UCONN

NEAG SCHOOL OF EDUCATION

MODERN MODELING METHODS

2016

May 24-25, 2016

2016 Modern Modeling Methods Conference

Sponsored by the **Measurement, Evaluation, and Assessment Program**, the **Educational Psychology Department** and the **Neag School of Education**, and the **University of Connecticut**.

Welcome to the 6th annual Modern Modeling Methods conference at the University of Connecticut. Special thanks go to Cheryl Lowe, Robbin Haboian-Demircan, Steve Rugens, Ravi Thumma Satyanarayana, and conference services for providing administrative and logistical support for the conference. Thanks also to Dr. Del Siegle, chair of the Educational Psychology department, Dr. Richard Schwab, Dean of the Neag School of Education, and Statistical Innovations for their continued support of this conference. In addition, thank you to all of the keynote speakers and concurrent presenters for making this wonderful program possible.

Finally, thank you to all of the 2016 Modern Modeling Methods conference attendees for coming and being a part of the sixth annual M^3 conference! I hope to see you all back in Storrs May 23-24 2017 for the seventh annual Modern Modeling Methods conference. Confirmed keynote speakers for 2017 include Dr. Steven Boker (UVA) and Dr. Kenneth A. Bollen (UNC). Steven Boker will be offering a full day Pre-Conference workshop on Dynamic SEM on May 22; and Ken Bollen will be offering a half day Post-Conference workshop on Model implied instrumental variables using MIIVsem on May 25. We hope to add a third keynote speaker to the line-up in the coming months. If you have suggestions, please be sure to fill out the conference evaluation, located on our website, www.modeling.uconn.edu or email them to me at betsy@uconn.edu

Proposals for concurrent sessions will be due February 1st, 2017, and can be submitted online at our website: www.modeling.uconn.edu

We hope you have a great conference – Happy modeling!

Contents

Modern Modeling Methods - 2016 Schedule

Monday – May 23

Pre-Conference Workshop: ***New features in Mplus***

School of Business – Room 106

Lecturers: Bengt Muthén, Tihomir Asparouhov and Ellen Hamaker

The pre-conference workshop runs from 8:30 am to 6:00 pm, and has two parts:

1. Analyses you probably didn't know that you could do in Mplus - with highlights from the forthcoming book, Regression and Mediation Analysis Using Mplus (Muthen, B., Muthen, L., & Asparouhov, T.)
2. Preview of the forthcoming Mplus Version 8

Morning Session

Part (1) will sample examples from a range of Mplus topics such as heteroscedasticity modeling for regression and mediation analysis using the CONSTRAINT option and random coefficient modeling; Heckman modeling; Bayesian approach to missing data on binary covariates; Monte Carlo studies of moderated mediation; sensitivity analysis of mediator-outcome confounding; counterfactually-defined causal effect mediation modeling with binary outcomes and binary and nominal mediators; and two-part mediation analysis with floor and ceiling effects.

Afternoon Session

Part (2) provides a preview of the forthcoming Mplus Version 8 which features new methods for analyzing intensive longitudinal data such as that obtained from ecological momentary assessments. Long time series for both n=1 and n >1 are discussed. The focus is on time-series analyses with auto-regressive and moving-average components both for observed-variable models such as regression and mediation analysis and for latent-variable models such as factor analysis and structural equation modeling.

The pre-requisites for this workshop are an understanding of Linear and Logistic regression; a computer is not needed.

Tuesday – May 24

Breakfast and registration ***8:00 am – 8:45 am***

Opening Keynote by Bengt Muthén

Which methods do we need for intensive longitudinal data?

8:45 am – 10:15 am in Laurel Hall 102

Concurrent paper session 1 10:30 am – 12:00 pm

Session 1A: Propensity Score and Group Balance methods
Laurel Hall 201

Paper	Authors
In the Pursuit of Group Balance on a Known Covariate in a Small-Scale Experiment: A Monte Carlo Study of the Efficacy of Standard Approaches	Tyler Hicks Jeffrey Kromrey
Examining Propensity Score Estimation Models and Conditioning Methods using Multilevel Modeling	Aarti Bellara
The Use of Nonparametric Propensity Score Estimation with Data Obtained Using a Complex Sampling Design	Ji An Laura M. Stapleton

Session 1B: New Directions in Item Response Theory
Laurel Hall 202

Paper	Authors
Rethinking Complexity in Item Response Theory Models	Wes Bonifay
Modeling Directional Local Item Dependence	Kaiwen Man Hong Jiao Meng Qiu
Methodological Illustration: Hyperparameters and Hyperpriors in Item Response Theory, Interpretations and Use	Allison Ames

Session 1C: Multiple Imputation Methods
Laurel Hall 205

Paper	Authors
Multiple Imputation in Finite Mixture Modeling	Daniel Y. Lee
Evaluation of Multiple-Imputation Procedures for Handling Missing Data in Multivariate Time Series Models	Linying Ji, Sy-Miin Chow, Nicholas C. Jacobson
Estimating Latent Variable Interactions with Incomplete Exogenous Indicators via CART in MICE using Unconstrained Product Indicator Approach	Evgeniya Reshetnyak Heining Cham

Session 1D: Innovations in Structural Equation Modeling
Laurel Hall 206

Paper	Authors
Non-Linear Growth Models as Measurement Models: A Second-Order Growth Curve Model for Measuring Potential	Dan McNeish Denis Dumas
Heteroscedastic-Consistent Confidence Intervals for Standardized Regression Coefficients	Paul Dudgeon

Session 1E: Network and Spatial Analysis
Laurel Hall 301

Paper	Authors
Modeling the Effects of Network Attributes on Subgroup Integration	Qiwen Zheng Tracy Sweet
Individuals as Dynamic Networks: A Merging of Intraindividual Variability, Network Analysis, and Experience Sampling	Xiao Yang David Lydon Nilam Ram

Session 1F: Longitudinal Mediation Models
Laurel Hall 302

Paper	Authors
Multilevel autoregressive mediation models: Specification, estimation, and applications	Qian Zhang Lijuan Wang Cindy Bergeman
Random-Effects Cross Lagged Panel Models for Longitudinal Mediation Analysis	Wei Wu

Session 1G: Symposium- Identifying Informative Subgroup Typologies in K-12 Education Leadership Data through the Application of Latent Class Analysis (LCA) and Multilevel LCA
Laurel Hall 305

Paper	Authors
Paper 1: *The influence of typologies of school leaders on teacher retention: A multilevel latent class analysis*	Angela Urick
Paper 2: *Principal Turnover: Are there Different Types of Principals Who Move From or Leave Their Schools? A Latent Class Analysis of the 2007-08 Schools and Staffing Survey and the 2008-09 Principal Follow-up Survey*	Jared Boyce
Paper 3: *How Leaders Agree with Teachers in Schools on Measures of Leadership Practice: A Two-Level Latent Class Analysis of the Comprehensive Assessment of Leadership for Learning*	Alex J. Bowers

Concurrent paper session 2 1:15 pm - 2:45 pm

Session 2A: Missing Data: Issues and Opportunities
Laurel Hall 201

Paper	Authors
Don't Be Fancy. Impute Your Dependent Variables!	Kyle M. Lang Todd D. Little
Missing Data in Dyadic Modeling: Issues and Opportunities	Alexander M. Schoemann John K. Sakaluk
Confirmatory Factor Analysis with Planned Missing Data: A Monte Carlo Simulation Study	Mine Dogucu Hui Jiang Pamela Soto-Ramirez

Session 2B: Repeated Measures Approaches and Issues
Laurel Hall 202

Paper	Authors
Misleading Results From Combining Predictions of Residualized and Simple Gain Scores in the Same Longitudinal Analyses	Robert E. Larzelere Isaac J. Washburn Mwarumba Mwavita Ronald B. Cox Jr. Taren Swindel
Difference-in-Difference Estimates of Demographic Processes	Lawrence L. Wu

Session 2C: Propensity Score Methods
Laurel Hall 205

Paper	Authors
Optimizing Random Forests Propensity Scores	Heining Cham, Landon Hurley Yue Teng
A simple fix for bias due to a latent/mismeasured covariate in propensity score weighting analysis: Factor scores from models inclusive treatment and other covariates	Trang Quynh Nguyen, Cyrus Ebnesajjad, Hwanhee Hong, Elizabeth A Stuart
Investigating Causal DIF via Propensity Scores Methods	Yan Liu, Chanmin Kim, Amery Wu, Paul Gustafson, Bruno Zumbo

Session 2D: State of the Art in Dynamic SEM
Laurel Hall 206

Paper	Authors
Latent class analysis for intensive longitudinal data, Hidden Markov processes, Regime switching models and Dynamic Structural Equations in Mplus 8	Tihomir Asparouhov

Session 2E: Mixture Modeling Approaches
Laurel Hall 301

Paper	Authors
Testing the Effectiveness of Three-Step Approaches for Auxiliary Variables in Latent Class Analysis	Zachary K. Collier Walter L. Leite
Aberrant Behavior Detection based on a Multivariate Mixture Modeling Approach	Kaiwen Man Jeffery Harring
Nonparametric Multilevel Latent Class Analysis: A Comparative Study of Teachers' Beliefs about Teaching and Learning in 24 Countries	Chi Chang

Session 2F: Measurement and Modeling
Laurel Hall 302

Paper	Authors
Assessing Item Measurement Invariance in Large Scale Cross-Country Surveys. Monte Carlo Simulation Study	Artur Pokropek
Modeling individual differences in terms of process model parameters in the context of large-scale reaction time data	Zita Oravecz
Assessing the convergent and discriminant validity of person x situation interaction effects: Development of a multi-method model for measuring situation-specific traits.	Fred Hintz, Christian Geiser Saul Shiffman

Session 2G: Applications of SEM and multilevel modeling
Laurel Hall 305

Paper	Authors
Two illustrations of meta-analytic structural equation modeling	Suzanne Jak
Using multi-membership multi-level VAM model to evaluate teacher and school effects	Eunkyeng Baek Chunhua Cao
Sensation Seeking and Risk Taking among Family Members - APIM approach	Adam Sagan Anna Siwy-Hudowska

Concurrent paper session 3 3:00 pm - 4:30 pm

Session 3A: Symposium: Latent variables and causal inference
Laurel Hall 201

Paper	Authors
Paper 1: *Causal inference using counterfactually-defined effects in mediation analysis with latent variables*	Bengt Muthén
Paper 2: *Causal effects using counterfactuals for mediation models with a two-part mediator*	Mårten Schultzberg
Paper 3: *Latent class covariate balancing for causal inference*	Katherine Masyn & Marcus Waldman
Paper 4: *Principal stratification when the intermediate variable is continuous and measured with error*	Marcus Waldman & Katherine Masyn

Session 3B: Symposium: Innovations in Planned Missing Data Designs
Laurel Hall 202

Paper	Authors
Paper 1: *How low can you go? Minimum sample size thresholds for the 6-form planned missing data design*	Jacob D. Curtis Samuel F. Meeks Kyle M. Lang Todd D. Little
Paper 2: *Take it to the next level! Planned missing data designs at Level-2 in multilevel models*	Britt K. Gorrall Rong Chang Kyle M. Lang Todd D. Little
Paper 3: *What is the ability to impute: Entire level 1 missingness based on level 2 group membership*	Luke Waggenspack Kyle M. Lang Todd D. Little

