey items.

Session 3C: Multilevel Modeling in Stata: A Teaching Demonstration (Room 205)

Multilevel Modeling in Stata: A Teaching Demonstration
Meghan Cain

Spend less time teaching syntax and more time teaching theory in your next multilevel modeling course. This presentation introduces Stata software: a complete, integrated software package that provides tools for data management, visualization, statistics, and automated reporting. It is affordable, easy to use for students with no programming experience, and flexible enough for advanced users to develop their own estimators and commands. This presentation will start with a brief overview of Stata, and then focus on its suite of commands for fitting, evaluating, and interpreting multilevel models. We'll begin with tools for visualizing nested and longitudinal data to teach the intuition behind MLMM. We will then fit some random-intercept and random-slope models, using point-and-click and through commands. There will be examples with different estimators and covariance structures. We will also compare models statistically and learn how to easily create model-comparison tables. Finally, I will demonstrate a few postestimation tools for diagnostics and calculating best linear unbiased predictions (BLUPs) of the random effects.

Session 3D: Understanding Composite-Based Methods via Regression Component Analysis (Room 206)

Understanding Composite-Based Methods via Regression Component Analysis
Edward Rigdon

Composite-based approaches to structural equation modeling, including partial least squares path modeling or PLS and generalized structured component analysis or GSCA have been growing in popularity, but differing literatures and even symbol palettes can make it hard for those familiar with

common factor models to understand these alternatives. Regression component analysis (RCA) transforms a common factor model into an equivalent composite-based model, adopting factor model parameter estimates and using them to calculate weights. RCA equates to modeling with "regression method" factor scores, but here those scores are consistent with the composite-based population though they are not consistent with a common factor model. When the original factor model is exactly correct in the population, RCA replicates results from regression weight forms of partial least squares (PLS) path modeling and generalized structured component analysis (GSCA), when RCA composites are standardized in the same way as PLS and GSCA composites. Other relationships emerge depending on the sampled population. Here, all analytical methods are described using one consistent palette of symbols. However, one important takeaway is that the equations used to describe some methods can mislead regarding the methods' actual behavior.

Session 3E: Issues in Factor Analysis (Room 301)

How Many Factors? Comparing Factor Retention Criteria in Exploratory Factor Analysis
Briana Oshiro, D. Betsy McCoach, Jessica Kay Flake

Exploratory factor analysis (EFA) has been an important method in psychological research for over a century. Despite its long history, there has been no consensus as to the best method to determine the number of underlying factors in the factor model. Previous simulation studies have focused on the accuracy and bias of various factor retention criteria (FRC), resulting in a variety of suggestions and criteria for deciding the number of factors in EFA. However, accuracy and bias do not tell the whole story of how well an FRC works. In this study we also consider the variance of results and the percentage of over and underestimates to neutrally compare the performance of nine FRC recently recommended in the literature. We found that the least biased FRC is not the most accurate nor the most consistent. Although parallel analysis, AIC, and sequential model chi-square test had the least bias, the extended version of exploratory graph analysis using the Gaussian graphical model (Golino et al., 2020) was one of the most accurate while also exhibiting the lowest variances in results.

Fitting CFA Models with a Mixture of Continuous and Categorical Observed Variables
Christine DiStefano, Dexin Shi, Guyin Zhang

Questionnaires are commonly used with measurement investigations, where data are often collected using a Likert scale. When analyzing a Likert scale, data for latent variable analyses (e.g., confirmatory factor analysis [CFA]) may be considered continuous if there are at least five ordered categories and categorical if four or fewer categories are included. However, situations may occur where data are mixed, such as including a battery of assessments, assessments which include items with different numbers of scale points, and/or respondents failing to use all categories for a Likert scale. This study investigates a variety of robust estimation methods with CFA that include a mixture of continuous and categorical data.

Concurrent Paper Session 4 — Tuesday 3:50 – 4:50 pm

Session 4A: A Framework for Modeling Dyadic Discrepancy (Room 201)

A Framework for Modeling Dyadic Discrepancy
Robert E. Wickham, Kathryn S. Macia

Researchers interested in dyadic phenomena often apply regression techniques aimed at partitioning the variances and covariances among dyad member responses into theoretically meaningful components. The predictors and outcomes in these analyses are observed or latent variables representing the level of Person or Dyad level attributes, and the regression relationships describe the degree to which individuals or dyads with higher scores on the predictor attribute(s) have higher or lower scores on the

outcome attribute(s). However, there are some research questions that may be better served by using the difference between Person level scores provided by each dyad member. As such, the present work re-casts the well-established latent difference score model for longitudinal panel data as a Dyadic Discrepancy Model (DDM) for capturing the difference between Person level scores for both distinguishable and indistinguishable dyads. The DDM is contrasted with established approaches for modeling dyadic matching and is incorporated into the existing framework of dyadic modeling approaches. A series of worked examples illustrate how the DDM may be combined with the actor-partner interdependence model, variations on the common-fate model, as well as actor-partner common-fate hybrid models.

Session 4B: Graphical Modeling (Room 202)

Application of Gaussian Graphical Models to Visualization and Prediction of Assessment Outcomes
James J. Thompson

With the use of computerized testing, ordinary assessments can capture both answer accuracy and answer response time. For the Canadian Programme for the International Assessment of Adult Competencies (PIAAC) numeracy and literacy subtests, person ability, person speed, question difficulty, question time intensity, person score discrimination, question time discrimination, and pace (rank of response time within question) were assessed. Undirected Gaussian Graphical Model networks based on partial correlations were predictive of the measures as nodes. The population-based model extrapolated well to person estimations. It was shown that the "training" Canadian model generalized with minor differences to four other English-speaking PIAAC assessments. Another avenue to network visualization with additional interpretive benefits exists (exploratory graph analysis, EGA). The EGA method is less conservative (more edges are included) than the previous methodology described. However, the EGA method assesses unidimensionality, assigns communities, recognizes redundant information, and by a loading comparison test suggests that the observed data was best explained by the network model rather than a factor model. Thus, the undirected network approach provides a heuristic that is both descriptive and predictive. Such models are not causal and can be taken as examples of "mutualism".

Session 4C: Factor Models for Dynamics (Room 205)

Beyond Cross-Sectional Factor Analysis: Dynamic Invariance of Items
Pascal Deboeck, Jonathan E. Butner, Ascher K. Munion, Brian R.W. Baucom, R. Chris Fraley, Omri Gillath

Commonly, longitudinal models use scales developed based on cross-sectional factor analyses. While cross-sectional factor analysis has been a mainstay for scale development, modeling individual correlations does not ensure that items change similarly over time. If items were to change in different ways over time, generating latent or summed scores may produce apparent changes unrelated to the phenomena of interest, like the beats that can be heard when two similar tones are played. Additionally, questions could be raised about whether a scale consists of a single underlying construct. This presentation uses factor analysis to consider whether items change similarly across time. Like measurement invariance, differing levels of dynamic invariance can be imposed on data to test whether items have similar dynamics. Beyond items having a similar scale, one can test whether the relations (e.g., regressions/covariances) between components of the dynamic model (e.g., derivatives/time-lagged observations) are the same across items, and whether the other values of the model components (e.g., initial value, perturbations) are the same above and beyond the dynamic relations. This presentation will highlight examples and consequences of items that do not have dynamic invariance, and explore substantive data for dynamic invariance.

Bayesian Estimation of Factor Models Characterizing Dynamics
Ascher Munion, Pascal Deboeck, Jonathan Butner

Bayesian estimation offers a wide array of contributions to modern dynamic multivariate modeling, as well as a history of integration with topological model estimation. Capitalizing on the intersection of this history, the utility of considering invariance of dynamic will be explored. We begin by presenting a Bayesian extension of the previously presented dynamic equivalence, based on a Latent Differential Equations model, including a discussion of applications for Bayesian approximate invariance. Then a brief review of alternative multivariate Bayesian dynamic models (e.g., dynamic factor models, network models, state-space models) will be presented, and the ways in which dynamic equivalence could be accessed within these models will be discussed. Potential advantages and limitations of Bayesian estimation are presented in terms of model parameterization, tests for equivalence and missing data management. Finally, we discuss potential utilities and limitations of tests of true topological equivalency, compared to dynamic equivalency, will be discussed, based on recent advances from computation topology and persistent homology.

Session 4D: Advances in Modeling for Causal Inference (Room 206)

Exploring Model-Based Causes for Effect-Size Shrinkage in Educational Research
M. Shane Tutwiler, Michael Carlozzi, Zoe Kao

In educational research, observed effect sizes often shrink as sample sizes increase; a phenomenon often attributed to design factors such as fidelity of implementation (Wolf & Harbatkin, 2023). In this study, we simulated cluster-randomized trials across a range of sample sizes (varying the number of clusters from 12 to 180) and showed that the effect-size shrinkage was recovered as an artifact of statistical power and selection via null-hypothesis significance testing, a phenomenon known as Type M Error (Gelman & Carlin, 2014). Results of this research should compel applied educational researchers to invest more resources into ensuring adequate statistical power and utilizing modern analytic approaches before extending additional resources toward more nuanced components of their designs.

Evaluating the Impact of Analytic Approaches in a Multilevel Regression Discontinuity Application
Jason Schoeneberger, Chris Rhoads, Faeze Safari

Regression discontinuity (RD) designs are increasingly used for causal evaluations. Current best practice for estimating treatment effects from RD designs suggests using a data-driven bandwidth (Cattaneo & Vazquez-Bare, 2016). However, existing work on bandwidth identification has focused on single-level data or has used cluster-robust standard errors to correct for correlated errors. No existing work has explored how asymptotically optimal bandwidth selectors would perform in a multi-level modeling context. A recent U.S. General Services Administration request for application provides a useful application in that it outlined a scenario where schools were identified to receive an intervention based on a school-level assignment variable but outcomes were measured at the student-level. The current paper describes a simulation study conducted to determine the performance of various approaches to bandwidth selection and treatment effect estimation for RD designs with a multi-level structure and a level-2 assignment variable. Approaches were evaluated under simulation conditions with varying within-cluster sample sizes, level-2 sample sizes and variance components, as well as levels of covariation between cluster size and outcome variance component. Additionally, the importance of modeling a level-1 covariate highly correlated with the outcome (e.g. a pretest) is explored.

Session 4E: Tools for Teaching Multilevel Modeling (Room 301)

Using R to Enhance Shared Understanding of Linear Mixed Effect Models with and without Data
Katherine Zavez, Ofer Harel

In many settings, data is collected by taking multiple samples on the same individual or group. This type of data is often referred to as multilevel and can be analyzed using linear mixed effect models

(LMMs). LMMs can be fit in R using the lmer function in lme4 or the lme function in nlme. For a fitted lme/lmer model, there are functions for extracting information about the model variables in the parameters package, the model equation in the equatiomatic package, and the model structure in the nlme package. While this information is often useful prior to data collection (e.g., writing an analysis plan or running a power analysis), we were unable to identify equivalent functions that work without data. Therefore, we created a collection of three R functions (model2Variables, model2Equation, and model2Structure), called learn2LMM, that only require a lme/lmer model formula to produce similar outputs to that of the data-dependent functions. The hope is that these functions will inform model and variable selection decisions prior to model fitting, as well as encourage the development of similar R functions that are focused on explaining the statistics behind existing functions. Future work aims to make the learn2LMM an R package.

A Communication-Focused Approach to Building Path Diagrams for Multilevel Models
Jeffrey M. Girard

Path diagrams visually communicate the structure (and, optionally, the results) of complex analyses in the structural equation modeling (SEM) framework (e.g., path analyses, measurement models, and structural models) and are valuable tools for statistical design, teaching, and communication. Given the close relationship between the SEM and multilevel modeling (MLM; also called mixed effects or random coefficient) frameworks, efforts have been made to adapt path diagrams for MLM as well (e.g., Curran & Bauer, 2007). However, these techniques have not seen widespread adoption and it is still rare to find path diagrams provided alongside MLM results (unlike in SEM, where such diagrams are ubiquitous). I argue that a contributing factor to this state-of-affairs is that previous efforts have emphasized comprehensiveness of the diagrams over clarity of communication. To explore alternatives, I propose a novel path diagram format for MLM that emphasizes visual clarity, intuitiveness, and accessibility to a wider set of viewers (e.g., students, reviewers, and non-scientists less familiar with path diagrams). This format can be applied to any MLM model, but the benefits are most pronounced for complex models (e.g., with more than two levels and/or cross-classification). Feedback on this format is welcome as I prepare a manuscript describing it.

Concurrent Paper Session 5 Wednesday 9:00 – 10:00 am

Session 5A: Dealing with Daily Data (Room 201)

How We Cycle: A Tutorial on Combining Day-to-Day Dynamics with Day-of-Week Effects and Weekly Dynamics
Mohammadhossein Manuel Haqiqatkhah, Ellen L. Hamaker

Our emotions follow the weekly rhythm of our activities, which is reflected in daily diary data in two ways: a) the day-of-week effects (DOWEs) as weekday-specific average levels of emotions; and b) week-to-week dynamics as lagged effects of weekday-specific activities on the same weekday a week later. The main body of psychological literature so far has either studied DOWEs in isolation from lagged dynamics or studied day-to-day dynamics using (lag-1) autoregressive models in isolation from DOWEs and other forms of dynamics (e.g., moving average and week-to-week dynamics)—and a few have combined the two approaches. Using a rich daily diary dataset, we demonstrate that the majority of individuals show a combination of DOWEs and various forms of day-to-day and week-to-week dynamics, which the current practices in psychological research fail to account for. We thus miss a lot of substantively relevant information about psychological processes and face model misspecifications and erroneous conclusions. We explain how various aspects of dynamics and patterns in empirical data can be visualized, introduce a family of seasonal autoregressive-moving average (SARMA) models that can accommodate them, discuss their behaviors and properties, and provide a tutorial on fitting them to empirical data in R.

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

Madison Shea Smith, Susan C. South

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.

Session 5B: Variable Selection in Building Generalized Linear Mixed Models (Room 202)

A New Algorithm for Variable Selection in Building Generalized Linear Mixed Models

Yutian T. Thompson, Yaqi Li, Hairong Song, David E. Bard

Variable selection is a novel way to prevent generalized linear mixed models (GLMMs) from model overfitting, nonconvergence, low external validity and estimation biases. Among selection approaches for GLMMs, the method of regularized parameter estimation has shown superiority in jointly selecting both fixed and random effects, but challenges such as high computational cost, the outnumbered predictors problem, and multicollinearity have created barriers to its wider usage in practice. We propose a new algorithm, called random rPQL that incorporates a regularized penalized quasi-likelihood estimation (rPQL) with a stability selection process to overcome the aforementioned challenges. Simulation results indicate that the random rPQL algorithm can select fixed and random effects with high accuracy and efficiency, even when the number of candidate variables exceeds within-group observations or when severe multicollinearity exists among predictors. Additionally, we present a new R package that allows researchers to appropriately and pragmatically select variables in GLMMs.

A New Method for Variable Selection in Building GLMMs with Incomplete Data

Yutian T. Thompson, Yaqi Li, David E. Bard

Addressing the challenge of missing data in variable selection, particularly in Generalized Linear Mixed Models (GLMMs), is an area with limited literature. Incomplete data hinder the accuracy in the selection of both important fixed and random effects. Despite the implementation of multiple imputations that can help to hold missing data, how to pool the selection results from imputed datasets will turn out to be a new issue. To our knowledge, a method to conduct a variable selection, and at the same time handle missing data in GLMM has not been well established. In the current study, we proposed MI-random-rPQL to perform variable selection in GLMMs with incomplete clustered data. The MI-random-rPQL integrates Random rPQL, a method that enables to accurately select important fixed and random effects in GLMM for complete data, with multiple imputations. The results from simulation studies demonstrated that the proposed method has achieved great success in selecting the important fixed and random effects when the data is incomplete.

Session 5C: Multilevel R-Squared Measures in Stata (Room 205)

Using the mlmr2 Package to Estimate Multilevel R-Squared Measures in Stata

Anthony J. Gambino

Rights and Sterba (2023) provided a comprehensive framework for computing R-squared measures for multilevel models with theoretically any number of nested levels. This presentation will introduce a Stata package I've written called mlmr2, which is currently available to all Stata users through

the SCC (simply run "ssc install mlmr2" in Stata). This package provides users with the mlmr2 command, which can be used after estimating a model with Stata's mixed command to compute all of the relevant R-squared measures from the Rights and Sterba framework for the model along with interpretations of the measures. I will demonstrate how mlmr2 can be used to easily produce R-squared measures for multilevel models, compare competing multilevel models using these measures, and produce bootstrap confidence intervals for these R-squared measures.

Session 5D: Disentangling Person-Dependent and Item-Dependent Causal Effects (Room 301)

Disentangling Person-Dependent and Item-Dependent Causal Effects: Applications of Item Response Theory to the Estimation of Treatment Effect Heterogeneity

Joshua Gilbert, Luke Miratrix, Mridul Joshi, Benjamin Domingue

Analyzing heterogeneous treatment effects plays a crucial role in understanding the impacts of educational interventions. A standard practice for heterogeneity analysis is to examine interactions between treatment status and pre-intervention participant characteristics, such as pretest scores, to identify how different groups respond to treatment. This study demonstrates that identical observed patterns of heterogeneity on test score outcomes can emerge from entirely distinct data-generating processes. Specifically, we describe scenarios in which treatment effect heterogeneity arises from either variation in treatment effects along a pre-intervention participant characteristic or from correlations between treatment effects and item easiness parameters. We demonstrate analytically and through simulation that these two scenarios cannot be distinguished if analysis is based on summary scores alone as such outcomes are insufficient to identify the relevant generating process. We then describe a novel approach that identifies the relevant data-generating process by leveraging item-level data. We apply our approach to a randomized trial of a reading intervention in second grade, and show that any apparent heterogeneity by pretest ability is driven by the correlation between treatment effect size and item easiness. Our results highlight the potential of employing measurement principles in causal analysis, beyond their common use in test construction.

Concurrent Paper Session 6 — Wednesday 10:20 – 11:50 am

Session 6A: Bayesian Models (Room 201)

Comparing the Accuracy of Three Predictive Information Criteria for Bayesian Linear Multilevel Model Selection

Sean Devine, Carl F. Falk, Ken A. Fujimoto

Due in part to recent advances in software, such as Stan and the popular brms package in R, Bayesian multilevel modeling techniques have become increasingly popular. As researchers leverage these techniques to fit new models, information criteria—fit indices which provide information about a model's fit to the data—play an important role in disambiguating between competing models. However, a systematic evaluation of these Bayesian criteria in a multilevel context has not yet been undertaken. Using simulation, we investigate the model selection accuracy of three popular information criteria: the deviance information criteria (DIC), Watanabe-Akaike information criterion (WAIC), and an approximation to the leave-one-out cross-validation information criterion (LOO-CV). We manipulated the following factors to determine how they affected these indices' accuracy in identifying the data generation model: 1) the number of groups, 2) number of observations, 3) variance of the random effect, 4) fixed effects magnitude, 5) model misspecification, and 6) model selection strategy. In general, WAIC and LOO-CV outperformed DIC and are recommended when computationally feasible. In addition, we argue that a selection strategy that simply chooses the model with the lowest information criteria—a practice that is frequently employed in applied research—may result in overfitting.

Modeling Misspecification as a Parameter in Bayesian Structural Equation Models
James Uanhoro

Addressing model misspecification in Bayesian structural equation models is an area of active investigation. We review a set of related Bayesian SEMs that (i) assume misspecification to be part of the data generation process, and (ii) incorporate a parameter for quantifying the extent of misspecification. This parameter offers a direct measure of absolute model fit (the CRMR or RMSEA) and facilitates comparisons among models fitted to identical data sets. Furthermore, the models enable the examination of local misspecification. And differently from extant SEMs, parameter uncertainty reflects the degree of model misspecification in addition to sampling error. Through extensive simulation studies, we validate these capacities of the models and demonstrate the models' effectiveness in accurately estimating structural parameters and supporting inference. We end with a demonstration of how to implement the approach using the minorbsem package in R.

Bayesian Semiparametric Item Response Theory Models: A Methodological Illustration
Meng Qiu, Sally Paganin

In item response theory (IRT) modeling, the typical assumption is that subject-specific latent traits follow a normal distribution, which is often violated in reality. Violation of this normality assumption can lead to biased parameter estimates. To relax this restrictive assumption, we introduce the use of a Bayesian nonparametric approach, specifically a Dirichlet process mixture (DPM), to approximate the latent trait distribution as a mixture of normal distributions. An advantage of a DPM over a finite mixture is that the number of mixture components can be automatically determined, rather than relying on class enumeration based on information criteria such as Akaike's Information Criterion. In this methodological illustration, we first introduce the idea behind DPM and explain its construction based on the stick-breaking process and the Chinese restaurant process. Second, we demonstrate the DPM-based semiparametric IRT model using a 2-parameter logistic model and how model identification can be achieved through constraints on either item parameters or latent traits. Third, we illustrate the implementation of the model in NIMBLE package. Finally, we use simulated datasets to demonstrate the performance of the model.

Session 6B: Novel Applications of Latent Variable Modeling (Room 202)

Decomposing the Effects of Suffering on Depression Using a Reparameterized SEM and Penalized Maximum Likelihood
Noah Padgett, Richard Cowden, Tyler J. VanderWeele

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. We present results from a novel decomposition of latent factor effects on future outcomes. The decomposition utilizes the theory of potential outcomes under multiple versions of treatment (VanderWeele & Hernan, 2013) to define the effects of factors and residuals on future outcomes. The decomposition of effects was attainable by utilizing penalized maximum likelihood to obtain identified effects and standard errors from the uniquely defined model. Our presentation will emphasize how modern approaches to estimating traditionally unidentified models can lead to conceptually meaningful results about the longitudinal relationship between suffering and depression.

Intensive Longitudinal Modeling of Big Social Media Data
Jeffrey M. Girard

This talk will discuss two recent papers in which we longitudinally analyzed the same dataset of professional Instagram influencers (13,606,905 uploads over 925 days from 5,469 influencers in 76 different countries) to explore two very different research questions. In the first paper, written for an audience in marketing science, we explored which influencer behaviors were associated with growth over time in number of followers. We used log-linear growth curve modeling within the multilevel modeling

framework to do so. In the second paper, written for an audience in psychological science, we explored the impact of COVID-19 on weekly and yearly rhythms in the influencers' smiling on Instagram. We used computer vision algorithms to computationally estimate the presence and intensity of smiles in each image uploaded by the influencers, and we used a two-level autoregressive model with moderated periodic effects within the dynamic structural equation modeling (DSEM) framework to explore the research question. In presenting both papers, we will describe the results and their implications for marketing and psychological science, but my main emphasis will be on the challenges faced while modeling this large and complicated dataset and the solutions we used.

Validation of the Evidence-Based Practices Attitudes Scale (EBPAS) using Dynamic Fit Index Cutoffs

Julian M. Hernandez-Torres, Natalia Giraldo-Santiago, Daniel McNeish

There is a lack of validated instruments and resources to measure the capacity to deliver and implement evidence-based interventions in Puerto Rico. This study investigated the factorial structure EBPAS in a sample of mental health professionals in Puerto Rico. Most studies use rules of thumb to evaluate fit, but their adequacy varies across different model characteristics. Dynamic fit index (DFI) cutoffs are a novel method of assessing model fit in Confirmatory Factor Analysis (CFA). The EBPAS-15 was administered to a sample (n=237) of Puerto Rican social workers, counselors, and psychologists working in diverse settings (e.g., schools, clinics, and communities). DFI cutoffs were derived for CFA model estimated using Diagonally Weighted Least Squares. A correlated-factor model with four latent factors adequately fit the data. DFI demonstrated misspecification consistent to that of an incorrectly omitted cross-loading was present. These results point to a degree of complexity not being fully captured by the model. This novel fit evaluation method allows for a more comprehensive assessment of a scale's strengths and limitations, as well as richer discussions to further explore alternative factor structures.

Session 6C: Reproducibility, Replicability, and Registration of Simulation Studies (Room 205)

Why Do Reproducibility and Replicability of Simulation Studies Matter?

Jessica Kay Flake

Data fraud cases, failed replications, and a greater understanding of how research practices can be exploited to achieve desired results have inspired a methodological reform movement. Reforms are aimed at making it easier to evaluate, quality check, replicate, and build off previous research. Though framed as a problem for substantive research, a small body of literature is emerging about how these reforms are needed in methodological literature. I will discuss how publication bias, reproducibility, and replicability are relevant problems for the methodological research community. Building off scholarly work from biostatistics and computational modeling, I will overview reforms that the methodological research community can take in brief, as a precursor to the more in-depth talks in the symposium.