Session 3C: Longitudinal mixture modeling
Laurel Hall 205

Paper	Authors
Right-Sizing Growth Mixture Models for Longitudinal Data	Phillip Wood Wolfgang Wiedermann Douglas Steinley
Child Types and Patterns in the Classroom: A Latent Trajectory Approach to Longitudinal Peer Sociometric Nominations over Grades	Richard Faldowski

Session 3D: Measurement Models and IRT
Laurel Hall 206

Paper	Authors
Using Balance Scores to Disentangle Differential Item Functioning, Item Impact, and Item Bias	Amery D. Wu Yan Liu Bruno D. Zumbo
Accounting for Rater Heterogeneity in Systematic and Random Errors with an Item Response Rater Model	Yao Xiong
A Multilevel Bifactor Approach to Construct Validation of Mixed-Format Scales	Yan Wang Eun Sook Kim

Session 3E: Symposium: New methodologies in multivariate modeling
Laurel Hall 301

Paper	Authors
Moderator: Ofer Harel	
Presentation 1: *Network Analysis Using a Random Walk in Sparse Graphs*	Haim Bar
Presentation 2: *Robust multivariate mixture model via mean-shift penalization*	Kun Chen
Presentation 3: *Robust analysis of multiple secondary outcomes in case-control studies*	Elizabeth Schifano

Session 3F: Symposium: Advanced Quantitative Methods in Educational Research
Laurel Hall 302

Paper	Authors
Paper 1: *Applying meta-analytic structural equation modeling (MA-SEM) to jointly model components of the Simple View of Reading (SVR) and the Not-so-Simple View of Writing (NSVW)*	Yusra Ahmed Richard Wagner Young-Suk Grace Kim W. Pat Taylor
Paper 2: *Is the treatment weak or the test insensitive: How explanatory item response models can elucidate the nature of intervention effects?*	Paulina A. Kulesz David J. Francis Jack M. Fletcher Sharon Vaughn
Paper 3: *Regression discontinuity and statistical power: A more efficient design for research with selected samples*	Jeremy Miciak W. Pat Taylor Yusra Ahmed Jack M. Fletcher

Poster session and Reception

4:45 pm – 6:45 pm

Student Union Ballroom

Wednesday – May 25

Concurrent paper session 4 8:00 am - 9:00 am

Session 4A: Bayesian SEM
Laurel Hall 205

Paper	Authors
The Impact of Bayesian Priors on Specification Search of Structural Equation Modeling	Xinya Liang, Yanyun Yang
An Investigation of Prior Distribution Accuracy and Robustness in SEM	James P Clifton

Session 4B: Generalized Structured Component Analysis
Laurel Hall 301

Paper	Authors
Application of Generalized Structured Component Analysis to Item Response Theory	Ji Hoon Ryoo
gesca: An R Package for Generalized Structured Component Analysis	Sunmee Kim

Session 4C: Introduction / Efficient Computational Strategy for Relational SEM
Laurel Hall 305

Paper	Authors
Introduction to Relational SEM and an Efficient Computational Strategy for Relational SEM	Joshua Pritikin

Session 4D: That's a State Space Model too!
Laurel Hall 306

Paper	Authors
That's a State Space Model too!	Michael D. Hunter

Session 4E: Regression Trees for Longitudinal and Clustered Data
Laurel Hall 106

Paper	Authors
Regression Trees for Longitudinal and Clustered Data: Methods, Applications, and Extensions	Jeffrey Simonoff

Concurrent paper session 5 9:15 am – 10:15 am

Session 5A: Bayesian Multilevel Modeling
Laurel Hall 205

Paper	Authors
Bayesian Model Averaging for Multilevel Models - Paper	June Yang
The matching effect of intra-class correlation (ICC) on the estimation of contextual effect: A Bayesian approach to multilevel modeling	Hawjeng Chiou

Session 5B: Analysis of multivariate longitudinal data in Stan
Laurel Hall 301

Paper	Authors
Analysis of multivariate longitudinal data in Stan	Tyler Matta Quinn Lathrop

Session 5C: Dealing with missing data in propensity score analyses
Laurel Hall 305

Paper	Authors
Evaluating Multiple Imputation Techniques for Treating Missing Data in Propensity Score Analyses	Seang-hwane Joo, Jessica Montgomery, Yan Wang Patricia Rodriguez de Gil Eun Sook Kim, Jeffrey D. Kromrey
A comparison of methods for imputation of missing covariate data prior to propensity score analysis	Walter Leite, Burak Aydin Sungur Gurel, Duygu Cetin-Berber

Session 5D: What's for dynr: A package for linear and nonlinear dynamic modeling in R
Laurel Hall 110

Paper	Authors
What's for dynr: A package for linear and nonlinear dynamic modeling in R	Lu Ou, Michael Hunter Sy-Miin Chow

Session 5E: Speed Dating with Regression Methods
Laurel Hall 106

Paper	Authors
Speed Dating with Regression Methods	David J Corliss

Session 5F: Bayesian Multivariate Multilevel Models for Pretest Posttest Data
Laurel Hall 306

Paper	Authors
Bayesian Multivariate Multilevel Models for Pretest Posttest Data	Jean-Paul Fox

Concurrent paper session 6 10:30 am – 12:00 pm

Session 6A: New Directions in Mediation
Laurel Hall 205

Paper	Authors
A novel effect size measure for the indirect effect in mediation analysis	Mark Lachowicz, Kristopher Preacher, Ken Kelley
Independence Tests to Determine the Direction of Effects in Mediation Models	Wolfgang Wiedermann
Is mediation still evolving? A review of recent advances in understanding and uncovering indirect effects	Emil Coman, Victor Villagra Maria Coman, Suzanne L. Suggs

Session 6B: Bayesian Item Response Theory
Laurel Hall 301

Paper	Authors
Bayesian Analysis of Joint Modeling of Response Times with Dynamic Latent Ability	Abhisek Saha Xiaojing Wang Dipak K. Dey
Bayesian Multilevel Modeling with Small Cluster Sizes	Jiaqi Zhang

Session 6C: Recent developments in latent variable modeling Symposium
Laurel Hall 305

Paper	Authors
Paper 1: *Evaluating the Performance of Latent Class Modeling vs. K-Means and Extended K-Means Clustering*	Jay Magidson
Paper 2: *Assigning New Cases to the Appropriate Latent Classes Using 1-step Scoring in Latent GOLD 5.1*	Jay Magidson Jeroen Vermunt
Paper 3: *Prior Sensitivity Analysis in Default Bayesian Structural Equation Modeling*	Sara van Erp Joris Mulder Daniel Oberski

Session 6D: Issues in Multilevel Modeling
Laurel Hall 306

Paper	Authors
What You Don't Know Can Hurt You: Adjusting for Endogeneity in Multilevel Modeling	Lynn Foster-Johnson Jeffrey D. Kromrey
The Effect of Multicollinearity on a Multilevel Structural Regression Model	Seda Can Oya Somer
Data Sparseness and Two-Level Logistic Multilevel Models: Impact of Cluster Size on Fixed and Random Effects	Bethany A. Bell Jason A. Schoeneberger Zhaoxia Guo Jeffrey D. Kromrey John M. Ferron

Session 6E: Mixture Modeling Applications
Laurel Hall 106

Paper	Authors
Does hearing fluctuate over time in Meniere's Disease? A growth mixture modeling approach for categorizing patterns of change over time.	Laurel Fisher Michael Hoa
Growth Mixture Modeling in Self-Rated Health and Depression	Joseph W. Jones Thomas Ledermann Ginger Lockhart
Modelling menstrual cycle length and variability at the approach of menopause by using hierarchical change point models	Michael Elliott Xiaobi Huang Sioban Harlow

Session 6F: Innovations in large data techniques
Laurel Hall 110

Paper	Authors
Evaluating Monte Carlo Simulation Through Data Mining Approaches	Oscar Gonzalez
Functional Parallel Factor Analysis for Functions of One- and Two-dimensional Arguments	Ji Yeh Choi
Asymptotic efficiency of the pseudo-maximum likelihood estimator in multi-group factor models with pooled data	Fan Yang Wallentin

---

Keynote by Andrew Gelman
1:00 pm – 2:30 pm
Laurel Hall 102

Concurrent paper session 7 2:45 pm – 4:15 pm

Session 7A: Modeling Innovations in Social Science
Laurel Hall 205

Paper	Authors
Structural Equation Modeling in Archival Accounting Research: An Application to Disclosure and Cost of Capital	Steven Utke Lisa Hinson
Illustration of the novel Mplus' SKEWT option for testing for randomized intervention effects when outcomes are skewed. How much did the eConsult trial reduced costs of care in treated patients	Emil Coman, Victor Villagra, Daren Anderson, Ianita Zlateva
Models for detecting nuanced health disparities (HD). Assessing the size of HDs and when and how they are reduced or not	Emil Coman, Victor Villagra, Daren Anderson, Ianita Zlateva

Session 7B: Tutorial: Conducting propensity score analysis with SEM using MPLUS
Laurel Hall 301

Paper	Authors
Tutorial: Conducting propensity score analysis with structural equation models using MPLUS	Walter Leite Laura Stapleton

Session 7C: Issues in Multilevel Modeling
Laurel Hall 305

Paper	Authors
Efficient estimation of variance components in nonparametric mixed-effects models with large samples	Nathaniel E. Helwig
Bootstrapping for Two-Level Synthesis of Single-Case Experimental Data	Mariola Moeyaert
The package does matter: A comparison of multilevel modeling in 5 common software packages	D. Betsy McCoach, Graham Rifenbark, Aarti Bellara, Xioran Li, Sarah Newton, Janice Kooken, Anthony Gambino, Dani Yomtov

Session 7D: Effect Sizes and Hypothesis Tests
Laurel Hall 306

Paper	Authors
Bayes Factor null hypothesis tests are still null hypothesis tests	Matt N. Williams
Asymptotic efficiency of the pseudo-maximum likelihood estimator in multi-group factor models with pooled data	Fan Yang Wallentin

Thursday – May 26
Post-Conference Workshop
8:30am- 4:30pm

Mixture modeling with correlated data

Jeff Harring, University of Maryland
Business Room 106

The post-conference workshop deals with model-based clustering methods, commonly referred to as finite mixture modeling. These methods have been applied to a wide variety of cross-sectional and longitudinal data to account for heterogeneity in population characteristics. This workshop is intended as both a theoretical and practical introduction to finite mixture modeling as it pertains to statistical methods used to model correlated data. It is assumed that workshop attendees are well-grounded in basic statistical methods and have had some exposure to more advanced quantitative methods such as structural equation modeling and multilevel modeling. Although finite mixture modeling has proven advantageous across many disciplines, the examples used in the workshop draw primarily from social science research, including the fields of education and psychology. Models will be presented in several formats — path diagrams, equations, and software syntax. Data, annotated Mplus script files along with annotated output for all of the examples will be provided. It is not required to bring Mplus on a laptop to attend the workshop — participants will still benefit from a comprehensive set of slides, input files and data so that they could practice running workshop examples on their own. Participants who have Mplus base package + combination add-on will be able to run the examples in real time during the workshop.