Replicating Simulation Research: A Case Study

Tristan D. Tibbe

Replication studies are often thought of in the context of substantive research, with research teams adhering to the sampling strategies, study protocols, and analysis plans of published projects as closely as possible to see if they reach the same conclusions. Although simulation studies that evaluate statistical methods involve synthetic data generated by the research team itself, these studies can (and should) be replicated as well. In this talk, I will describe my own experience replicating a highly cited simulation study, I will discuss the process of replication from gleaning information in the original article to running my own simulation. I also highlight aspects of the published study that both facilitated and hindered replication. Attendees can use the information to conduct their own simulation replication and write their own simulation articles in a way that ensures future researchers can replicate their findings.

Registered Reports for Simulation Studies

Amanda Kay Montoya

Registered reports are a new publication method where the decision to publish a paper is made prior to data collection/analysis using a two-stage submission process. Registered reports are new for simulation studies and the benefits of this approach may be distinct from those posited for human-subjects or animal research. This talk provides recommended best practices for undertaking registered reports for simulation research based on first-hand and editorial experience. The registered reports process highlights the importance of planning how simulation results will be interpreted, acknowledgement of constraints on generality of simulation research, and avoiding questionable simulation practices. I provide guidance for how to prepare a Stage 1 submission, common issues which arise in peer-review, and how to plan for the adjusted workflow of registered reports. Additionally, I provide guidance for peer reviewers of registered reports: what aspects of the study should be evaluated at each stage, and how the peer review process differs under this new publication workflow.

How to Register Your Simulation Study: Our Lessons Learned

Lindsay Alley, Mairead Shaw

Preregistering or submitting a simulation study as a registered report is a complex undertaking with many possible approaches. To illuminate some of these possible approaches, we present concrete examples from our experiences registering two simulation studies: one focused on multilevel modeling, and one on invariance testing under model misspecification. We will discuss the decision points involved in registering a simulation study, the biggest challenges that we encountered, and ways that the registration of simulation studies differs from registration of other kinds of research. The registration process entails three broad steps: (1) developing a research and analysis plan; (2) communicating this plan, using narrative and code, so that every step is clear to reviewers; and (3) considering how you will interpret future results. For each step of this process, we describe the approach we took in our projects and share specific examples from our code and stage 1 manuscripts. With these concrete examples from our experiences, attendees can learn from our mistakes and successes to try out registering their simulation studies.

Session 6D: Missing Data: Problems and Solutions (Room 301)

Evaluating the Effect of Change on Change in Cross-Domain Latent Growth Curve Analysis with Missing Data

Parisa Rafiee, Manshu Yang

Cross-Domain Latent Growth Curve (CD-LGC) modeling is a method of analyzing longitudinal data where the goal is to determine whether the change trend over time in an outcome is associated with the change trend in a time-varying predictor. An almost inevitable challenge in CD-LGC is the occurrence of missing data. If inappropriately handled, missing data can result in biased and less precise estimates, reduced statistical power, and diminished generalizability of study results. Although the Full Information Maximum Likelihood (FIML) estimation has been widely used for conducting CD-LGC analyses, its performance in the presence of missing data has not been systematically examined, especially when the missingness occurs in the time-varying predictor. The Sequential Full Bayesian (SFB) estimation is another method for dealing with missing data in CD-LGC analyses and has been increasingly used in recent years. Via a simulation study, we aimed to evaluate the performance of FIML and SFB estimation for handling Missing at Random (MAR) data in CD-LGC analysis, across conditions with different missing data scenarios and varying sample sizes, numbers of measurement occasions, and missing data proportions. Results suggest that SFB outperformed FIML in estimating parameters from CD-LGC models, especially when the time-varying predictor had missing values.

A Solution for Including Auxiliary Variables with Categorical Dependent Variable Estimation in SEM

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

Full information maximum likelihood (FIML) estimation for structural equation models (SEMs) provides preferable parameter estimates and standard errors compared with other possible missing data treatments (e.g., listwise deletion). When data are not at least missing at random (MAR), biased estimates or standard errors may occur, but these biases may be mitigated or partially mitigated when auxiliary variables are taken into account. Software for SEM, such as Mplus or the R package lavaan have features that allow for the inclusion of auxiliary variables in missing data estimation, but these features are limited to use only with models with continuous dependent variables. Based on the "extra DV model" approach suggested by Graham (Graham et al., 1997; Graham, 2003), we describe a method of including auxiliary variables in SEMs with binary dependent variables. Using a simulated data set and a latent growth curve model with binary dependent variables, we illustrate this new approach and show that the mean and variance estimates and standard errors for the intercept and slope parameters are equal to the results when using the auxiliary software feature. We then illustrate a growth curve model with a binary variable for the likely diagnosis of cognitive impairment over 10 waves.

Dummy Variable Adjustment Technique: An Alternative to Maximum Likelihood and Multiple Imputation

Roula Aldib, Lee Branum-Martin

This study examines the efficacy of the Dummy Variable Adjustment (DVA) technique in handling missing data, as compared to the more commonly used Maximum Likelihood (ML) and Multiple Imputation (MI) methods. Despite historical criticism of DVA for biasing regression estimates (Jones, 1996), a recent discussion by Allison (2022) suggests there is a place for DVA in certain study designs where some variables or survey items may not be applicable to some participants. In particular, participants may respond that an item does not apply to them and is therefore a "valid skip." Such data are not missing at random but are systematically missing based on participant characteristics. Using the Program for the International Assessment of Adult Competencies (PIAAC), this study compares DVA, ML, and MI in addressing valid skips or "does not apply" instances. Preliminary findings suggest that baseline regression parameters are underestimated under the assumption of missing at random in ML and MI, and they are more appropriately modeled through DVA. Therefore, it is crucial to further examine whether DVA may reveal information not captured by ML and MI, particularly in cases where "valid skip" responses provide substantive information related to the outcome variable.

Concurrent Paper Session 7 — Wednesday 1:00 – 2:30 pm

Session 7A: Modeling Individual Differences (Room 201)

Measurement Invariance of the Big Five across Socioeconomic Background: Multigroup Confirmatory Factor Analysis and Alignment Optimization

Emilija Meier-Faust, Sandra Bohmann

The Big Five personality dimensions are related to a variety of life outcomes. Most investigations use the Big Five without explicitly addressing whether the Big Five can be measured invariantly in diverse samples. Previous investigations found evidence for measurement non-invariance in relation to respondents' sociodemographic and socioeconomic characteristics. The current investigation tested the measurement invariance of a widely used short questionnaire â€“ the BFI-S â€“ regarding different aspects of respondents' socioeconomic background: education, occupational prestige, income, and cultural capital. Using data from the German Socio-Economic Panel (SOEP) in a sample (N = 16,298) interviewed in 2017, we applied multigroup CFA testing as well as the recently recommended alignment optimization (Asparouhov & Muthén, 2014). Results indicated scalar non-invariance of Openness and

Neuroticism across all indicators of socioeconomic background, of Conscientiousness across four out of five indicators of Agreeableness across three indicators, and Extraversion across one. Four items with substantial group differences in intercepts were identified â€“ two for Openness and one for Extraversion and Neuroticism each. Approximate alignment could not be reached for these items. These results offer guidance for modelling the Big Five in diverse samples and highlight the importance of taking differences in socioeconomic background into account when constructing personality questionnaire.

Applying SEM Based Person-Fit to the Wechsler Adult Intelligence Scale IV Demonstrate the Validity of Measurement at the Individual Level

Jared Block, Steven Reise

The WAIS-IV is one of the leading tests of cognitive ability in adults and is used alongside other measures to support in the diagnoses of individuals during a neuropsychological examination. Using person fit statistics which measure how well an individual's response pattern aligns with model expectations, we can assess the validity of measurement at the level of the individual. Applying this measure to a real clinical sample, we find that both the test and statistic perform as expected with most individuals' response patterns aligning with the strong correlations between cognitive domains that is found across neurocognitive exams. However, a select number of individuals failed to fit the model, indicating that their scores should be interpreted with extreme caution. This application provides evidence for the potential utility of such person fit measures in order to assure that scores are valid at both global and individual levels.

Session 7B: Network Applications (Room 202)

Exploring the Emotional Well-Being of Young Adults through Network Psychometrics

Daniel Hernández-Torrano

The importance of emotional well-being (EWB) in public health is increasingly recognized. However, research has been delayed due to the lack of consensus on the fundamental components of EWB, effective ways to measure them, and mechanisms to inform intervention theory. The present study aims to advance the conceptualization and measurement of EWB by applying a network psychometrics modeling approach to estimate and visualize the network structure of experiential and reflective features of EWB (e.g., positive emotions, life satisfaction, sense of meaning, and goal realization) in an incidental sample of young adults. The results extend previous studies relying on traditional latent variable approaches (e.g., CFA), enriching the understanding of the complex relationships between the different facets of EWB and better explaining how individuals feel generally and about life overall. Moreover, the study identifies the most essential EWB components that are optimal target indicators to inform intervention theory and practice that promote positive states of mind.

Exploring the Dynamics of Motivation in Physical Activity among Older Adults Through Panel Network Approach

Tommaso Palombi, Denny Borsboom, René Freichel, Elisa Cavicchiolo, Fabio Lucidi, Fabio Alivernini

Self-Determination Theory distinguishes different forms of human motivation: intrinsic motivation and extrinsic motivations (external, introjected, identified, integrated). This study explores the dynamic interplay of these different forms of motivation in physical activity among older adults. The research involved 384 older adults, collecting data over a month through four assessment waves. The data were analyzed using graphical vector autoregressive (GVAR) models that allowed the separation of within- and between-person effects and the exploration of temporal associations. Centrality measures were employed to determine the influence and predictability of various motivational factors. The findings reveal a complex interplay of motivational states over time. Intrinsic motivation was uniquely important in predicting physical activity, suggesting the need for personalized, enjoyment-focused strategies to encourage physical activity among older adults. The study highlighted the dynamic nature of motivation in this age group, with significant transitions observed between different forms of motivation, such as

from external to introjected, and introjected to integrated regulation. Particularly noteworthy is the centrality of integrated regulation that aligns activity with personal values and goals, playing a crucial role among older adults. The study underscores motivational states' complexity and temporal dynamics, offering valuable insights for designing more effective and sustainable health promotion strategies.

Estimating the Group Differences of Longitudinal Network Analysis: An Example of Eating Disorder Psychopathology

Jihong Zhang, Jinbo He

This study, titled "Estimating the Group Differences of Longitudinal Network Analysis: An Example of Eating Disorder Psychopathology," employs multigroup longitudinal network analysis to examine gender differences in the development of disordered eating behaviors. The data comes from a 18-month longitudinal study of 1540 students, the research integrates measures of eating disorders, interpersonal problems, and emotion dysregulation. The analysis compares the temporal network structures of two gender groups through likelihood ratio tests and estimates gender differences in centrality measures via bootstrapping. Results indicate significant gender disparities in centrality indices of temporal networks. Specifically, behaviors like vomiting or taking laxatives (EDE-VT) show lower InStrength in boys than in girls, while girls exhibit higher OutStrength in behaviors such as long periods without eating (EDE-WE) and food preoccupation (EDE-FP). Additionally, certain nodes like Awareness (Awr) and Goals (Gls) from the Difficulties in Emotion Regulation Scale, and Weight/Shape Preoccupation (EDE-WP) and Long periods of binge eating (EDE-BE) from the Eating Disorder Examination Questionnaire, demonstrate higher bridge strength in boys' networks. The study's findings reveal critical gender-specific patterns in the development of disordered eating behaviors, highlighting the need for tailored intervention strategies. The observed differences in network structures and central symptoms between boys and girls offer valuable insights for more effective prevention and treatment approaches in addressing eating disorders.

Session 7C: Advances in Multilevel Mixture Modeling (Room 205)

Examining the Effect of Nested Data on Class Enumeration and Model Fit in Latent Profile Analysis

Angela Starrett, Katherine Masyn

Increased interest in Latent Profile Analysis (LPA) stems from its ability to discern useful subgroups for targeted interventions, manage multidimensional data, and simultaneously model complex interactions. This approach is particularly pertinent in educational settings, where data often display nested structures, such as students within classrooms and schools. With more traditional latent variable models, nesting can potentially distort parameter estimation and bias standard errors; however, the specific effects of nesting LPA have not been thoroughly investigated. Prior research has explored mixture modeling with nested data, notably in latent class analysis (Kaplan & Keller, 2011) and growth mixture modeling (Chen et al., 2017), but there remains a research gap regarding LPA in nested contexts.

Literature on LPA methodology highlights various approaches to handling nested data, ranging from disregarding it to adopting multilevel models. Mplus, a software known for its robust capabilities in managing complex data structures, is frequently used for LPAs. Within Mplus, the "type=complex" option is utilized for analyzing complex survey data, adjusting standard errors and chi-square statistics to reflect the nested nature of the data. However, the effectiveness of this option specifically for LPAs with nested data is yet to be established. This study evaluates the impact of nesting on enumeration and model fit in the context of LPA. The research will simulate continuous data from four indicators across three known subpopulations within a two-level model. The simulation focuses on three design factors: conditional intraclass correlation (ICC), degree of separation, and mixing proportions. These factors are crucial as they significantly influence class enumeration and the likelihood of convergence in LPAs (Henson et al., 2007; Tofighi & Enders, 2008). A total of 13,500 datasets (27 conditions * 500 replications) were generated, varying ICCs (.05, .15, .25), mixing proportions (balanced, unbalanced, and

unbalanced with one class less than 10%), and degree of separation (all well separated, none well separated, and one class well separated from two not well separated classes). The study employs robust maximum likelihood estimation to compare estimation convergence, performance of fit indices for class enumeration, and quality of parameter estimates across three scenarios: the true model, a model ignoring nesting, and a model using the "type=complex" option in Mplus to determine the efficacy of the type=complex option for accurate class enumeration and parameter estimates in LPAs with nested data. It also assesses the option's ability to replicate model separation and homogeneity. The findings provide valuable insights into best practices for conducting LPAs with multilevel data in Mplus, contributing to the existing body of knowledge and methodology in this area.

Application of a Multilevel Latent Class Analysis with Cross-Classified Data
Audrey Leroux, Katherine Masyn

Despite the ubiquity of multilevel data in these same settings, multilevel modeling techniques for Latent class analysis (LCA) are still in their infancy, with the initial applications of multilevel LCA occurring only within the last 15 years (e.g., Henry & Muthén, 2010). More complicated multilevel modeling extensions involving cross-classified data structures (Leroux & Beretvas, 2022) has never been applied to LCA. This paper demonstrates the specification and estimation of a cross-classified multilevel LCA (CC-LCA) model using real data and explores the consequences of ignoring the cross-classified data structure for measurement and structural parameter estimates as well as for individual classification.

Data for this study are from the Longitudinal Study of American Youth (LSAY; Miller). Following prior examples of latent class analysis of high school students' attitudes towards mathematics and science (Ing & Nylund-Gibson, 2013, 2017; Masyn, 2017), we used ten dichotomized student questionnaire items about attitudes towards science and math as indicators of a 4-class latent multinomial variable. Students are also clustered within high school and their prior middle school attended, thus resulting in a cross-classified data structure. Student sex (binary) was considered as a student-level predictor of latent class membership, while school-level predictors such as enrollment and urbanicity were considered as middle and high school predictors of latent class. We use a cross-classified random effects extension of the multilevel LCA assuming cluster effects can be adequately represented by latent class random effects, i.e., randomly-varying multinomial intercepts for the distribution of the latent class variable with fixed thresholds for the latent class indicators. Preliminary findings suggest that failure to account for the cross-classified data structure in a multilevel LCA produce biased variance estimates for the random multinomial intercepts, consistent with previous work on cross-classified modeling (Luo & Kwok, 2009; Meyers & Beretvas, 2006). Furthermore, latent class composition at the student level is distorted. Implications of our findings and future extensions into latent transition models with complex nesting are discussed.

Moderated Nonlinear Mixture Analysis for Longitudinal Invariance Testing in Latent Transition Analyses
Katherine Masyn, Boshi Wang

In the last decade, mixture modeling methodologists have been actively playing catch-up to the vast and ever-growing psychometric research corpus regarding measurement invariance (MI) and differential item functioning (DIF) in the SEM and IRT spheres, e.g., Cole (2019); Finch (2015); Masyn (2017); Olivera-Aguilar and Rikoon (2018); and Vermunt and Magidson (2021). Most of these techniques involve adapting one of the more established SEM-based MI testing approaches (i.e., MIMIC or multiple-group modeling) to the mixture modeling in a cross-sectional data setting. However, scant work has been done adapting longitudinal SEM MI testing to longitudinal mixture models, such as latent transition analysis (LTA) (cf., Di Mari, Dotto, Farcomeni, and Punzo, 2022). Additionally, there is a newer approach to MI testing in SEM—moderated nonlinear factor analysis (MNLFA) (Bauer, 2017)—that has yet to be adapted for mixture models or compared to MIMIC and MG MI testing. Initially developed in the context of integrative data analysis, MNLFA has been shown to offer greater flexibility and overcome many of the oft-unacknowledged limitations of the MIMIC and multiple-group (MG)

approaches to MI testing. This paper addresses those two gaps by proposing a modification of MNLFA, moderated nonlinear mixture analysis (MNLMA), for longitudinal MI testing in LTA. We compare two alternate modeling processes for MNLMA. In the first variation, directly translating the MNLFA steps for longitudinal data, a calibration sample is selected using a randomly selected time point of measure for each individual. The MNLMA is done on the calibration sample for MI testing, including time as one of the individual-level sources of DIF. The estimated measurement parameters (including the DIF effects of all covariates) are fixed and then the full sample is "scored" using the resultant model. Finally, the scores (modal class assignments and classification error) are used with a 3-step LTA (a la Nylund-Gibson, Grimm, Quirk, and Furlong, 2014) to estimate the latent transition probabilities. In the second variation, a fully multilevel MNLMA is specified, with time as a within-level source of DIF and individual-level covariates as between-level sources of DIF. The scores from the multilevel MNLMA are then used with a 3-step LTA. The proposed model variations will be illustrated with data from the Trajectories in Early Career Research: Data Repository (Feldon, Litson, Roksa, and Griffin, 2023). We apply our two alternate MNLMA approaches to test for longitudinal MI and then estimate LTAs for the first six years of doctoral study. We compare our LTA results, accounting for DIF, to three LTA models more likely to be found in the current applied literature: (i) LTA assuming longitudinal MI; (ii) 3-step LTA assuming full non-invariance; and (iii) random intercept LTA (RI-LTA) assuming longitudinal MI.

Session 7D: Structural Equation Modeling Methods (Room 301)

Comparing Approaches to Examine Multiple Binary Moderators in Latent Variable Models
Kaylee Litson, Amanda Kay Montoya, Yiwei Wang

Many methods currently exist to examine interaction effects that include latent variables, yet these methods have not been widely adopted, despite the prevalence of latent variable methods in social-behavioral research. In practice, researchers often aim to examine multiple moderators simultaneously, which can increase the complexity of latent variable interaction models and lead some to wonder how to appropriately model these effects. The process of extending existing latent interaction methods to cases with multiple moderators has not been explained clearly or evaluated empirically. We demonstrate how to fit models that include latent interactions with multiple categorical moderators across seven methods: unadjusted means, reliability-adjusted means, product indicator (with and without constraints), multi-group (with and without invariance testing), and latent moderated structural equations (LMS). We demonstrate how to fit these models using an empirical dataset from the Emerging Adulthood Measured at Multiple Institutions 2 (EAMMI-2) data. We highlight strengths and weaknesses of each method and provide resources and example code for fitting latent moderation models with multiple moderators.

Quantile Structural Equation Modeling: Testing a Novel Distance Based Approach
Jeffrey Shero, Zhixin Zhu, Jessica Logan

Quantile-sensitive designs, like conditional and unconditional quantile regression, can assess how an association between constructs varies based on wherein the distribution of the outcome of interest an individual lies. These methods are useful for assessing associations that vary across the distributions of the variables analyzed but come with their own set of issues and considerations. Beyond these issues, these methods are limited in the ways they can be applied, with no current methods existing for assessing more complex structural equation modeling (SEM) based methods in a quantile-sensitive framework. The current paper aims to address this, developing, introducing, and empirically testing multiple approaches to constructing quantile-specific covariance matrices that can be subsequently applied to SEM models. These approaches included extensions of conditional and unconditional quantile regression, the estimation of quantile-specific variance terms and subsequent transformation into quantile covariance terms, and a novel approach extending Euclidean distance methods to allow for quantile-sensitive SEM analyses. Each method was tested and compared using a simulation design, simulating data with heterogeneous effects and applying each method. Results indicated that the newly developed Euclidean distance-based

extension was most effective in uncovering the simulated heterogeneous effects, effectively doing so for regression, factor, and other SEM-associated effects.

Concurrent Paper Session 8 Wednesday 3:00 – 4:00 pm

Session 8A: Innovations in Mixed Modeling (Room 201)

Model Selection of GLMMs in the Analysis of Count Data in SCEDs: A Monte Carlo Simulation
Haoran Li

Generalized linear mixed models (GLMMs) have great potential to deal with count data in single-case experimental designs (SCEDs). However, applied researchers faced challenges in making various statistical decisions when using such advanced statistical techniques in their own research. This study focused on a critical issue by investigating the selection of an appropriate distribution to handle different types of count data in SCEDs due to overdispersion and/or zero-inflation. To achieve this, I proposed two model selection frameworks, one based on calculating information criteria and another based on utilizing a multistage-model selection procedure. Four data scenarios were simulated including Poisson, negative binominal (NB), zero-inflated Poisson (ZIP), and zero-inflated negative binomial (ZINB). The same set of models (i.e., Poisson, NB, ZIP, and ZINB) for each scenario were fitted. In the simulation, I evaluated 10 model selection strategies within the two frameworks by assessing the model selection bias, and its consequences on the accuracy of the treatment effect estimates and inferential statistics. Based on the simulation results and previous work, I provide recommendations regarding which model selection methods should be adopted in different scenarios and discuss implications, limitations, and future research directions.

Latent Class Clustering of Random Coefficient Estimates Obtained from a Multilevel Analysis
Jay Magidson, Jeroen Vermunt

A common approach in marketing research is to estimate a multilevel choice model applying Bayesian MCMC to preference items elicited in an online experiment. This yields 10-20 random effect estimates per respondent (step one) which are then subjected to a cluster analysis to obtain market segments (step two). The question we address is whether such a two-step approach, where random coefficient estimates are used as indicators in a latent class (LC) cluster analysis, performs as well as applying a LC model to the original choices (one-step). Based on a small simulation study, surprisingly, the answer is yes. Other issues: To obtain segments differing in preferences, one needs to separate scale heterogeneity (similar to a response style) from preference heterogeneity. This was achieved by including latent scale classes in addition to the clusters. In the two-step approach, BIC yielded many more classes than true, whereas the use of CHull worked well.

Session 8B: Innovations in Structural Equation Modeling (Room 202)

Introducing the Deleted One Covariance Residual Measure to the Structural Equation Modeling
Fathima Jaffari, Jennifer Koran

This study introduces a model-free case influence measure (DOCR) to SEM and evaluates its performance compared to that of Mahalanobis Distance and Generalized Cook's Distance when the sample size, proportion of target cases to non-target cases, and type of model used to generate the data are manipulated. The findings suggest DOCR generally performed better than the other measures in identifying target cases across all simulated conditions. However, the performance of DOCR under small sample size was not satisfactory, and results suggested DOCR is sensitive to sample size. Recommendations for current use of DOCR and for future research are provided.

Ruling Out Latent Time Varying Confounders in Two-Variable Multi-Wave Studies
David Kenny, D. Betsy McCoach

There has been considerable interest in estimating cross-causal effects in two-variable, multi-wave designs. However, there does not currently exist a strategy for ruling out unmeasured time-varying covariates that may act as confounders. In this paper, we propose a new strategy for testing whether an unmeasured time-varying covariate could explain the covariation in the data. That model, called the Latent Time-Varying Covariate (LTVC) model, can be tested with observations of two variables across three or more times. We discuss the LTVC under the condition that certain parameters to be invariant across time. We explore factors that affect the power to reject the LTVC. We propose that researchers should attempt to rule out the LTVC model before estimating a model with causal effects between the variables. If the LTVC model can be rejected, the estimates from the cross-causal model become more plausible. Several extended examples of this approach are discussed.