DATIC Workshops 2016

We are proud to announce two new additions to this year's workshops at the Data Analytic Training Institute of Connecticut, held every June at the University of Connecticut.

Generalized Linear Mixed Models (GLMM)

Instructor: *Ann A. O'Connell* **June 27-28, 2016**

This short-course (2 days) covers extensions of mixed and hierarchical linear models for outcome variables that represent dichotomous, ordinal, multinomial, or count data. We being with a review of single-level generalized linear models in terms of estimation, interpretation, and model fit. We then expand on this foundation to build and interpret generalized linear mixed models. Emphasis will be on model building, interpretation, comparison, and the use of analysis adjustments for the limited nature of these kinds of dependent variables. Emphasis is on application and interpretation, with hands-on analyses and examples from the education, health, and social/behavioral sciences literature. Participant background should include regression, analysis of variance, and some exposure to multilevel modeling. Software for examples will include HLM, SAS, and SPSS.

DATIC

Longitudinal Modeling using Mplus

Longitudinal Modeling using Mplus
Instructor: *D. Betsy McCoach* **June 29-July 1, 2016**

During this three day workshop, students will learn how to model longitudinal data using Mplus. The main focus of the workshop focuses is on fitting growth curve models in Mplus. Specifically, we will cover linear, polynomial, multiphase, non-linear growth curve models, and multivariate growth curve models, and latent change score models for both observed variables and latent constructs.

Topics to be covered include

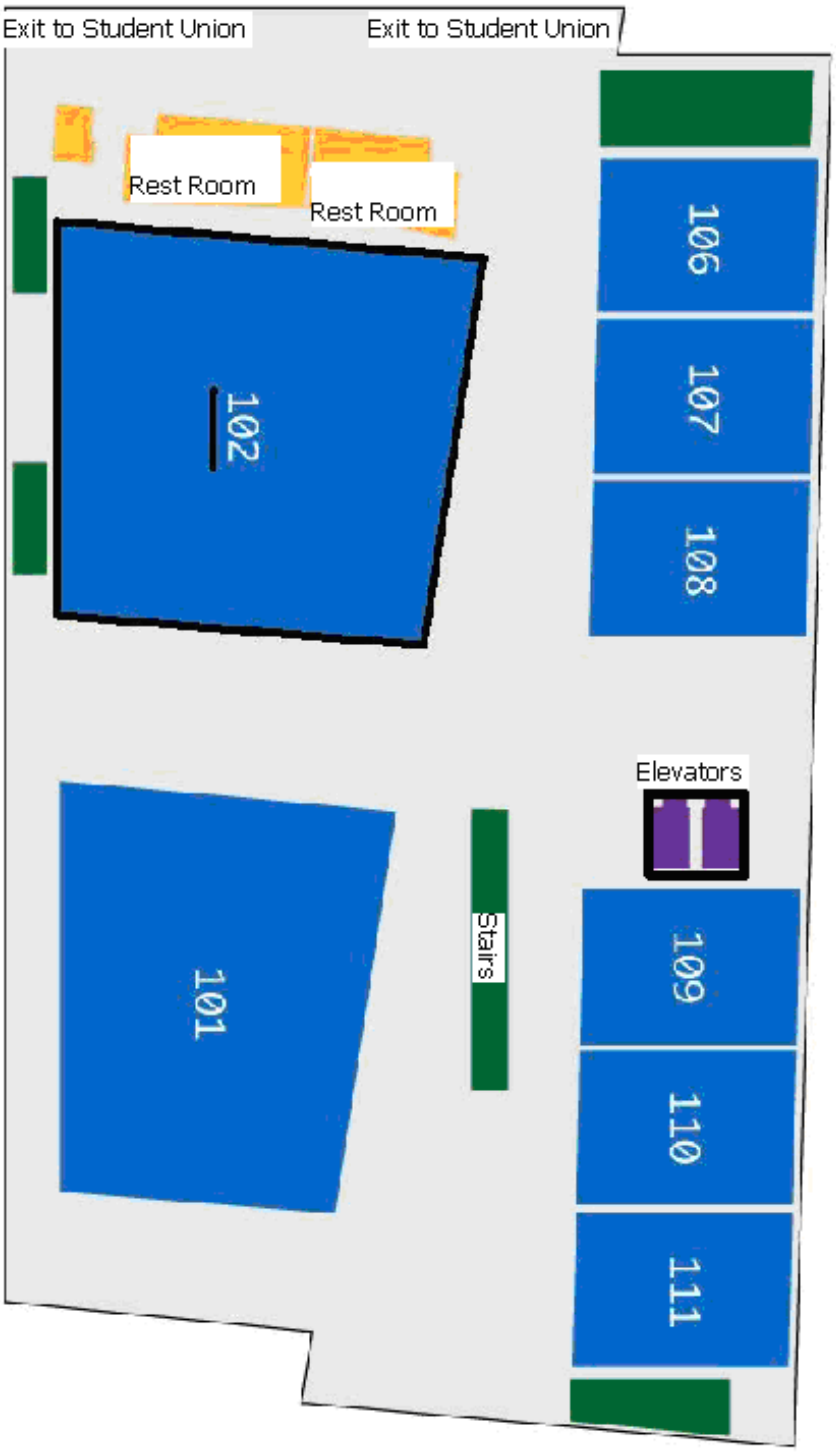
- Time unstructured vs. Time structured data
- Centering and coding variables for longitudinal models
- Estimating Linear growth curve models
- Time Varying Covariates
- Estimating multiphase growth curve models
- Estimating Polynomial growth curve models
- Latent Basis growth curve Models
- Estimating other types of non-linear growth curve models
- Latent Change Score Models
- Multivariate Models
- Factorial invariance across time

UCONN
NEAG SCHOOL OF EDUCATION

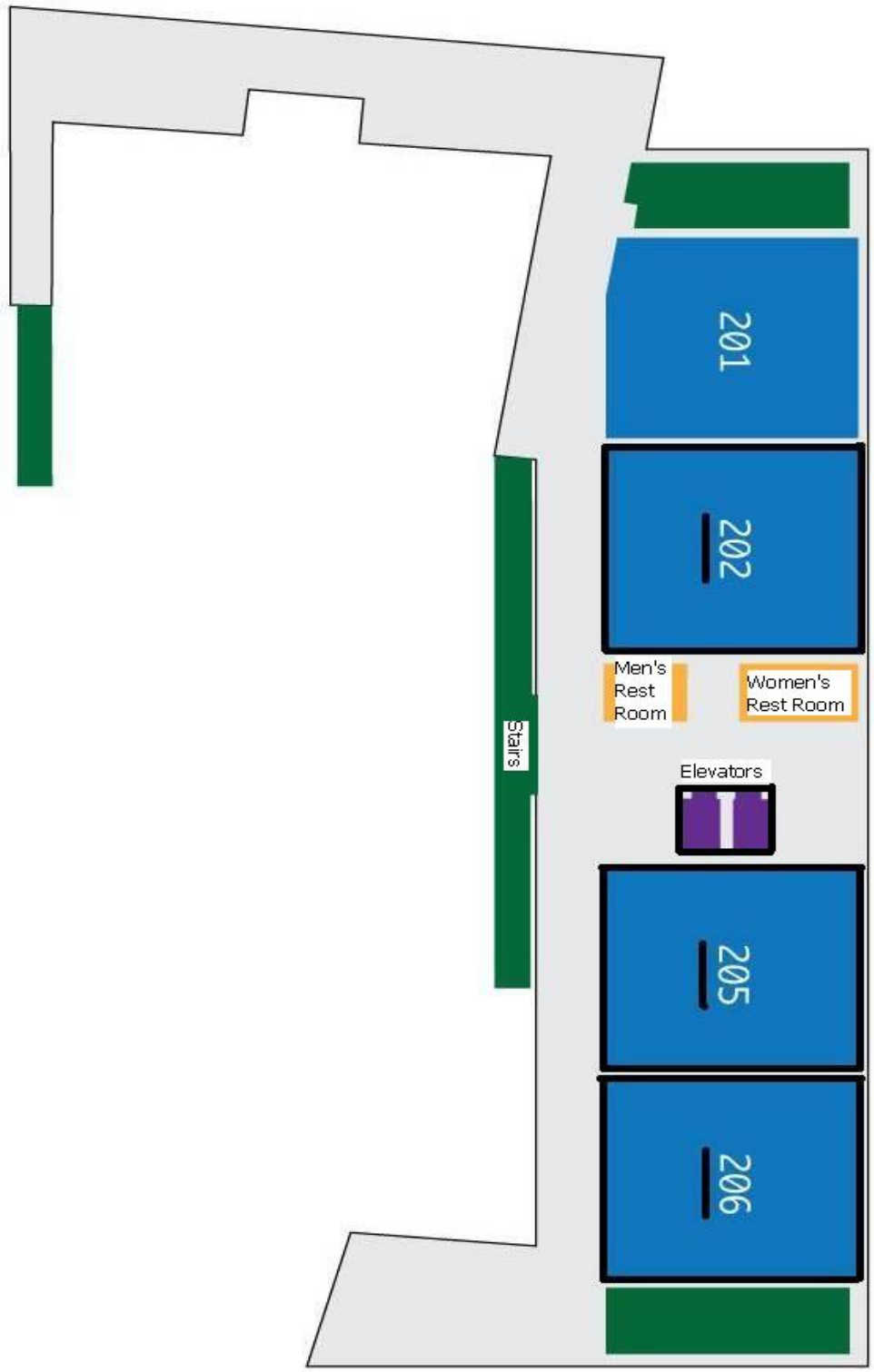
*Register online at www.DATIC.UCONN.edu

Maps

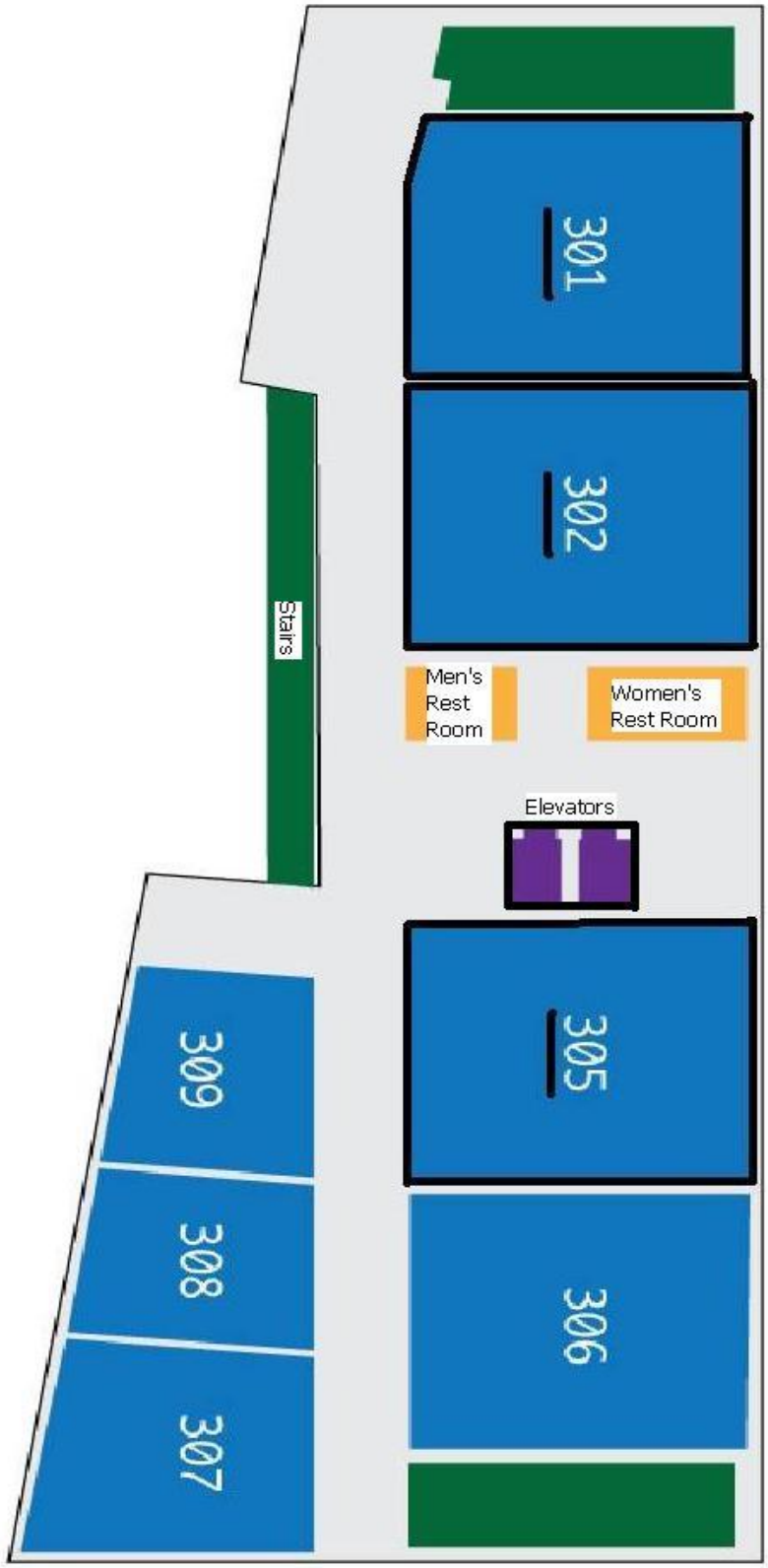
Laurel Hall (Classroom Building) Floor Plans:
1st Floor



Laurel Hall (Classroom Building) Floor Plans:
2nd Floor



Laurel Hall (Classroom Building) Floor Plans:
3rd Floor



UCONN campus:

Paper Abstracts

Concurrent paper session 1 10:30 am – 12:00 pm

Session 1A: Propensity Score and Group Balance methods

In the Pursuit of Group Balance on a Known Covariate in a Small-Scale Experiment: A Monte Carlo Study of the Efficacy of Standard Approaches

Tyler Hicks, Jeffrey Kromrey

Abstract: In multilevel studies, a contextual variable reflects a group-level characteristic (such as the organizational climate) that is measured by an individual-level variable (such as the perceived climate). In addition, the group-mean of the individual-level variable is used as the proxy of the group-level characteristic. Accordingly, the contextual effect is defined as the pure effect of the contextual variable (Z) on the outcome (Y) after removing the impact of the explanative variable at individual level (X). The multilevel model with contextual effect is often of particular interest in the fields of education and psychology in that the group-mean of the predictor at the individual level reflects the situational influence. In practice and technical consideration of multilevel modeling, a strong intra-class correlation (ICC) at the measure of outcome is the basic requirement. However, in terms of the definition of the contextual effect, both the ICC of outcome (ICCy) and the ICC of the predictor (ICCx) need to be considered simultaneously in a multilevel model. This issue has not been mentioned in the literature. In the present study, a Monte Carlo simulation along with an empirical data containing 38 companies and 1200 employees are adapted to explore the influences of both ICCx and ICCy on the estimation of contextual effect. In particular, the Bayesian estimation is applied to compare with the traditional maximum likelihood method. Results of simulation study reveal that, in the cases of small sample size, a smaller ICCx combined with a higher ICCy is more efficient for the parameter estimation; in contrast, a higher ICCx combined with a lower ICCy has poor performance. The performance of the Bayesian estimation is similar to the maximum likelihood method. However, the Bayesian estimation shows the superiority of predicting the true value of the parameters, especially when the Ncluster is low, revealing that the Bayesian method is a good alternative to the maximum likelihood method for estimating the contextual effects in the multilevel models. This paper shows the importance of appropriately matching ICCx with ICCy in the multilevel model. Limitations and suggestions for further study are also discussed.

Examining Propensity Score Estimation Models and Conditioning Methods using Multilevel Modeling

Aarti Bellara

Abstract: Propensity score methods have been used to minimize the selection bias in observational studies to identify causal relationships. Propensity score analysis aims to use the estimate to create balanced groups, akin to a randomized experiment. This study used Monte Carlo methods to examine the appropriateness of using propensity score methods to achieve balance between groups on observed covariates and reproduce treatment effect estimates in multilevel studies. The results of this study suggest that the degree to which propensity score analyses are able to create balanced groups and reproduce treatment effect estimates with clustered data is largely dependent upon the propensity score estimation model and conditioning method selected.

The Use of Nonparametric Propensity Score Estimation with Data Obtained Using a Complex Sampling Design

Ji An, Laura M. Stapleton

Abstract: The propensity score (PS) method is a widely used approach in making causal inferences in education. The application of the PS method in data selected based on complex sampling (CS) designs has attracted increased attention, however, no research has explored the use of nonparametric methods (i.e., data-mining methods) for estimation in CS data settings. The purpose of this study is to evaluate (through simulation) the performance of random forests and boosted regression trees for PS estimation, with the hope of finding alternatives that are more accurate and efficient for CS data.

Session 1B: New Directions in Item Response Theory

Rethinking Complexity in Item Response Theory Models
Wes Bonifay
Abstract: The notion of complexity commonly refers to the number of freely estimated parameters in a model. An investigation of five popular measurement models suggests that complexity in item response theory (IRT) should be defined not by the number of parameters, but instead by the functional form of the model.

Modeling Directional Local Item Dependence
Kaiwen Man, Hong Jiao, Meng Qiu
Abstract: Local item dependence (LID) is likely to be present when a test is constructed of items that are clustered around a testlet. A testlet (Wainer & Kiely, 1987) is defined as a common stimulus where multiple items are constructed based on like in a passage-based reading comprehension test. In testlets, local item dependence is most often non-directional. Another scenario is the multipart items where the answer to the second part of an item depends on the answer to the first part of the item. This directional local item dependence is less studied in literature. This study proposes a new approach to accounting for directional local dependence in multipart items. The new method models directional local item dependence between two items by estimating the correlation of difficulty parameters of two items that are dependent by assuming the bivariate normal relationship among the difficulty parameters. Markov Chain Monte Carlo method is used for model parameter estimation. Model parameter recovery is evaluated in a simulation study in terms of item and ability parameter estimation.

Hyperparameters and Hyperpriors in Item Response Theory: Interpretations and Use
Allison Ames
Abstract: Bayesian estimation methods require specification of a likelihood model as well as specification of model parameter priors. Parameters of the prior distribution, hyperparameters, can have priors themselves, called hyperpriors. This methodological illustration presents estimation of hyperparameters via hyperpriors for IRT models, as well as their use and interpretation. Several considerations will influence the estimation of hyperparameters (e.g., sample size, number of items, model-fit) and resulting parameter shrinkage, and these scenarios are also presented. Little technical knowledge of Bayes methods will be required to follow the illustration, which will be presented almost entirely using graphical methods and interpretation of output.

Session 1C: Multiple Imputation Methods

Multiple Imputation in Finite Mixture Modeling
Daniel Y. Lee
Abstract: In situations where one might opt to use multiple imputation (MI) for dealing with missing data, suggestions have been made on how to appropriately impute data in the context of multiple groups. The literature, however, does not offer much information about the consequences of MI techniques in instances where groups are unknown but are suspected to exist, as is the case with finite mixture models. The purpose of this study was to examine the accuracy of parameter recovery using MI for a finite mixture factor model. Through Monte Carlo simulation, it was shown that MI produced bias estimates even when the correct number of groups was known. The simulation also revealed that the results may have been subject to label switching, making results unreliable and unverifiable.

Evaluation of Multiple-Imputation Procedures for Handling Missing Data in Multivariate Time Series Models
Linying Ji, Sy-Miin Chow, Nicholas C. Jacobson
Abstract: The availability of intensive longitudinal data has helped spurred the use more sophisticated methods for studying change. Unfortunately, missing data issues also arise frequently in such studies. Conventional missing data approaches are fraught with additional computational challenges when applied to intensive longitudinal data, and may not always be applicable due to the broad-ranging measurement characteristics of the covariates. In this study, we consider and illustrate the use of two approaches for implementing multiple imputing (MI) to cope with the missingness in fitting multivariate time series models, including a full MI approach, in which all missing cases are imputed simultaneously, and a partial MI approach, in which missing covariates are imputed with multiple imputation, while missingness in dependent variables are handled with full information maximum likelihood estimation. The advantages

and limitations of each approach under different missing data mechanisms are evaluated using a simulation study. Implementation of the procedure in R is illustrated using empirical data.

Estimating Latent Variable Interactions with Incomplete Exogenous Indicators via CART in MICE using Unconstrained Product Indicator Approach

Evgeniya Reshetnyak, Heining Cham

Abstract: The purpose of our simulation-based research is to address the gap in the existing methodological non-linear SEM literature, and to evaluate the performance of CART in MICE in handling missing data in the context of estimation of latent variables interaction effect using Unconstrained Product Indicator method (UPI; Marsh et al., 2004). Another goal of our study is to evaluate the sensitivity of the proposed approach to the deviations from normality of exogenous items. Because CART automatically captures non-linear effects in the data, CART incorporated in MICE may be an attractive alternative to existing missing data handling methods in non-linear SEM. The results of our study demonstrate good performance of the proposed procedure. We provide practical recommendations based on our findings.

Session 1D: Innovations in Structural Equation Modeling

Non-Linear Growth Models as Measurement Models: A Second-Order Growth Curve Model for Measuring Potential

Dan McNeish, Denis Dumas

Abstract: Recent methodological work has highlighted the promise of non-linear growth models for substantive problems in behavioral sciences. In this paper, we outline such a model to measure developing constructs – which are notoriously difficult concepts to measure because they cannot be assessed at a single time-point – focusing on human potential, in particular. Much research in substantive literatures has conceptually explored issues surrounding what potential is and how one might begin to measure it, but empirical work on potential has been relatively limited from a methodological perspective. this paper looks to vastly expand upon previous attempts to rigorously measure potential. We first review the key literature on the measurement of potential and describe a general second-order, non-linear growth model that estimates these relevant components. We provide a simple illustrative example using the public-use Early Childhood Longitudinal Study – Kindergarten dataset to demonstrate our model as applied to measuring academic potential.