Session 8C: Restructuring Basic Statistical Curricula: Mixing Older Analytic Methods with Modern Software Tools in Psychological Research (Room 205)

Restructuring Basic Statistical Curricula: Mixing Older Analytic Methods with Modern Software Tools in Psychological Research
Emil Coman, Dakota Cintron, James Jaccard, Sabrina Uva, Ana-Maria Cazan

We illustrate a modern approach to introduce statistical concepts that is centered on research methodology and structural thinking, rather than on a formal mathematical basis. We provide worked examples of basic statistical tests, their plain English inquiry formulations, and their graphical view in Onyx, Jamovi, R\lavaan, and dagitty.net, as simple path models. We provide the basics of path analysis, its tracing rule, and of covariance algebra rules. We provide quick visual calculations using the tracing rule for the regression coefficient in simple regression and the instrumental variable (IV) models, to illustrate its intuitive power compared to mathematical derivations. We finally present modern uses of the tracing rule in causal analytic works. We re-cast topics like instrumental variables, feedback-loop models, and "causal" mediation, generally believed to be "advanced" subjects (as path analysis and its descendant method structural equation modeling, SEM, are also improperly perceived) into a mold set by simple research questions like "Is religiosity influencing anxiety? Is this an association, or an actual causal effect? Is this effect stable, or does it vary? How? What mechanisms explain it?"

Session 8D: InclusiMetrics (Room 301)

Advancing Research on Methodology: A Panel Discussion on the Creation of a Minority-Centered Methodological Conference, InclusiMetrics
Marcus Harris, Zachary Collier

There exists a noticeable lack of diversity in our field, spanning those involved in creating, evaluating, and disseminating research and statistical methods across various levels—be it undergraduates and graduate students, professors, industry experts, departments of education, or think tanks. This panel discussion seeks to delve into the conception and establishment of a novel minority-centered methodological conference, Inclusimetrics. The overarching goals of this conference are to (1) actively recruit underrepresented minorities into Ph.D. Methodological programs, (2) formulate frameworks and perspectives for conducting rigorous methodological research, (3) elucidate the implicit rules and strategies for advancing careers as methodological researchers, and (4) foster networks that challenge the existing size, scope, scale, and significance of methodological research. At the core of the conference's name, this panel discussion will intricately explore the dimensions of diversity and inclusion, dismantling barriers while thoroughly examining the pivotal role and responsibility of non-minorities within our discipline. We invite you to join us in the creation of Inclusimetrics, where the vision is to reshape the landscape of methodological research through a lens of inclusivity and diversity.

Poster Abstracts

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

1. Incorporating Think-Aloud Interviews into Instrument Development: An Applied Example
Lindsay J. Alley, Cole Johnson, Kristy A. Robinson, Jessica K. Flake

Advice on the development and validation of psychological instruments emphasizes the importance of collecting validity evidence from varied sources (AERA et al., 2014), but response processes evidence is reported in less than 4% of studies (Cizek et al., 2008; Hogan & Agnello, 2004; Villalobos Coronel, 2015). However, response processes encompass "the mechanisms that underlie what people do, think, or feel when interacting with, and responding to, the item or task" (Hubley & Zumbo, 2017, p. 2), which cannot be gleaned from quantitative analyses alone. One approach to collecting response processes validity evidence is the think-aloud interview, where respondents state their thoughts out loud while completing the instrument of interest. After observing poorer than expected fit from confirmatory factor analysis, we conducted a think-aloud interview study to examine the validity of candidate items for a measure of student perceptions of motivational climate in STEM university classrooms. We demonstrate the kind of validity insights that can be gained using think-aloud interviews and how these insights can be used to aid instrument development and illuminate psychometric results.

2. Predictors of Cause Specific Survival and Prognosis of Pharyngeal Squamous Cell Carcinoma Using Type I Generalized Half Logistic Distribution: Comparison with Cox Regression Model and Random Forest
Phillip Awodutire, Michael Kattan, Oladimeji Akadiri

Objective: To develop and validate an accelerated failure time model to build a predictive model for cause-specific survival and prognosis of pharyngeal Squamous Cell Carcinoma.

Method: We screened pharyngeal cancer patients diagnosed with Squamous Cell Carcinoma from the Surveillance Epidemiology and End Results database between 2010 and 2020 with the patient's demographic characteristics (Age, gender, Race), tumor primary site, the total number of in situ malignant tumors, the month of diagnosis to treatment, clinical stage of the tumor, histological Grade of the tumor, and treatment modality as factors included for the study. An accelerated failure time model using the Type I generalized half logistic distribution (TIGHLD) was used to determine independent prognostic factors affecting the survival time of patients with pharyngeal squamous carcinoma. We used the Concordance Index (C-index) and Root Mean Square Error (RMSE) and calibration curve for predictability, to compare the TIGHLD model to the Cox proportional hazard and random forest survival models. All tests are conducted at a 0.05 level of significance.

Results: The accelerated failure time models demonstrated superior effectiveness in modeling (fit and predictive accuracy) the cause-specific survival (CSS) of pharyngeal squamous cell carcinoma compared to the Cox and random forest model due to lowest RMSE(0.12) and highest C-Index(0.875). All variables under consideration in this study demonstrated significance at the 0.05 level for cause specific survival of the patients.

Conclusions: This study shows how comparative evaluation of modelling techniques in large, clustered healthcare datasets may provide insight into relative strengths of different strategies for clinical prognostication. The accelerated failure time model using the TIGHLD provides a relatively accurate method to predict the prognosis of pharyngeal cancer in squamous cell carcinoma patients and can be preferred to the Cox proportional hazard (Cox PH) and random forest model.

3. Assessing Genetic Algorithms for Variable Selection in Predictive Modeling Based on Classification: Comparing Loss Functions and Internal Models through a Simulation Study
Catherine Bain, Dingjing Shi

In two-group classification predictive modeling, selecting variables is a crucial step, aiming to reduce the data collection burden, enhance model efficiency, and improve generalizability. Among the variable selection methods available to researchers, the genetic algorithm (GA) stands out, offering customization through various loss functions and internal models. While existing research assesses GA implementations with fixed loss functions and models, there is a gap in literature exploring the algorithm's independent performance as a variable selection strategy. This study employs a Monte Carlo simulation to evaluate the performance of the GA using various assessment measures. Within the GA, we compared three commonly-used classification internal models: support vector machines which could separate non-linearly separable data, random forest which could handle mixed typed (i.e., both binary and continuous) data, and logistic regression for predicting binary outcomes along with three loss functions: log loss, hinge loss, and the adjusted rand index. This research assesses the impact of loss functions and internal models on the performance of the GA as a variable selection method for two-group classification predictive models, significantly contributing to the field by clarifying whether the power of the GA lies in its selection strategy or in the internal elements of a specific implementation.

4. Dynamic Modeling of Physiological Reactions: A Markov Chain Approach to Affect Dynamics
Francesca Borghesi, Pietro Cipresso

Traditionally, affective states encompass behavioral, subjective, and neuropsychophysiological changes associated with emotional stimuli. Affect Dynamics field highlighted the need to consider time in analyses. However, the understanding of temporal and physiological dimensions in a lab setting remains underexplored. We developed an experimental design to expose individuals to all the possible transitions from one affect to another (indeed, affect dynamics) in a lab setting.

Participants were exposed to stimuli from the International Affective Picture System (IAPS) database, with images categorized into randomized arousal-valence blocks. We recorded psychophysiological responses, by using facial electromyography (EMG) for emotional valence and galvanic skin response (GSR) for physiological arousal. This approach allowed for capturing continuous emotional responses and analyzing affective state transitions in pairs.

To model these affect transitions, we utilized a Markov discrete chain. This method created a probability matrix for each participant, indicating the likelihood of moving from one affective state to another. Employing the Markov chain enabled us to quantitatively express physiological variability in emotional transitions, enhancing our understanding of the affect dynamics in a laboratory setting.

5. Missing Data Methods for Functional Causal Mediation Analysis with Applications to Smoking Cessation
Cody Campen, Donna Coffman

Ecological momentary assessment (EMA) techniques have enabled researchers to measure processes as they occur in real-time. In functional causal mediation analysis, EMA enables the modeling of time-varying mediators through functional regression techniques â€“ for instance, how pharmacotherapy and smoking cessation outcomes may be mediated day-to-day through participant self-efficacy. However, the intensive and intrusive nature of EMA creates a non-ignorable burden on participants, resulting in both substantial participant attrition and sporadic missingness. Multiple imputation procedures accommodating functional response models have only recently been developed, thus applied researchers have almost exclusively utilized complete case analyses. Here, we present an assessment of various state-of-the-art and common-practice methods for handling missing data in functional response models including a functional extension of multiple imputation by chained equations, complete cases analysis, and complete case analysis with participant compliance thresholds common in applied research. To assess these methods, we test performance across common missing data patterns including sporadic missingness and participant dropout as well as numerous EMA data characteristics

such as small and large sample sizes and varying rates of participant compliance. These results are illustrated in an application to a smoking cessation study investigating the time-varying mediating effect of self-efficacy on pharmacotherapy and cessation outcomes.

6. Investigating the Development of Academic Motivations in Adolescents Living in Poverty: A Latent Growth Model Analysis

Elisa Cavicchiolo, Sara Manganelli, Tommaso Palombi, Fabio Lucidi, Fabio Alivernini

The present study, rooted in Self-Determination Theory (SDT), aims to investigate the trajectories of academic motivations in adolescents from a very low-income background using latent growth modeling. The influence of parental and teacher support, as well as peer relationships, on these motivations was investigated. The sample comprises 228 adolescents from families living in poverty, a population that has been inadequately studied. Data collection occurred over four waves spanning a 2-year period. To assess changes in different types of motivations (i.e. intrinsic, identified, introjected and external), unconditional growth curve models were performed using a multiple indicator approach and latent variables. Major determinants of motivation according to SDT and sociocultural factors were examined through conditional latent growth models with time-invariant and time-varying covariates.

The results revealed a significant decrease in levels of identified regulation over time, while other forms of motivation remained stable. This decline was particularly pronounced among second-generation and male adolescents. Nevertheless, academic motivation benefits from parental and teacher support. Notably, peer acceptance appeared to sustain identified regulation, while peer friendships seemed to exert a negative impact. Overall, using a parsimonious variable-center approach, our study provides valuable insights into the nuanced development of academic motivation in adolescents living in poverty.

7. A Project-Based Instruction Approach to Improving Student Lunar Phases Learning Outcomes: A Quantitative Inquiry

Merryn Cole, Hongwei Yang, Jennifer Anne Wilhelm

We investigated how students' lunar phases learning outcomes were affected by student and teacher demographic characteristics (gender, race/ethnicity, spatial thinking ability, and content knowledge). The study identified moderately strong correlations both between students' spatial thinking ability and understanding of lunar phases, as well as between the pre-to-post-intervention scores of the two measures. Multilevel modeling showed significant predictors of learning outcomes from both student and teacher variables. This study furthers works on establishing a connection between student learning outcomes and the content knowledge and spatial ability for themselves as well as their teachers, and shows promise for a project-based instruction approach in aiding in lunar phases understanding.

8. The Impact of Group Size Ratio and Model Size on the Sensitivity of Fit Measures in Measurement Invariance Testing: A Monte Carlo Simulation Study

Ruiqin Gao, Christine DiStefano, Jin Liu, Ning Jiang, Jiali Zheng

A Monte Carlo simulation study was conducted to evaluate the impact of group size ratio in combination with model size on the sensitivity of the three fit measures (i.e., CFI, RMSEA, and SRMR), which are commonly used to detect a lack of metric invariance or scalar invariance. Design factors included model size, group size ratio, location of noninvariance, and location of noninvariant items. Results suggested that CFI failed to detect metric noninvariance when invariance tests involved a group ratio of 1:4 in combination with the location of non-invariant items only on one factor. RMSEA failed to detect either metric or scalar invariance with a large model size (e.g., 2 factors/ 30 indicators), especially when the group size ratio was 1:4 and the location of non-invariant items on one factor was involved. SRMR failed to detect metric invariance when a group ratio of 1:2 was involved. SRMR was not adequate for detecting scalar invariance. Recommendations for empirical researchers are provided based on the findings.

9. Airbag Moderation: A New Hypothesis and Range of Statistical Methods that Capture Many Hitherto Neglected Types of Process

James Hall, Lars-Erik Malmberg, Ariel Lindorff, Nicole Baumann, Pamela Sammons

The airbag in your car, your immune system, the provision of free meals in schools—these systems (and many others) are different yet have elements that operate in the same manner. Until recently however there was no methodological term that described this underlying common functionality and therefore also no commonly understood way to statistically model their effects. Instead, researchers both conceived of, and tested, only part of these systems (via application of the hypothesis, and statistical methods for, Moderation) with the consequence that this partial-testing has introduced sizeable gaps in knowledge within many fields of research.

This presentation: 1. introduces a hypothesis that describes these systems/functions, "Airbag Moderation", 2. Describes a range of statistical methods that exist to help empirically test for their presence; 3. Provides an empirical demonstration of data from 2608 families to show the effectiveness of UK Sure Start Children's Centres.

The take-home message is that: Airbag Moderation is a novel hypothesis that is demonstrably more suitable than Moderation for conceptualizing and testing a wide range of theories, interventions, and social policies. It is also easily implemented via existing statistical techniques. Potential applications of Airbag Moderation suggest future directions for research in substantive and methodological areas.

10. Selection Effects for Inequity in Education: Identifying and Evaluating a New Type of Educational Effect via Application of Airbag Moderation

James Hall, Lars-Erik Malmberg, Gregory Palardy

Selection effects in education—who goes where, gets what, and how much—play a pivotal role in the formation and persistence of educational inequities. These include inequalities from preschool to higher education and as regards both student progress and attainment as well as the funding and resources of teachers, institutions, and districts.

However, current statistical methods can struggle to simultaneously evaluate both the presence of a selection effect in education and its consequence for educational inequity. For example, where one group has a different access to an educational resource and this differential access may be the reason for group differences in educational outcome. Say for example geographic differences in access to educational opportunities.

This paper responds with two empirical examples that overcome current methodological difficulties via the application of the new Hypothesis of "Airbag Moderation". These examples use data from the USA and the UK and from two different phases of education (USA high school, UK preschool). Evidence is found of selection effects in education that both widen and narrow differences between groups of students in their education outcomes. Discussion focusses upon methodological requirements for extensions to this work, as well as implications for educational research, policy, and practice.

11. Artifact Corrections for Effect Sizes: Seeing Reality for What It Is

Matthew B. Jané, Edward Kroc, Brenton M. Wiernik, Frederick L. Oswald, Blair T. Johnson

Study artifacts are sources of contamination in research methodology that can induce severe bias in effect size estimates. My new book, Artifact Corrections for Effect Sizes is a living open-source book that covers various types of artifacts including measurement error, indirect/direct selection effects, artificial dichotomization, group misclassification, small sample bias, treatment noncompliance, and scale coarseness. Each artifact is discussed in detail regarding how it biases effect sizes (e.g., correlations and standardized mean differences), along with the equations and R code necessary to correct biased effect sizes. The book also discusses how to incorporate artifact corrections into meta-analyses to obtain unbiased estimates of the target quantity in the context of evidence synthesis.

12. What's the Issue with Cut-Scores? A QuantCrit Perspective

Madeline Klotz, Lori Skibbe, Ryan Bowles

A tenet central to the QuantCrit framework is the idea that numbers lack neutrality (Castillo & Gillborn, 2022). This framework encourages researchers to be thoughtful regarding the construction of their analytical models, paying attention to who is included (or not included) in their research, which impacts research findings, subsequent intervention, and policies resulting from the research. Fraught with subjective interpretation and biased labeling, cut-scores are commonly utilized on standardized assessments for determining intervention decisions. We consider the implications of using cut-scores when assessing children's phonological awareness (PA), as the understanding of the sound structure of language is a strong predictor of later reading. Composite PA scores are typically interpreted using cutoffs which provide labels that are used to make decisions about interventions. This is problematic due to variation across PA assessments in cutoffs and subsequent labeling, creating space for bias in both interpretation and intervention. To further exemplify the subjectivity of cut-scores, equipercentile test equating was used to put assessments on the same scale, allowing for direct comparison of cut-offs. The current analysis compares interpretational cutoffs across multiple standardized assessments of PA, all utilized when making intervention decisions, and identifies discrepancies and areas for subjective bias through a QuantCrit lens.

13. Moderator Under- and Overspecification in Multiple Regression Analysis

Noah Koehler, Wen Luo, Olukayode Apata

Research regarding under- and overspecification within the context of multiple regression models has typically focused on the wrongful inclusion or exclusion of individual predictor variables. However, little attention has been given to the analogous issues of under- and overspecification of moderator terms in these models. This gap is particularly relevant when researchers lack clear theoretical or empirical guidance on including moderator effects, potentially leading to under- or over-specification problems. The present study conducts a series of Monte Carlo simulations to assess the consequences of under- and overspecified moderated multiple regression models on biases in parameter estimates, type-I error rates, and statistical power relative to correctly specified models. Sample size, correlations among the predictors, and the effect size of the true interaction effects were varied across the simulations at levels commonly found within social science literature. Results are discussed in the context of both exploratory and confirmatory research studies, emphasizing practical implications for decision-making in the absence of refined theories or empirical studies.

14. Exploring Estimates of Multilevel Reliability for School Based Behavioral Measures

Katie Lane, D. Betsy McCoach

Many schools utilize universal behavior screening to quickly evaluate all students to determine who may need additional support. Most commonly, a single teacher will rate all students in their homeroom class or a specified hour of the day for secondary schools. As a result, screening data has a clearly nested structure and previous estimates of intra-class correlations show non-negligible cluster effects. In addition to the multilevel structure, behavior screening data is typically ordinal, most often with four scale points, and follows a non-normal distribution. In this study, we simulated data with characteristics similar to behavior screening data with levels of measurement error and between-cluster variability. We conducted single-level and multi-level reliability analyses to understand the impacts of ignoring nesting in reliability estimates and more robust identify method(s) for estimating reliability for future studies of behavior screening tools and other similar instruments.

15. Improving the Accuracy in Distinguishing between Second-Order and Bifactor Models Using Penalized Structural Equation Modeling

Hyeryung Lee, Xi Xu

Bifactor models frequently exhibit superior model fit compared to second-order models, even when data are generated from a second-order model. Previous research has indicated that fit indices and

chi-square difference tests tend to be biased in favor of bifactor models. However, when adjustments for cross-loadings and correlated residuals are incorporated post-hoc, these biases in the chi-square test diminish. Nevertheless, such post-hoc model adjustments based on modification indices may lead to issues of unreliability and lack of replicability. To address these concerns, this study introduces Penalized Structural Equation Modeling (PSEM) as a robust alternative. PSEM enhances Confirmatory Factor Analysis (CFA) by integrating all possible cross-loadings and correlated residuals with a zero-mean alignment prior. Through a simulation study, where data generated from a second-order model including cross-loadings and correlated residuals, the effectiveness of PSEM is evaluated against basic CFA. The chi-square model difference test, AIC, and BIC serve as the primary criteria for assessing model fit and reveal a significant distinction: while the basic CFA approach tends to favor the bifactor model, PSEM accurately identifies the underlying second-order structure. This finding underscores the precision of PSEM in distinguishing between second-order and bifactor models, highlighting its potential in advancing measurement modeling practices in psychometrics.

16. Bayesian Analysis of Ordinal Response Variables in Educational Research: A Comparison of Different Priors

Xing Liu

Although there is increasing popularity of applying Bayesian methods in educational research, Bayesian methods for ordinal response variables have been underutilized. In addition, the impact of prior distributions on model estimates has not been fully investigated. The purpose of this study was to apply Bayesian regression methods for ordinal response data in educational research. It also compared the Bayesian proportional odds (PO) models with different priors including noninformative, weakly informative, and informative priors. The empirical data from the High School Longitudinal Study of 2009 (HSLS:09) were employed to conduct the data analysis. Two Markov chains were used with the default number of 12,500 iterations and 2,500 burn-in draws for each chain, so the posterior sample size was 10,000 for each MCMC chain. The trace plots, histograms, autocorrelation plots, and density plots were used to evaluate MCMC convergence. The results of the conventional PO model and the Bayesian PO model with noninformative priors were compared. Bayesian PO models with different priors for the predictor variables were also discussed. In addition, suggestions on the use of priors were included. This study clarified misconceptions on the choice of priors and would help researchers become familiar with the Bayesian PO models.

17. Unpacking Subgroup Differences in Treatment Effects: A Causal Decomposition Approach for Mediated Moderation Analysis

Xiao Liu

Assessing differences between demographic subgroups (e.g., female and male) in the causal effect of a treatment (e.g., an intervention, a predictor of interest)—or, moderation analysis—plays important roles in population sciences. In moderation analysis, besides quantifying how much subgroups differ in treatment effect ("total moderation"), it is often useful to examine why the effect differences between the subgroups arise—such as by examining intermediate variables that may contribute to the effect difference—or, mediated moderation analysis.

Conventionally, estimands in mediated moderation analysis are often defined using parametric modeling approaches (e.g., Baron & Kenny, 1986), but such approaches have limitations in causal interpretation. For causal interpretation involving intermediate variables, causal mediation methods are fast-growing, but have limited development for mediated moderation analysis; a particular challenge is that the subgroups are often defined by demographic characteristics, which are nonmanipulable.

This study extends the causal decomposition approach in causal mediation literature, and develop methods for conducting mediated moderation analyses with causal interpretations. With our methods, the total moderation is decomposed into the causal estimands that capture how much the subgroup difference in treatment effect is attributable to the subgroup difference in an intermediate variable ("mediated moderation") and how much is not ("remaining moderation"). For the defined causal estimands, we also

develop multiply-robust estimators, which facilitate using machine learning techniques for causal inference in mediated moderation analyses. We evaluate performance of the methods through simulations, and illustrate the applications in an empirical mediated moderation analysis for unpacking gender differences in the effects of an intervention for youth behavioral problems.

18. Multilevel Mediation with Unmeasured Cluster-Level Confounders: Evaluating Propensity Score Models

Cameron McCann, Xiao Liu

Observational studies play an important role in educational research, as randomized assignment is often infeasible. Data in this context are often clustered within a multilevel hierarchical structure. While propensity score (PS) methods have been evaluated for studying treatment-outcome relations with multilevel data in observational studies, the performance of these methods for assessing mediation effects has not been thoroughly investigated before. Prior research evaluating PS methods for studying treatment-outcome relations in multilevel data suggests that the PS methods should take the multilevel structure of the data into account. Drawing on this research, the present study aims to assess the performance of three different PS models for estimating mediation effects in multilevel mediation analysis when treatment assignment is nonrandomized and an unmeasured cluster-level confounder exists. Through two simulations, we compare the following three PS models: single-level, fixed-effect, and random-effect. In the first simulation the unmeasured cluster-level confounder confounds all relations (i.e., treatment-outcome, treatment-mediator, and mediator-outcome) and in the second simulation it only confounds the treatment-outcome relation. Preliminary results highlight the importance of accounting for the multilevel structure of the data in the PS model.

19. Addressing the Effects of Home Resource Variables on Achievement in International Large Scale Assessment: Are Latent Variables the Right Approach?

Lionel Meng, Daniel Bolt

When controlling for contextual variables on achievement outcomes for purposes of comparing countries in international large-scale international assessments, investigators often define "contextual" latent variables. Such latent variables (or their corresponding latent variable estimates) become the controlled source(s) of influence on achievement, and portray the observed indicator variables as simple reflections of the contextual latent variable(s). Following recent work by VanderWeele (2022), we show how this approach is insensitive to the unique forms of causal influence associated with the indicators. For illustration, we consider the use of a home resources for learning index (HRL) to improve country comparisons using the TIMSS and PIRLS assessments. Using structural equation modeling techniques, we show how the correlations of the home resource variables with achievement are inconsistent with their loadings on the HRL latent variable, implying indicator uniqueness effects on achievement. One significant implication is to question the role of measurement invariance studies, and in particular the application of partial invariance models, for improving cross-country comparisons. Using simulation and empirical data, we revisit findings from a paper by Wendt et al. (2017) to illustrate how introducing country-specific loadings/discrimination parameters of the HRL indicators appear to harm, as opposed to improve, cross-country comparisons of achievement.