Heteroscedastic-Consistent Confidence Intervals for Standardized Regression Coefficients

Paul Dudgeon

Abstract: Yuan and Chan (2011) rigorously derived consistent confidence intervals for standardized regression coefficients under general conditions. They demonstrated that large-sample confidence intervals for unstandardized regression coefficients could not in general be simply rescaled and applied to standardized coefficients when normality assumptions hold. Jones and Waller (2015) extended these development to circumstances where data are non-normal by applying Browne's (1984) asymptotic distribution-free (ADF) theory. This talk proposes heteroscedastic-consistent (HC) estimators as potentially better solutions for constructing confidence intervals on standardized regression coefficients under non-normality. A new method for estimating HC standard errors based on misspecified structural equation modelling is proposed (Yuan & Hayashi, 2006). Both the ADF and HC estimators are evaluated in a Monte Carlo simulation. Findings confirm the superiority of a subset of HC estimators over the ADF estimator and over Yuan and Chan's normal theory method. Possible extensions of the HC estimator method to other effect sizes in linear regression and to other modelling contexts will be considered.

Session 1E: Network and Spatial Analysis

Modeling the Effects of Network Attributes on Subgroup Integration

Qiwen Zheng, Tracy Sweet

Abstract: Subgroup structure is an inherent characteristic of most types of social networks in which ties are more common among actors within a subgroup than between actors in any two different subgroups. Subgroup integration describes the degree of separation across subgroups; more ties for actors in different subgroups indicates higher subgroup integration. In this paper, we incorporate network level covariates in the hierarchical mixed membership stochastic block model (HMMSBM) to estimate the effects of network attributes on subgroup integration. In addition, we present several simulation studies to evaluate the performance of HMMSBM with network-level covariates as well as a real-world data

example to relate teacher practices on students' friendship network integration and provide insight of how this model can be used in practice.

Individuals as Dynamic Networks: A Merging of Intraindividual Variability, Network Analysis, and Experience Sampling

Xiao Yang, David Lydon, Nilam Ram

Abstract: Lifespan developmental theories view persons as dynamic systems, with thoughts, feelings, and actions that are interconnected and change over time. The predominant approaches for analyzing experience sampling data, however, typically only model relations among two or three variables. In this paper we forward an approach that merges intraindividual variability/change and social network methods in order to chart and describe the interplay among many aspects of function and how those dynamics change over time. Data come from "stable" and "turbulent" individuals who reported on their feelings, behaviors, and actions over 9 weeks that were or were not disrupted by a major life event. Person-specific system dynamics are modeled with a network of time-lagged and contemporaneous relations. Results illustrate how the method is used to describe dynamics of high-dimensional systems, chart how those systems develop or remain stable over the long-term and quantify the extent to which the systems differ in structure and extent of change. As more experience sampling data become available the combined intraindividual variability/change + social network analysis approach provides new opportunities to articulate and test hypotheses about individuals as complex dynamic systems – including how behavioral dynamics differ across context, change with age, and are disrupted by major health events.

Session 1F: Longitudinal Mediation Models

Multilevel autoregressive mediation models: Specification, estimation, and applications

Qian Zhang, Lijuan Wang, Cindy Bergeman

Abstract: Traditional mediation models may either overlook the multilevel structure of the data set, or the temporal sequence of X, M, and Y for causation to take place. In the study, extending from the cross-lagged panel models (CLPMs) in Maxwell, Cole, & Mitchell (2011), we proposed the multilevel autoregressive mediation models (MAMMs) by allowing all the coefficients to be random across individuals. In addition, Level-2 covariates can be included to partly explain the inter-individual differences of mediation effects. Given the complexity of the model structure, Bayesian estimation was used. To illustrate the estimation procedure, both CLPM and MAMM were fitted to a daily diary data set. It was found that the two models yielded different statistical conclusions regarding the average mediation effect. In light of this, a simulation study was then conducted to examine the estimation accuracy of Bayesian estimation for MAMMs and consequences of model mis-specifications. Factors considered include sample sizes (N), numbers of time points (T), fixed indirect and direct effect sizes, and Level-2 variances and covariances. Results indicated that estimation of fixed effects for the indirect effect components and the conditional effects of a Level-2 covariate were accurate for $N \geq 50$ and $T \geq 5$. In terms of the estimates for Level-2 variances and covariances, they were accurate provided a sufficiently large N and T (e.g., N>400 and $T \geq 50$), and less accurate when N and T became smaller. Estimates of the average mediation effect were mostly accurate with moderately large $N \times T$ (e.g., 1,000). Furthermore, with sufficiently large N and T (e.g., N=400 and T=50), fixed-effect estimates and the average mediation effect were accurate when Level-2 variances were zero and MAMMs were fitted, and they could be inaccurate when CLPM was used and random effects existed; in addition, DIC can be used to select the correct MAMM with non-zero Level-2 variances. Limitations and future directions were discussed.

Random-Effects Cross Lagged Panel Models for Longitudinal Mediation Analysis

Wei Wu

Abstract: This presentation propose a random-effects cross lagged panel model (RECLPM) to model longitudinal mediation, by taking advantage of the recent advances on multilevel SEM. RECLPM is a directly extension of the fixed-effects cross lagged panel model (FECLPM) by allowing any of the effects involved in a mediation analysis to be random. A simulation study is conducted to compare the performance of RECLPM to FECLPM and a random-effects model proposed in Bauer, Preacher, and Gil (2006). The preliminary result suggested that RECLPM outperformed both models in recovering the mean mediational effect when the effect varied across individuals.

Session 1G: Identifying Informative Subgroup Typologies in K-12 Education Leadership Data through the Application of Latent Class Analysis (LCA) and Multilevel LCA

Paper 1: The influence of typologies of school leaders on teacher retention: A multilevel latent class analysis by *Angela Urick*
Paper 2: Principal Turnover: Are there Different Types of Principals Who Move From or Leave Their Schools? A Latent Class Analysis of the 2007-08 Schools and Staffing Survey and the 2008-09 Principal Follow-up Survey by *Jared Boyce*
Paper 3: How Leaders Agree with Teachers in Schools on Measures of Leadership Practice: A Two-Level Latent Class Analysis of the Comprehensive Assessment of Leadership for Learning by *Alex J. Bowers*

In this symposium, the authors will present three research studies that apply latent class analysis (LCA) and multilevel LCA to the issue of identifying significantly different subgroup typologies of people in K-12 education data, including teachers and principals. The session will first open with a brief overview of applying LCA to typology research in K-12 education, and then proceed with a discussion of three different applications of LCA in this domain: 1) to what extent are there different subgroups of principals in their perception of their leadership, how do these different subgroups influence different subgroups of teacher perceptions of leadership in their school and to what extent do these subgroups influence teacher turnover; 2) to what extent are there different types of principals who exit their schools; 3) and to what extent are there different levels of alignment between different subgroups of teachers' and principals' perception of leadership in K-12 schools. The session will then end with a discussion between the authors and the audience on novel applications of LCA to these types of social science data and research questions.

Concurrent paper session 2 — 1:15 pm - 2:45 pm

Session 2A: Missing Data: Issues and Opportunities

Don't Be Fancy. Impute Your Dependent Variables!
Kyle M. Lang, Todd D. Little
Abstract: Common wisdom among missing data researchers prescribes the general use of multiple imputation (MI) or full information maximum likelihood (FIML) to treat missing data in applied research (Enders, 2010; Little & Rubin, 2002). Yet, special situations in multiple linear regression (MLR) modeling (e.g., listwise deletion [LWD]-based slopes being unbiased when missingness depends only on the predictors) continue to fuel debate regarding exactly how to address missing data in MLR models (e.g., Little, 1992; von Hippel, 2007). Some work has suggested that observations with imputed DVs should be excluded from the inferential regression models (MI and Delete, MID; von Hippel, 2007) because these observations contain no information about the model's slope coefficients. Although this argument holds in specialized situations (i.e., when the missingness predictors are fully represented within the analysis model's predictor set or when the true predictors are independent), we use Monte Carlo simulation to demonstrate why this approach, and LWD, are unsound in real-world MLR modeling. We also show that employing a sufficient number of imputations (e.g., m = 100) removes potential problems caused by retaining observations with imputed DVs.

Missing Data in Dyadic Modeling: Issues and Opportunities
Alexander M. Schoemann, John K. Sakaluk
Abstract: Missing data is a problem almost all researchers in psychology must confront when designing studies and analyzing results. Modern missing data techniques such as full information maximum likelihood (FIML) and multiple imputation (MI) provide researchers with tools to accurately and powerfully estimate parameters in the presence of ignorable missing data. Despite these advances modern missing data techniques, they are rarely discussed by researchers engaged in dyadic data analyses. Dyadic data provide unique challenges (e.g. indistinguishable dyads) and opportunities (e.g. unique planned missing data designs) in regards to missing data. In this talk we will discuss these challenges and opportunities and illustrate them with simulation studies and analysis of real world data.

Confirmatory Factor Analysis with Planned Missing Data: A Monte Carlo Simulation Study

Mine Dogucu, Hui Jiang, Pamela Soto-Ramirez

Abstract: Three planned missing designs (3-form, fractional block and a control design) were compared in a confirmatory factor analysis model based on a simulated data. Multiple imputations and full information maximum likelihood were the two missing data treatment methods used. Different sample size conditions were tested. Missing designs and treatment methods were compared based on model convergence rates and parameter estimations.

Session 2B: Repeated Measures Approaches and Issues

Misleading Results From Combining Predictions of Residualized and Simple Gain Scores in the Same Longitudinal Analyses

Robert E. Larzelere, Isaac J. Washburn, Mwarumba Mwavita, Ronald B. Cox, Jr. Taren Swindel

Abstract: This paper illustrates some problems that can occur when longitudinal analyses combine predictions of simple gain scores and residualized gain scores in the same analysis. First, when including the initial score as a predictor of a simple gain score, the regression coefficient for any purported cause will be identical to the coefficient predicting residualized gain scores. More complex models that combine latent growth modeling with cross-lagged coefficients can produce misleading coefficients, because each time-ordered coefficient is controlling for the other one. This is illustrated with the autoregressive latent trajectory model and the bidirectional dual change model.

Difference-in-Difference Estimates of Demographic Processes

Lawrence L. Wu

Abstract: In this paper, we consider the use of difference-in-difference (DD) estimation procedures when a natural experiment is thought to influence a demographic process observed for successive cohorts in a population. We contrast the estimates reported in most existing studies with what we argue is a more natural DD life-table estimator specified within a proportional hazard framework. We show that the former does not possess a natural life-table interpretation, but instead involves weighted averages of underlying life table parameters. These weighted averages will moreover evolve with time as new cohorts are born and as existing cohorts undergo differential exposure to treatment, thus complicating the interpretation of standard DD estimates of the effect of treatment. Under proportionality, we show how the standard DD estimator will yield different estimates depending on the form of age or duration dependence in the underlying baseline hazard. Our proposed DD life-table estimator does not suffer from these issues and thus may be useful to researchers wishing to exploit a natural experiment to study demographic processes such as fertility, mortality, migration, or the formation and dissolution of marital and cohabiting unions.

Session 2C: Propensity Score Methods

Optimizing Random Forests Propensity Scores

Heining Cham, Landon Hurley, Yue Teng

Abstract: Random Forests is a statistical learning method which has been proposed for propensity score estimation models that involve complex interactions, nonlinear relationships, or both of the covariates (Austin, 2012; Lee, Lessler, Stuart, 2010). Cham (2013) investigated the effects of three random forests model tuning parameters on propensity score estimation: (1) decision rules to select the covariate and its split value in a classification tree, (2) the number of covariates randomly sampled for selection, and (3) methods of estimating. The results concluded that the optimal specification with the least bias of the average treatment effect estimates depend on the nature of data and different propensity scores adjustment methods. This proposed study extends Cham (2013) findings and investigates various propensity score models and propensity score equating methods. We hypothesize that the optimal application of random forests propensity scores will be with the propensity score equating methods that are more robust to misspecified propensity scores (genetic matching and weighted residual bias corrections). The expected research findings will provide a more automatic procedures for propensity score analysis.

A simple fix for bias due to a latent/mismeasured covariate in propensity score weighting analysis: Factor scores from models inclusive treatment and other covariates

Trang Quynh Nguyen, Cyrus Ebnesajjad, Hwanhee Hong, Elizabeth A Stuart

Abstract: Propensity score modeling assume that covariates are measured without error. This is problematic when a covariate X is measured with error or is a latent variable. It has been suggested that if multiple measurements Ws of X are available, rather than using Ws in propensity score analysis, it is better to use Ws in a latent factor model, estimate a factor score (FS) and use this FS to represent X in modeling the treatment and outcome. Evaluation of this approach to date is limited. In the context of propensity score weighting for a binary treatment T, two continuous covariates including one (Z) observed directly and error-free and one (X) observed indirectly through error-prone measurements Ws, we investigate this method (the simple FS) and two other types of FSs generated from structural equation models (SEM) (or mathematically comparable factor models) that also include T (treatment-inclusive FSs) and Z (treatment-&-covariates-inclusive FSs). We investigate situations with continuous and ordinal Ws, where the measurement errors are normally distributed, non-differential with respect to T, Z and the outcome Y, and uncorrelated across the Ws. We consider logit, probit, and identity links for T and ordinal Ws. For all methods, the larger the number of Ws and the more correlated they are with X, the smaller the bias; continuous Ws result in less bias than ordinal Ws. Comparing methods, the simple FS method does not improve over the direct Ws method in terms of bias; the treatment-inclusive FS is an improvement but is still biased, especially when X and Z are uncorrelated; and the treatment-&-covariates-inclusive FS method effectively removes bias. Focusing on the treatment-&-covariates-inclusive method, we investigate situations where there is differentiation in performance among the different FSs; and evaluate bootstrapping as a strategy to improve interval estimation in situations where coverage of the outcome model estimated confidence interval is less than nominal, including small numbers of Ws, low W-X correlations, large effect of X on T, and treatment prevalence far from .5. We conclude that a simple fix for the bias due to a latent/mismeasured covariate in propensity score weighting analyses is to represent the covariate using a treatment-&-covariates-inclusive FS. For illustration, we apply the method to estimating the effect of smoking in adolescence on later life depression where a covariate is baseline depressive symptom severity, a latent construct measured using a multi-item scale.

Investigating Causal DIF via Propensity Scores Methods

Yan Liu, Chanmin Kim, Amery Wu, Paul Gustafson, Bruno Zumbo

Abstract: Once a method flags an item as DIF, it is difficult to conclude that the grouping variable (test language) is responsible for the DIF result because there may exist many confounding variables that lead to the DIF result. The present study is to demonstrate the application of propensity score methods in psychometric research on DIF for day-to-day researchers. In addition, the follow-up simulation study will help researchers to understand how different scenarios of model misspecification will affect the performance of logistic regression DIF analysis with propensity score matching.

Session 2D: State of the Art in Dynamic SEM

Latent class analysis for intensive longitudinal data, Hidden Markov processes, Regime switching models and Dynamic Structural Equations in Mplus 8

Tihomir Asparouhov

Abstract: In this talk we describe the general framework implemented for release in Mplus Version 8.0 and Version 8.1 which simultaneously incorporates time series modeling, hidden Markov modeling, multilevel modeling and mixture modeling. The framework can be useful to study continuous and categorical latent variable development across time. Specific examples and simulation studies are discussed such as: Multilevel Mixture Models, Multilevel Latent Transition Models with cluster specific transition probabilities, Dynamic Structural Equation Models, Multilevel Mixtures of Dynamic Structural Equation Models, Dynamic Latent Class Analysis, Multilevel Markov Switching Autoregressive and Kalman Filter Models.

Session 2E: Mixture Modeling Approaches

Testing the Effectiveness of Three-Step Approaches for Auxiliary Variables in Latent Class Analysis

Zachary K. Collier, Walter L. Leite

Abstract: Latent class analysis (LCA) is a clustering method that identifies subgroups within a given population. The inclusion of auxiliary variables in LCA has contributed to application studies by investigating the types of individuals/subjects that belong to classes. Auxiliary variables can be identified as either a predictor of the latent categorical variable or as a distal outcome. The primary purpose of this study was to determine the results of three, 3- step

methods for analyzing auxiliary variables in LCA. The 3- step approach is most notable due to its empirical ability to ensure that either type of auxiliary variable does not influence the measurement parameters of the LCA model. In addition, 3-step methods have shown to be more efficient in computation time and outputs correct estimates and standard errors compared to the alternative, single-step approaches. In this study, 3- step methods for auxiliary variables were analyzed under varying levels of sample size, effect pertaining to the type of auxiliary variable and latent categorical variable, and entropy. Each manipulated condition imitated previous LCA studies in education, sociology, psychology, survey, and medical research. A Monte Carlo simulation was conducted to compare consistency of estimates between two approaches in two-class LCA models with latent class predictors. Each of the three methods was compared in two-class models with distal outcomes. Additionally, relative bias, relative bias standard error, coverage, power, and type I error are provided since these were not given in previous published explorations on 3- step approaches. This study provided more information on the literature with reference to required sample sizes for 3- step methods, how they compared with respect to power, and whether they provided unbiased estimates across a variety of conditions. The results of this study conclude that one of the two methods for latent predictor variables yields better outcomes. However, another method for distal outcome variables showed to be the most robust to sample sizes below 500 and produced lower type one-error rates. Additionally, this study analyzes applied data using each 3- step approach.

Aberrant Behavior Detection based on a Multivariate Mixture Modeling Approach
Kaiwen Man, Jeffery Harring
Abstract: Recently, response times have been used to measure and evaluate many latent characteristics of students, and particular to this research, have been utilized to identify aberrant response behavior of examinees on educational and psychological tests. Statistical procedures based on response times provide a more flexible framework to study aberrant testing behavior of respondents than traditional IRT-based modeling. The focus of this research is to extend current response time methodology for detecting aberrant testing behavior to a multivariate mixture modeling framework, which incorporates two response time based person-fit indices. Detection of aberrant testing behavior using this new mixture modeling technique will be evaluated vis-à-vis a Monte Carlo simulation.

Nonparametric Multilevel Latent Class Analysis: A Comparative Study of Teachers' Beliefs about Teaching and Learning in 24 Countries
Chi Chang
Abstract: The purpose of this paper is to apply nonparametric multilevel latent class analysis (NP-MLCA) to the simultaneous identification of 1) subtypes of teachers' beliefs and 2) variation in these subtypes across countries. In total, 35,263 teachers in 24 countries were investigated. Seven indicators of beliefs were measured and utilized for classifying teachers. This study identified two classes of teachers and three classes of countries based on teachers' beliefs about teaching and learning. This study seeks to provide a more rigorous means of conducting comparative studies, and its adoption can be expected to provide higher quality information for policy makers.

Session 2F: Measurement and Modeling

Assessing Item Measurement Invariance in Large Scale Cross-Country Surveys. Monte Carlo Simulation Study
Artur Pokropek
Abstract: Detection of items non-equivalence in large cross-country surveys is a challenging problem as the number of groups and sample sizes are very large. Five methods of detection are discusses and evaluated using Monte Carlo simulations: MGCFA, generalized regression approach, multilevel CFA, BSEM approach and alignment optimization. The methods are assessed according to power, type I error and speed of computations. Detection of nonequivalent items is being tested in different scenarios.

Modeling individual differences in terms of process model parameters in the context of large-scale reaction time data
Zita Oravecz
Abstract: Process models propose concrete underlying mechanisms when modeling observed data. The Linear Approach to Threshold with Ergodic Rate model describes reaction time data in terms of two distinct aspects of cognitive functioning: speed of information accumulation and caution. It will be described how the model can be extended to capture individual differences in terms of these cognitive process model parameters and how predictor variables can be included to model individual differences in these. The parameter estimation is implemented in the Bayesian framework,

which allows for the cognitive process model parameters and regression coefficients to be estimated simultaneously. Parallel processing algorithms for handling large data volumes will also be discussed. The approach will be illustrated with data from a large-scale research project that uses web-based crowdsourcing to study Alzheimer's Dementia.

Assessing the convergent and discriminant validity of person x situation interaction effects: Development of a multi-method model for measuring situation-specific traits.

Fred Hintz, Christian Geiser, Saul Shiffman

Abstract: Multi-method designs are preferable in many research applications because they allow researchers to obtain multiple points of view, and determine the extent of differences between them. For situation-specific constructs, the random-fixed effects model of Latent State-trait theory (LST-RF) allows researchers to differentiate between the main effects of a situation on observations and person x situation interaction effects on observations. The proposed paper extends the LST-RF model with a multi-method component that allows researchers to analyze the extent to which method effects are stable across situations. With the LST-RF-MTMM method researchers are able to to determine method-specificity of traits, and changes in method-specificity across fixed situations, in addition to situation-specificity and person x situation interactions. An application will be presented regarding smoking cessation and mood.

Session 2G: Applications of SEM and multilevel modeling

Two illustrations of meta-analytic structural equation modeling

Suzanne Jak

Abstract: Meta-analytic structural equation modeling (MASEM) is a statistical technique to pool correlation matrices and test structural equation models on the pooled correlation matrix. In Stage 1 of MASEM, correlation matrices from independent studies are combined to obtain a pooled correlation matrix, using fixed- or random effects modeling. In Stage 2, a structural model is fitted to the pooled correlation matrix. The aim of the current presentation is to provide guidance and examples on how to apply MASEM, and how one can test moderator hypotheses about specific model parameters. I will illustrate the procedure using both fixed- and random-effects subgroup analysis with two real datasets.