20. Assessing Differential Item Functioning of the PISA 2018 Academic Resilience Scale

Valerie Ofori Aboah, Latif Kadir, Ann O'Connell

Differential Item Functioning (DIF) is a common measurement issue that plagues measurement scales, rendering results from such scales biased as it yields an unfair advantage to a particular group. This study conducted DIF analysis on the USA PISA 2018 Academic Resilience Scale, with a sample size of 4548 (49.9% female, 50.1% male). We analyzed the 5-item Academic Resilience Scale, focusing on both Differential Item Functioning (DIF) and Differential Test Functioning (DTF). We applied Item Response Theory with a Graded Response Model to both reference and focal groups, using females as the reference. DIF analysis followed Meade's (2010) two-stage approach, employing Likelihood Ratio Tests

to compare baseline and comparison models. Effect sizes were quantified using various indices recommended by Meade (2010), including Signed Item Differences (SIDS), Unsigned Item Differences (UIDS), Expected Score Standardized Differences (ESSD) at the item level, and Signed Test Difference in the sample (STDS) and Expected Test Score Standardized Difference (ETSSD) at the scale level. While DIF was noted in over half of the items with medium effect sizes, scale-level effects for both groups were minimal, suggesting no significant DTF. Users can confidently use the resilience scale for gender comparisons without worrying about results being influenced by DIF.

21. Exploring Fit Indices Using Many-Facet Rasch Analysis Model for Validating Economics Test Items

Daniel O. Oyeniran, Enoch O. Olayori, Mopelola F. Oyeniran

The Rasch model is a part of the Item Response modeling family which improves on the shortcomings of the Classical test model. It focuses on the difficulty level of the items in relation to the ability of test takers. Many extensions have been proposed to the Rasch family among which is the Many-facet Rasch model which allows analysis of data beyond only item responses but other attribute variables. In this study, 40 Economics achievement items were administered to 498 Senior Secondary School II (SSS 2) students enrolled in an economics class in Ibadan metropolis of Oyo states, Nigeria. The facets considered in this study are students, items, school types (private: 61; public: 437), classes (science: 269; commercial: 127; art: 102) and sex (male: 221; female: 277). On average, student facet (-0.53) has a lower location on the logit scale and higher standard error (M = 0.36) when compared with the other facets which stand at 0-logits. All the facets on average are within the acceptable mean-square infit and outfit estimates of one while the standardized infits and outfits are close to zero, but students' facets are closer than other facets. There is a need for better targeting of achievement test items to students.

22. Comparing Model Complexity of Item Response Theory Models with Randomly Generated Data

Kirsten Reyna, Eric Loken

This study builds on the model complexity work of Bonifay and Cai (2017). We compare the fitting propensity of several categorical data models to randomly generated response patterns, extending the prior work in three ways. First, we add a 3-class latent class model, which yields a substantially better fit than all IRT models. Second, we fit multiple bifactor models to the data and show that the set of bifactor models will perform (nearly) as well as the factor model. Finally, we explore the data that are randomly generated from the simplex for the 128 data patterns, noting that the first and second eigenvalues are always small relative to what might be expected for clinical or social science measurement instruments.

23. LMMCov: An Interactive Research Tool for Efficiently Selecting Covariance Structures in Linear Mixed Models Using Insights from Time Series Analysis

Perseverence Savieri, Kurt Barbé, Lara Stas

Linear mixed models (LMMs) are commonly used for analysing multilevel, repeated measures, or longitudinal data where observations are correlated. Modeling the covariance structure of random effects is critical since any misspecification can negatively affect the estimates of the standard errors for the fixed effects, leading to incorrect statistical inferences and tests. Despite the use of information criteria like AIC, BIC, CAIC, HQIC and AICC for model selection, misspecification of covariance structures remains a challenge. We introduce LMMCov, an interactive Shiny app that aims to assist researchers in targeting the correct covariance structure by incorporating interactive visualisations of residuals. Time-series concepts, such as autoregressive models, are incorporated to select more complex structures. Users can: 1) get instructions and insights from the example database; 2) analyse their own data; and 3) create dynamic, systematic reports that can be downloaded. This poster outlines the methodology, theoretical background, and user interface of LMMCov. We highlight its potential to enhance covariance structure selection in LMMs.

24. Contributions of Evolution Knowledge, Conflict, and Religiosity on Longitudinal Patterns of Evolution Acceptance

Gena Sbeglia, Ross Nehm

Evolutionary biology is one of the least diverse STEM fields. Although the reasons for this extreme lack of diversity are poorly understood, one hypothesized contributing factor is low evolution acceptance, which is a nationwide challenge. A major limiter to the development of interventions to improve evolution acceptance is a lack of consensus regarding how evolution knowledge, religiosity, and perceptions of conflict impact it. In particular, researchers have found conflicting results regarding whether or not evolution knowledge and acceptance are related. These conflicting results are likely due to several methodological problems including low sample sizes, lack of longitudinal designs, and poor quality measurement instruments. Our study seeks to address these limitations. Using a sample of >5400 gateway biology students and well-validated instruments, we ask: Does evolution knowledge, religiosity, and conflict, account for evolution acceptance patterns? Using HLM, we report that religiosity explained a remarkably small (<4%) amount of the overall (within + between person) variance in acceptance patterns while evolution knowledge (5.5%-17.5%) and perceptions of conflict (24%-35%) had larger explained variance, especially for microevolution and macroevolution acceptance. The results support the need for causal work examining whether decreasing perceptions of conflict and increasing evolution learning impact evolution acceptance.

25. An Illustration of Advanced Intraclass Correlations for Inter-Rater Reliability

Aaron Matthew Simmons, Jeffrey Girard

Inter-rater reliability can be defined as the extent to which raters assigned similar scores to objects of measurement (e.g., diagnoses to patients, types to documents, points to athletes, or stars to movies). Intraclass correlation coefficients (ICCs) are derived from fractions of variance components and have long been used to quantify inter-rater reliability for continuous and near-continuous scores. Recent advances have extended ICCs to accommodate missing data, unbalanced designs, and multilevel structures (ten Hove, Jorgensen, & van der Ark, 2021, 2022). We illustrate the use of these advanced ICC formulations to answer important questions about the reliability of ratings in real-world settings. We also showcase novel software (the varde R package) that we created to allow users to easily implement these tools and techniques. Finally, we propose and solicit feedback on our plans for new extensions of the ICC framework including the use of generalized linear models to accommodate ratings that are binary, categorical, or otherwise bounded.

26. Measurement Invariance of Attachment and Perceived Stress across Six Intersectional Race/Ethnicity and Gender Identities in Emerging Adults

Jodi Sutherland Charvis, Liam Rozum, Sydney Iacoi, Chrystal Vergara Lopez, Hector Lopez Vergara

Early life exposures such as attachment styles hard wire one's appraisal of stressors, determining an outcome of psychological distress and chosen coping strategies. Knowledge of the interplay of these mechanisms for groups of emerging adults is therefore of great importance. Of note, these group differences are often reported according to race/ethnicity and gender. However, most of these group comparisons assume equal psychometric functioning of measures, negating tests for measurement invariance prior to making inferences about minoritized populations.

Multigroup confirmatory factor analysis (CFA) across both race/ethnicity and gender was applied to a measure of perceived stress and Adult Attachment for an online sample of n=1,187 (33% Black, 35% Latino/a, 32% White; 50% Women; aged 18-26). Bias in measurement was found for both measures, corrected using partial metric and scalar invariance for perceived stress and attachment respectively. For mean differences the two highest and least scores according to intersectional groups were as follows: distress: White Women, Latina Women, White Men; coping: Latino Men, Latina Women, White Men; avoidance: White Men, Latino Men, Latina Women; anxiety: White Men, Black Men, White Women.

Moving the literature towards a better understanding of intersectional group differences may necessitate models that can test for bias in measurement.

27. Planting Decision Trees: Human-Friendly Interpretation of Monte Carlo Simulations, Multiverse Analyses and Multivariate Posterior Distributions

Michael S. Truong, Gabriel Crone, Udi Alter, Ji Yeh Choi

Monte Carlo simulations, multiverse analyses and multivariate prior distributions are all core elements of modern data science. However, interpreting each of these elements is frequently a cumbersome ordeal that leaves much to be desired. Monte Carlo simulations are frequently presented as a large multi-way table that can be both intimidating and also difficult to extract key findings from. Similarly, multiverse analyses may have so many branching decisions that it becomes difficult to parse which decision may have been most critical to the final results. Multivariate posterior distributions feature heavily in Bayesian data analysis, but selecting the dimensions to visualize the largest changes in probability is challenging due to their high-dimensional nature. In response to these challenges, we investigate and illustrate the use of the classic machine learning technique of decision trees to extract the key aspects of each of these elements. In a typical machine learning context, decision trees are a powerful way of discovering the largest main effects and higher order interactions in an interpretable fashion. The adoption of effective applications of decision trees to each of these elements promises the human-friendly interpretation of Monte Carlo simulations, multiverse analyses and multivariate prior distributions.

28. Adjusting for Measurement Error in Learning Sciences Data via Bayesian Multilevel Modeling

M. Shane Tutwiler, Alana D. Newell

Precision adjustment, correction for measurement error in the modeling of outcomes, is a commonly used approach in the analysis of state standardized test scores (e.g., Reardon et al., 2021). However, the impact of measurement error is even more pronounced in the types of small, noisy data often generated in the types of discovery research conducted in the learning sciences. In this poster we first demonstrate the utility of using Bayesian Multilevel Modeling to adjust for measurement error in simulated data before using it on a set of data tracking the learning of students who used a science education curriculum. Limitations and future research are also discussed.

29. Response Time Modeling for PISA 2018 Math Items Using the Joint Model

Claudia Ventura, Eric Loken

Response time (RT) data offers significant insights into students' cognitive processes and test-taking strategies. By analyzing 31 math items from the 2018 PISA assessment, we investigate the joint modeling of accuracy and RT data. We first descriptively explore the association between item level RT and test score, showing that the correlation changes direction during the test. We also explore the dimensionality of the log RT data and consider the implications for joint models that include only one dimension for speed. We investigate the fit of the log normal IRT model (LNIRT) and relate diagnostics to our earlier explorations.

30. Applying a Stepwise MIMIC Modeling Approach to Assess Measurement Invariance in a Latent Class Analysis of Doctoral Socialization

Boshi Wang, Katherine Masyn

Although the manual maximum likelihood (ML) three-step approach (Asparouhov & Muthén, 2013) is currently one of the best developments to examine the relationships between covariates and latent class membership, recent studies indicate that the approach is not robust to specification errors because of the omission of direct effects (i.e., measurement noninvariance or differential item functioning) from covariates to the latent class indicators. To address this methodological issue, Masyn (2017) developed a 7-step, logical and principled approach of stepwise MIMIC modeling to assess the measurement invariance in latent class models. In the current study, we applied this approach in a 4-class, 8-indicator latent class model characterizing the socialization patterns of first-year Biology doctoral

students in the U.S. (N = 325) with faculty and peers. We evaluate self-identified gender, international status, first generation student status, and the intersections of these social identities as potential sources of differential item functioning (DIF) in the model. This poster summarizes the application and results of the DIF analysis using the stepwise MIMIC modeling approach, highlighting the substantive findings related measurement noninvariance across intersecting social identity groups and the consequences for understanding the differences in latent class membership for those groups. Future research directions are discussed.

31. Examination of Methods of Bayesian Hypothesis Testing for Direct Replication Studies in a Meta-Analytic Framework

Naike Wang

While the field of psychology is on the verge of a paradigmatic shift toward increased transparency and more replication attempts, a new methodological challenge arises: How do we determine if a multi-lab replication study is successful? The predominant frequentist method (comparing p-values of original studies with their replications) has been increasingly seen as inadequate. To fill this gap, this study integrates Bayesian methods into the researcher's toolkits by evaluating four Bayesian hypothesis tests (meta-analytic Bayes factor, or MABF) concerning their ability to identify false original findings based on their direct replications. To achieve this, the study adopts a two-phase simulation strategy. The first phase entails generating original findings based on underlying population effect sizes within a simulated research environment shaped by varying degrees of p-hacking, publication bias, and statistical power. The second phase then replicates these findings, applying MABFs for synthesis and analysis. By drawing connections between distinct MABFs that have never been directly compared, this study comprehensively reexamines the potential of Bayesian hypothesis testing as a replication success metric. The study reveals the most robust method(s) for evaluating replication success while also highlighting situations where all methods perform poorly, offering vital methodological guidance for future large-scale, multi-lab replication initiatives.