Using multi-membership multi-level VAM model to evaluate teacher and school effects

Eunkyeng Baek, Chunhua Cao

Abstract: Value-added modeling (VAM) has been proposed and used to estimate school and teacher effects in many school districts in nationwide. Multilevel modeling (MLM) is one of the modeling methods that has been often implemented in VAMs. It provides much flexibility that allows to capture important factors to be controlled to estimates accurate teacher and school effects (Bryk & Raudenbush, 1987; Laird & Ware, 1982; Raudenbush & Bryk, 2002). Although simple MLM provides a great flexibility to develop VAMs, it still raises some technical issues in that it is hard to apply for some specific data structures that are commonly seen in educational settings. One of the most common data structures that can be seen in educational settings are a co-teaching setting. Using a simple MLM, it is hard to take into account this setting since the lower levels (students) are nested simultaneously within more than one upper levels (teachers). By using multi-membership multilevel modeling (MMML), this setting can be handled. MMML allows the lower levels to be nested simultaneously one or more upper levels. However, due to the complexity of the statistical technique, it has not been implemented in developing VAMs with practical student learning data by the school districts. In this paper, we will illustrate how to implement MMML to develop VAMs in educational settings that students are nested within one or more teachers.

Sensation Seeking and Risk Taking among Family Members - APIM approach

Adam Sagan, Anna Siwy-Hudowska

Abstract: The purpose of the paper is to identify the family members (mother, father and oldest child) mutual influence of sensation seeking and physical, social, financial and ethical risk taking with multilevel structural equation modeling. Several dyadic models [M-C], F-C] and [F-M] were estimated. Covariates included age, gender and socio-economic status (SES) at the family level.

Concurrent paper session 3 **3:00 pm - 4:30 pm**

Session 3A: Latent variables and causal inference

Symposium

Paper 1: Causal inference using counterfactually-defined effects in mediation analysis with latent variables. *Bengt Muthén*

Paper 2: Causal effects using counterfactuals for mediation models with a two-part mediator. *Mårten Schultzberg*

Paper 3: Latent class covariate balancing for causal inference by *Katherine Masyn & Marcus Waldman*

Paper 4: Principal stratification when the intermediate variable is continuous and measured with error by *Marcus Waldman & Katherine Masyn*

Abstract: Advances in causal inference methodology have been emerging at a rapid pace over the last three decades, with no signs of slowing. In the world of structural equation modeling and latent class analysis, one notable area of recent development has been the integration of causal inference techniques, such as propensity score matching or inverse weighting, with latent variable models, enabling causal inference when exposure, confounders, mediators, or the outcomes themselves are latent variables. Another important area of development is the use of latent variables in the service of causal inference, e.g., CACE modeling using a latent class variable to capture complier status. This paper symposium pulls together four papers that present novel approaches to causal inference with creative utilization of a general latent variable modeling framework. The first paper describes the use of a latent class variable in place of an observed nominal mediator variable to estimate causal direct and indirect effects. The second paper presents an approach that integrates two-part modeling of a mediator with floor effects into a causal mediation process. The third paper explores the use of a latent class analysis of confounders and treatment as a semi-parametric alternative to propensity scoring for obtaining subgroups balanced on covariates and homogeneous with regards to the conditional probability of treatment assignment. The fourth paper concludes the symposium, demonstrating the use of a latent class variable in the principal stratification approach to estimating associative and dis-associative effects of a continuous intervening variable in a causal mediation process.

Session 3B: Symposium - Innovations in Planned Missing Data Designs

Paper 1: How low can you go? Minimum sample size thresholds for the 6-form planned missing data design by Jacob D. Curtis, Samuel F. Meeks, Kyle M. Lang & Todd D. Little

Paper 2: Take it to the next level! Planned missing data designs at Level-2 in multilevel models by *Britt K. Gorrall, Rong Chang, Kyle M. Lang, & Todd D. Little*

Paper 3: What is the ability to impute: Entire level 1 missingness based on level 2 group membership by *Luke Waggenspack, Kyle M. Lang, & Todd D. Little*

Abstract: Research in the social and behavioral sciences in recent years has been advanced due to the development and accessibility of sophisticated modeling techniques. Presently, there is a growing expectation placed on researchers to not only collect larger amounts of data, but to do so in an efficient and fiscally-responsible manner. Due to the unavoidable obstacle of the occurrence missing data, researchers need to take a preventive approach to handle missing data. A proven and widely-known strategy to combat missing data is planned missing data (PMD) designs. However, innovations in PMD designs need to emerge in order to meet the growing demands of efficiency, time and cost reduction in modern research. To address these growing research demands, the proposed symposium complies three papers which seek to push the boundaries of planned missing data designs in a novel direction. The first paper expands current knowledge on the three-form PMD into the six-form PMD using full information maximum likelihood (FIML). The remaining two papers introduce new concepts related to form assignment in planned missing data designs. The primary questions that will be addressed are: 1. For six-form planned missing data designs, what is the minimum threshold for total sample size? 2. In multilevel models that incorporate a planned missing data design, can form assignment occur at Level-2 without adversely effecting parameter estimated and model outcomes? 3. In multilevel models that incorporate a planned missing data design, what happens when Level-1 information is missing based off Level-2 membership?

Session 3C: Longitudinal mixture modeling

Right-Sizing Growth Mixture Models for Longitudinal Data

Phillip Wood, Wolfgang Wiedermann, Douglas Steinley

Abstract: Growth mixture models are frequently operationalized as finite mixture model factor analyzers in which estimated factor variances and covariances are constrained to equality across latent subgroups and factor loadings are fixed at numeric values. In many applications, the latent subgroups so analyzed often differ only in relative elevation or a four-group "cat's cradle" solution. Such patterns may be an artifact of the misspecification of the functional form of growth over time used as the basis function for identification of latent subgroups. Factor analyzers in which factor variances and covariances as well as loadings vary across subgroups are considered. A four-step procedure based on the approach outlined by Wood, Steinley and Jackson (2015) for unconditional growth models is proposed. In the first step, the maximum dimensionality of the data is considered. In the second, a variety of growth curve models are considered as candidate basis measurement models for the identification of latent subgroups. In the third step, additional parsimony of the proposed mixture model is explored by constraining parameter estimates of mixture subgroups to equality or excising components of the measurement model. Under this procedure, traditional polynomial growth mixture models are considered as a special case although other simpler or more complex models can also be considered. Examples of the approach are illustrated with a four-group simulated data set and with data from a longitudinal study of alcohol use.

Child Types and Patterns in the Classroom: A Latent Trajectory Approach to Longitudinal Peer Sociometric Nominations over Grades

Richard Faldowski

Abstract: Sociometric nominations are one of the most widely used peer-report method for eliciting information on children's peer relationships and social behavior in classroom settings. They are typically obtained by having students in classrooms nominate their classmates on various social attributes such as "Often quiet" or "Usually plays alone," then summing the numbers of nominations each child receives for each item. Although my colleagues and I have recently proposed Poisson and negative binomial methods that can more appropriately account for the unique properties peer nomination data modeled both within and over grades, our methods, to this point, only characterize children's patterns of change post-hoc. In this presentation, we describe a new method of empirically characterizing children's patterns of change over time using Poisson or negative binomial latent growth trajectory methods. This technique may have important predictive advantages for identifying correlates and concomitants associated with children who show differing patterns of change in anxious solitude or exclusion over time.

Session 3D: Measurement Models and IRT

Using Balance Scores to Disentangle Differential Item Functioning, Item Impact, and Item Bias

Amery D. Wu, Yan Liu, Bruno D. Zumbo

Abstract: This paper proposes a unified methodology for investigating and disentangling the phenomena of differential item functioning, item impact, and item bias. The methodology utilizes the technique of matching on the balance scores, the propensity scores in the present study, to aid in ruling out the alternative explanations for the causal inferences embedded in the definitions of item impact and item bias. The methodology was illustrated with data from TIMSS 2007 Grade-8 Mathematics booklet-one items. The results showed that, of the 25 item studied, 18 items (72%) were classified as having no impact, three items (12%) were classified as having impact, and four items (16%) were classified as biased.

Accounting for Rater Heterogeneity in Systematic and Random Errors with an Item Response Rater Model

Yao Xiong

Abstract: Human raters are widely involved in educational assessment, employment performance evaluation, and licensure exams. As human ratings are inevitably contaminated with errors, studying and monitoring rater performance is important in order to provide feedback to individual raters for the improvement of rating quality. To date, the vast majority of statistical rater models designed to study rater effects are built on the Item Response Theory (IRT) framework, which have considered only rater severity/leniency parameters in the models. Other rater effects are generally not built in these models as parameters to be estimated. In the current study, an Item Response Rater Model with three rater effect parameters, namely rater severity/leniency, rater centrality, and rater random error, is proposed and evaluated in a simulation study. The preliminary simulation results show that the model is robust in detecting various rater effects.

A Multilevel Bifactor Approach to Construct Validation of Mixed-Format Scales
Yan Wang, Eun Sook Kim
Abstract: Mixed-format scales are commonly used in educational and psychological measurement. Although utilizing both positively and negatively worded items might reduce the acquiescence or agreement bias, they might involve different cognitive processes and lead to systematic bias. This study applies the multilevel bifactor approach to handling wording effects with the nested data structure. We used the 2011 TIMSS USA sample and the scale measuring students' attitudes towards mathematics. The multilevel bifactor model (bifactor at the within level) was compared with alternative models (e.g., unidimensional, correlated uniqueness models). The predictive validity of the scale was examined and compared with that when wording effects were ignored. Substantive interpretations of wording effects are discussed. Recommendations and implications for practice are provided as well.

Session 3E: Symposium - New methodologies in multivariate modeling
Moderator: Ofer Harel

Presentation 1: Network Analysis Using a Random Walk in Sparse Graphs by *Haim Ba*
Abstract: Network analysis is used to investigate and reveal certain structures and connections in society. The term "society" is interpreted broadly, and it can mean a population of individuals, components in an infrastructure grid, or genes. A network corresponds to an adjacency matrix in which element (m; n) represents the strength of the connection between nodes m and n. We are particularly interested in large networks and the challenges they present. We deal with networks in which edge weights correspond to correlations between nodes (e.g. co-expression levels in genetics). We use a random effect mixture model in order to identify spurious correlations and remove the corresponding edges from the graph. The model yields shrinkage estimation, which leads to increased power to detect significant connections between pairs of genes. We obtain a sparse adjacency matrix in which the vast majority of spurious correlations are removed. This truncated matrix is used to compute the `average commuting times' as distances, which account for indirect connections between genes. It also allows us to use random walk models in order to analyze graph-theoretic properties of nodes and of the network as a whole, as well as to test for structural differences between networks obtained from different populations. This is joint work with Seojin Bang.

Presentation 2: Robust multivariate mixture model via mean-shift penalization by *Kun Chen*
Abstract: Finite mixture models have been widely used for modeling clustered and heterogeneous populations. In the univariate case, a mixture model of three components is commonly used for modeling heavy-tailed distributions arising from multiple testing problems, in which one component represents the null and the other two capture the departures from the null at the two tails. In the multivariate case, however, a three-component mixture model may be far from adequate to characterize the entire heavy-tail behavior of a multivariate distribution, since the two extreme components now only account for two specific corners of the multivariate space. In general, failing to accommodate certain rare, nuisance but extreme observations deviating from an expected mixture structure may greatly jeopardize model estimation and inference. We propose a robust multivariate Gaussian mixture approach based on a sparse and case-specific mean-shift mixture model parameterization, for simultaneously conducting anomaly detection and robust estimation/clustering. A penalized likelihood approach is adopted to induce sparsity among the mean-shift parameters in order to distinguish the anomalies from the normal observations, and a generalized Expectation-Maximization (EM) algorithm is developed to enable stable and efficient computation. The proposed approach is shown to have interesting connections with other robust methods including the trimmed likelihood method and M-estimation. We demonstrate the efficacy of our approach in a proteomics application to detect proteins with homogeneously changed intensity levels across multiple disease-producing strains, in which our method is tailored to incorporate prior biological knowledge and is able to handle replicated and incomplete data. This is joint work with Neha Mishra, Joan Smyth, Haim Bar, Elizabeth Schifano, Lynn Kuo and Ming-Hui Chen.