32. Underpowered Studies and Overrepresented Significant Findings in Educational Psychology: A Comprehensive Examination of Empirical Evidence

Naike Wang

Systemically underpowered research persists in behavioral science. A critical concern arises when meta-analyses synthesize findings in a literature dominated by underpowered studies, leading to strongly biased summary effects. No study has, however, examined the credibility of meta-analyses in educational psychology. We conduct a comprehensive assessment of meta-analytic outcomes by calculating the median retrospective power (MRP) based upon selected meta-analyses published in the top five educational psychology journals between 2012 and 2022. The MRP is a novel power analysis method, serving as an indicator of the credibility of meta-analyses and the replicability of typical studies within the field. Preliminary results reveal a substantial difference in power between meta-analyses involving experimental studies and those focusing on correlational studies: 31% vs 95%, highlighting that the power of typical experimental studies in educational psychology falls well below the commonly accepted threshold. Consequently, these studies are at a significantly increased risk of replication failures. We conduct further investigation to explore the methodological factors contributing to this pronounced power discrepancy. In light of these findings, we strongly advise researchers to interpret positive meta-analytic findings with great caution. This study underscores the pressing need to strengthen the credibility and replicability of experimental research in educational psychology.

33. When to Standardize Your Data? A Comparison of Percentile Bootstrap Confidence Interval for Standardized Indirect Effects

Yiwei Wang, Amanda Kay Montoya

Reporting standardized effect sizes have become increasingly important in psychological research including accompanying confidence intervals. In mediation analysis, this refers to constructing a

confidence interval for standardized indirect effects. The present study compares two methods for standardization using the percentile bootstrapping method. The intuitive approach, naive standardization, standardizes the predictors, mediators, and outcomes before bootstrapping. An alternative, repeated standardization, standardizes the above variables in each bootstrapped sample. Previous research suggests that naive standardization tends to be more conservative than repeated standardization (Cheung, 2009). Using a Monte Carlo simulation, we compared naive and repeated standardization across various levels of confidence intervals (90%, 95%, 99%), sample sizes (n=20, 50, 100, 200, 500, 1000), sizes of indirect effects, and distributions (normal, skewed, kurtosis). Our outcome variables are width of the confidence intervals, Type I and II error rates, and empirical coverage probability. We find that naive standardization has an empirical coverage probability closer to the nominal value in small sample sizes. We replicate the results of Cheung (2009) and further identify the thresholds when one method outperforms the other in practical use.

34. Modeling the Dynamics of Pain and Emotion Interactions in Fibromyalgia Using the Valved Bayesian Reservoir Model

Mirinda Whitaker, Akiko Okifuji, Jeanine Stefanucci, Pascal Deboeck

Pain experiences in idiopathic chronic pain conditions like fibromyalgia likely arise from dynamic interactions between aberrant pain signaling, deficiencies in the endogenous pain modulation (opioid) system, and emotional dysregulation (Clauw, 2014). While there are many options for modeling dynamics, modeling approaches are often disconnected from the processes they are applied to. Here we took a substance first approach and brought together affective scientists, pain researchers, and quantitative methodologists to craft a tailored model to decompose pain experiences in fibromyalgia patients based on ecological momentary assessment data on fibromyalgia symptoms. The current model extends previous work on the Reservoir Model (Deboeck & Bergeman, 2013; Whitaker, Bergeman, & Deboeck, under review) to include how dynamics may change over time as the result of time varying moderating parameters.

35. A Methodological Review of Intersectionality in Differential Item Functioning, Issues, and Challenges

Winifred G. Wilberforce, Ann A. O'Connell

The intersectional approach to differential item functioning (I-DIF) provides a new lens for psychometricians and quantitative researchers to contextualize the multiple identities of test-takers that may exist in their sample of interest. Compared to traditional DIF methods which focus on isolated group differences, i.e. due to race or gender, the I-DIF approach allows for a broader and more accurate understanding of potential item bias, as it captures a more realistic intersection of test-taker identity and lived experience. As researchers explore ways to successfully model this idea rooted in Quantitative Critical Race Theory (QuantCrit), there are questions about the effectiveness of I-DIF detection methods and the limitations of existing software packages. For example, the categorization of these intersectional identities raises questions about sample size as it can result in some groups having very small sample sizes that may affect statistical power. Because this idea is new, there is limited literature available on these issues. The purpose of this presentation is to review the available literature on the methods and software that have been used in modeling I-DIF, highlight the methodological challenges that exist with current methods, and make recommendations to assist quantitative researchers in promoting test fairness and limiting item bias.

36. Assessing Psychometric Properties of the Japanese Version of the Perceptions of Inclusion Questionnaire (PIQ) in Student and Teacher Samples

Akie Yada, Susanne Schwab, Kanako Korenaga, Carmen L. A. Zurbriggen

The Japanese version of the Perception of Inclusion Questionnaire (PIQ), aimed to assess students' subjective well-being at school, was analyzed using classical test theory (CTT) and item response theory (IRT) frameworks. The scale was originally developed in German and translated into

Japanese, where significant linguistic and cultural differences exist between the two languages. It is particularly important to establish the reliability and validity of the scale for its effective utilization in future research and practices. Data were collected from 713 students (grades 5-9) and their teachers (N = 35). First, the three-dimensional structure of both the student and teacher versions was confirmed through exploratory and confirmatory factor analyses. Second, strong measurement invariance between primary and lower-secondary school student samples was found. Third, a graded response model was conducted to provide detailed information about how each of the items functions at every level of the construct. Finally, results from multi-trait multi-method analysis indicated the PIQ's convergent and discriminant validity of the traits and discriminant validity of the methods. The results of our study highlight the benefits of testing psychometric properties with both CTT and IRT frameworks and demonstrate the reliability and validity of the scale in the Japanese school context.

37. Item Sensitivity Matters for Subtle Changes of Emerging Skills

Eleanor Fang Yan, Ryan Bowles, Gary E. Bingham, Xiao Zhang, and Hope Gerde

Multi-dimensional literacy skills, such as reading and writing, are often found to bear group differences due to certain aspects of students' demographic backgrounds, such as socioeconomic status or parents' educational background. These differences could be further complicated by various basal and ceiling rules during administration. We investigate the psychometric properties of the Test of Early Writing Language, 3rd edition (TEWL-3), collected as a part of a larger project. The test items that are not straightly ordered by their difficulty not only add testing fatigue to both students and administrators due to prolonged response time, but they potentially weaken the sensitivity of items that assess the subtle changes of emerging writing skills of children from 4-0 (4 years, 0 months) through 6-11. This study uses a multi-dimensional IRT framework to position the test in its theoretical models to assess the item sensitivity. We also follow up with differential item functioning to examine the group differences in our sample. Difficulties in structure, design alternatives, estimation, and educators' feedback are discussed.

38. Assessing Model Fit Indices in Ordinal Factor Analysis Models: ML vs. ULS

Guyin Zhang, Dexin Shi, Amanda Fairchild

In structural equation modeling, a variety of comparative model fit indices such as the root mean square error of approximation (RMSEA), comparative fit index (CFI), Tucker-Lewis index (TLI), goodness-of-fit index (GFI), adjusted goodness-of-fit index (AGFI), and parsimonious goodness-of-fit index (PGFI) are often used to evaluate model fit. While the cutoff values for fit indices obtained through normal theory-based maximum likelihood (ML) are well-established, there remains a lack of comprehensive understanding regarding the behavior of the indices under unweighted least squares (ULS) estimation. The objective of this study was to assess the correlation between model fit indices and various incidental parameters across different forms of model misspecification. Results showed that in comparison to ML, ULS yielded higher values across all fit indices except for RMSEA. The performance of fit indices exhibited variability in the presence of model misspecification under different estimators, and all ULS-based fit indices were comparatively unaffected by model size under misspecified dimensionality, irrespective of factor loadings and misspecification levels. RMSEA demonstrated greater sensitivity to factor loadings than other indices. ULS-based fit indices (except for PGFI) displayed lower sensitivity to model misspecification compared to ML.

39. Comparing the Psychometric Performance of Likert Scales and Sliders

Guyin Zhang, Dexin Shi, Amanda Fairchild

The Likert scale remains one of the most widely used response options for self-report data, despite being controversial. Concerns have been expressed about the nature of the data generated from Likert scales and the unequal distance between adjacent measurement points. Continuous rating scales, such as sliders, represent an alternative response option that has begun to gain attention recently. However, existing research on sliders has predominantly concentrated on completion rates and response times rather than exploring their measurement performance. This study conducts a comparison between

Likert scales and four different slider formats, using structural equation modeling and individual scores to anchor evaluation perspectives.
The slider demonstrates higher reliability compared to the Likert scale, with consistent reliability values observed across various slider formats. There is no substantial evidence indicating a significant difference in model fit indices between sliders and Likert scales in either confirmatory factor or measurement invariance models. Likert scales and sliders exhibit comparable performance in correlation coefficients. However, disparities arise in factor scores, means, and latent mean differences when comparing male and female groups. Moreover, a mismatch is evident in subjects' scores for the same item across different response formats.

40. A Feature Selection Approach to Improve Subgrouping Accuracy in Multivariate Dynamic Processes

Di Jody Zhou, Sebastian Castro-Alvarez, Siwei Liu

Dynamic psychological processes are typically heterogeneous across individuals. Gaussian mixture models (GMM) have become increasingly popular for detecting such heterogeneity. However, applying GMM in the context of dynamic modeling can be challenging because dynamic models are often over-parameterized. As a result, the performance of GMM is likely to deteriorate notably due to many unnecessary parameters that complicate the subgrouping algorithm. Reducing the number of non-informative features has been shown to improve subgrouping accuracy. In this study, we propose integrating regularization, a feature selection technique, into GMM to discard non-informative features in vector autoregressive (VAR) based dynamic modeling. We first introduce the rationale behind this penalized dynamic Gaussian mixture method. We then present simulation results to evaluate the performance of this method compared to a traditional mixture subgrouping method without feature selection under data conditions commonly encountered in studies of dynamic psychological systems.

2026 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 11th M^3 conference will be held in late June, 2026 at the University of Connecticut.

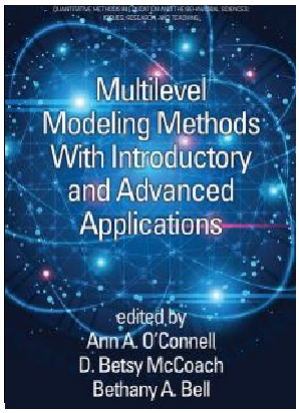
Submissions for the 2026 conference are due 1/30/26. We welcome both methodological research papers and papers that illustrate novel modeling applications of modeling. Papers related to latent variable modeling, multilevel modeling, mixture modeling, longitudinal modeling, dynamic modeling, and psychometric 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 2026 conference are due January 30, 2026. Notifications of presentation status will be emailed by February 20th, 2026. For more information about the 2026 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.

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]

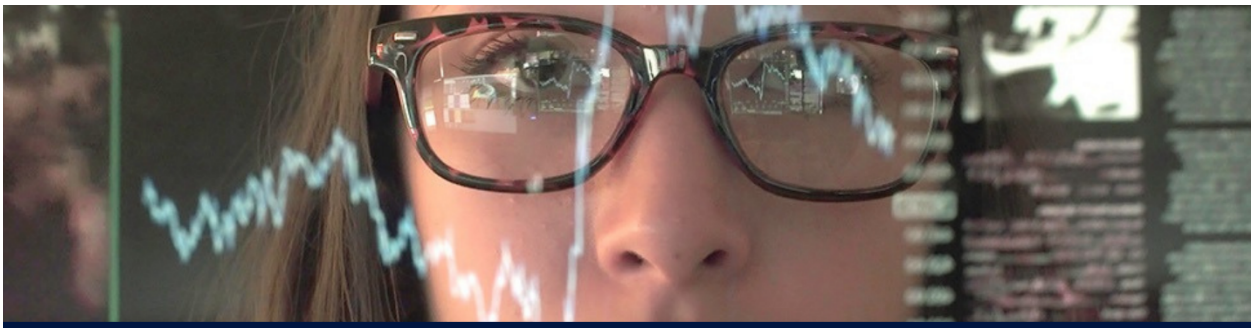
There is a new community space [OpenSciEdu-L] reserved for sharing information and engaging in discussions about Open Science practices with a focus on promoting Open Science, especially in the field of Educational Research. This listserv was created by members of the *Federation* of education hubs. 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*