Presentation 3: Robust analysis of multiple secondary outcomes in case-control studies by *Elizabeth Schifano*
Abstract: There is increasing interest in the joint analysis of multiple outcomes in the social and biological sciences, including the analysis of multiple secondary outcomes originating from case-control studies. When multiple continuous outcomes measure the same underlying trait, one can gain statistical power by assuming the exposure effect is shared across all outcomes using a scaled mean model. For secondary outcomes obtained from case-control studies, the over-sampling of the cases relative to the primary (e.g., disease) outcome can induce bias in estimated relationships between

the exposure variable(s) and secondary outcomes. We propose a weighted estimating equations approach for fitting the scaled mean model, in which the weights account for the case-control design to prevent sampling bias. The proposed weight is a population-based measure, defined as the inverse of the conditional probability of case-control status. The proposed method is robust to departures from normality of multiple continuous outcomes and the misspecification of within-individual correlation across the multiple outcomes. We show through simulation that the proposed method out-performs other methods and other weighting strategies. We illustrate our approach in a lung cancer case-control study when investigating genetic associations with multiple secondary smoking outcomes. Joint work with Chuanhua Xing(XPrecision) and Xihong Lin (Harvard School of Public Health).

Session 3F: Symposium- Advanced Quantitative Methods in Educational Research

Paper 1: Applying meta-analytic structural equation modeling (MA-SEM) to jointly model components of the Simple View of Reading (SVR) and the Not-so-Simple View of Writing (NSVW) by *Yusra Ahmed, Richard Wagner, Young-Suk Grace Kim, W. Pat Taylor*
Paper 2: Is the treatment weak or the test insensitive: How explanatory item response models can elucidate the nature of intervention effects? By *Paulina A. Kulesz, David J. Francis, Jack M. Fletcher, & Sharon Vaughn*
Paper 3: Regression discontinuity and statistical power: A more efficient design for research with selected samples by *Jeremy Miciak, W. Pat Taylor, Yusra Ahmed, & Jack M. Fletcher*
The purpose of this symposium is to describe advanced quantitative methods in educational research that hold potential for improved understanding of the relation and development of academic processes, particularly in the context of targeted academic intervention for struggling students. The three presentations emanate from statistical, conceptual, and practical challenges inherent to educational research, including: (1) poor statistical power, (2) variation in test properties, (3) individual differences in academic competency that may impact response to intervention, and (4) inadequately specified theoretical models of underlying academic processes. The presentations conclude with a discussion of possible implications and potential solutions to underlying challenges in educational research.

Concurrent paper session 4 — 8:00 am - 9:00 am

Session 4A: Bayesian Structural Equation Modeling

The Impact of Bayesian Priors on Specification Search of Structural Equation Modeling
Xinya Liang, Yanyun Yang
Abstract: Bayesian Structural Equation Modeling (BSEM; Muthén & Asparouhov, 2012) may be used in line with modification indices in specification search of structural equation models. The impact of Bayesian priors on specification search has not been thoroughly investigated. This study investigated the impact of prior distributions in specification search of small cross-loadings in confirmatory factor analysis. The design factors included sample sizes, factor structures, magnitude of primary loadings, and magnitude of cross-loadings. Seven informative prior distributions on cross-loadings were investigated. Evaluation was conducted based on model fit and model recovery. Results indicate that if the upper bound of the prior credibility interval is around the population cross-loading value, the model recovery rate is the highest and the coverage rate is also good.

An Investigation of Prior Distribution Accuracy and Robustness in SEM
James P Clifton
Abstract: The purpose of this presentation is to discuss results from a Monte Carlo study examining the impact of robust priors as a means of resolving data-prior conflicts in structural equation models (SEMs). More specifically, I plan to present findings from a simulation investigating whether the specification of priors based on heavy-tailed t-distributions can recover population parameters in confirmatory factor analysis (CFA) models more accurately than the use of conjugate normal priors, the latter of which are most commonly implemented in practical applications of SEMs (e.g., Golay, Reverte, Rossier, Favez, & Lecerf, 2013; Payandeh, Najafabadi, Najafabadi, & Farid-Rohani, 2013; Yanuar, Ibrahim, & Jemain, 2013).

Session 4B: Generalized Structured Component Analysis

Application of Generalized Structured Component Analysis to Item Response Theory
Ji Hoon Ryoo
Abstract: Item response theory (IRT) is widely used as a statistical modeling method in psychometrics, primarily to estimate item and person parameters to render the true ability interpretable. These estimations generally utilize one of three maximum likelihood estimation (MLE) methods: conditional, full, and marginal MLE. Although popular, all three depend upon distributional assumptions such as multivariate normality and therefore require large amounts of data to achieve stable estimation. In this paper, we will incorporate generalized structured component analysis (GSCA) into the common IRT model to help address this small sample and estimation issue. Moreover, GSCA differs from the conventional IRT by using the alternating least square estimation method to determine a composite component score rather than a factor score for ability.

gesca: An R Package for Generalized Structured Component Analysis
Sunmee Kim
Abstract: An R package, named gesca, was recently developed to apply Generalized Structured Component Analysis (GSCA) that is a component-based approach to structural equation modeling. In this presentation, we begin by briefly introducing the basic conceptual and technical underpinnings of GSCA and then illustrate the use of the gesca package. We will provide examples to demonstrate how the package handles the basic features of GSCA and its extensions, including constrained single- and multiple-group analysis, second-order latent variable modeling, the calculation of total and indirect effects, and missing data imputation.

Session 4C: Introduction to Relational SEM and an Efficient Computational Strategy for Relational SEM

Introduction to Relational SEM and an Efficient Computational Strategy for Relational SEM
Joshua Pritikin
Abstract: We introduce relational SEM, an adaptation of structural equation modeling to relational databases. Relational SEM is a superset of the mixed model and multilevel SEM. In addition, we introduce Rampart, a new computational strategy for frequently encountered relational SEM models. Rampart is inspired by the fact that the multivariate normal density is transparent to orthogonal rotation. Well suited to big data, Rampart becomes more effective as the size of the data set increases.

Session 4D: That's a State Space Model too!

That's a State Space Model too!
Michael D. Hunter
Abstract: State space models (SSMs) have seen heavier use in the last few decades within the social and behavioral sciences, due at least in part to the more ready availability of intensive longitudinal data. However, long methodological traditions in these sciences have used structural equation models (SEMs), and various special cases of them, to study both longitudinal processes and contemporaneous structures. This presentation unites SSM and SEM by showing that all continuous-variable SEMs are a special case of SSMs. Moreover, for SEMs that model change processes over time the SSM framework has both conceptual and computational advantages. We explore several example models in their SSM and SEM form to show their identical likelihood functions and parameter estimates while noticing the differences in their specification, interpretation, and computational time. With the firm footing that SEMs have provided and the link between SEMs and SSMs, methodological researchers can take the next step forward to extend their models in time.

Session 4E: Regression Trees for Longitudinal and Clustered Data: Methods, Applications, and Extensions

Regression Trees for Longitudinal and Clustered Data: Methods, Applications, and Extensions
Jeffrey Simonoff
Abstract: Longitudinal data, in which we observe many individuals over multiple periods are common in many applications, including economics, education, marketing, medicine, psychology, and sociology. In this talk, I present a flexible modeling approach that is specialized for longitudinal data, based on the regression tree, a regression method

that accounts for locally structured behavior and interactions among the covariates. It is shown how tree methods for non-longitudinal data can be adapted to account for the correlated structure of longitudinal data through an application of multilevel modeling. The tree can be based on the classic CART methodology, but can also use a conditional inference tree that does not suffer from selection bias (in which covariates with more possible split points are preferred). The method can be generalized to allow for parametric structure at the terminal nodes of the tree, such as linear growth curves in time, and can be used to construct diagnostics for testing the fit of a multilevel linear model, all while accounting for the longitudinal structure of the data. This is illustrated using real data examples.

Concurrent paper session 5 **9:15 am – 10:15 am**

Session 5A: Bayesian Multilevel Modeling

Bayesian Model Averaging for Multilevel Models

June Yang

Abstract: Model selection can be challenging when the relationship between the outcome and the predictors is unknown. Bayesian model averaging (BMA) may help with the decision on which predictors to include in the model. BMA does not select a single "best" model. Rather, it assumes a list of possible models, computes the posterior of each possible model, and uses it as the weight to average the posterior of the parameter from each model. This produces a posterior of the parameter to be estimated. Then the posterior mean and standard deviation can represent the final estimate of the parameter. BMA is widely applied in a variety of areas within statistics but has yet to be extending into multilevel modeling (MLM) in the literature. We combine the BMA approach and MLM and examine the performance of model averaging over a list of MLMs. Instead of implementing BMA on the full MLMs, we used a two-stage regression model as a starting point. The two-stage model estimates the regression coefficients of each model separately. It produces essentially the same results compared to the full MLM but is simpler to implement. Method: We illustrated how to conduct a BMA on a two-stage regression model using the empirical data from the State Court Processing database. This dataset contains 8446 observations nested within 39 counties. A 2-level logistic regression model with random intercepts and slopes will be estimated for these data. The model consisted of three level-1 predictors and six level-2 predictors. We first estimated the regression coefficients on level-1 models. Then we conducted BMA across level-2 models for each level-1 predictor. Finally, the averaged estimates across all level-2 models act as the final estimates of level-2 predictors. Data analysis was implemented using R and the BMA package. For each level-1 predictor, the BMA technique selected the best five models out of 27 total possible models. The cumulative model posterior probability for each level-1 model was 1. Results indicated that for the level-2 predictor CUNR had the highest probability to be included in the model for the level-1 predictors ICVS (83.6%), ITRI (56.9%) and IDET (49.1%). Other level-2 predictors had relatively lower probability (0~20.9%) to be included in the models for level-1 predictors. Combining the BMA approach and the MLM technique provided a conceptually synthesized statistical tool. Results of the empirical example suggested that BMA is an efficient way to obtain accurate parameter estimates. It incorporates the information from multiple possible models and accounts for the uncertainty from the models as well as from the parameters.

The matching effect of intra-class correlation (ICC) on the estimation of contextual effect: A Bayesian approach of multilevel modeling

Hawjeng Chiou

Abstract: In multilevel studies, a contextual variable reflects a group-level characteristic (such as the organizational climate) that is measured by an individual-level variable (such as the perceived climate). In addition, the group-mean of the individual-level variable is used as the proxy of the group-level characteristic. Accordingly, the contextual effect is defined as the pure effect of the contextual variable ( ) on the outcome (Y) after removing the impact of the explanative variable at individual level (X). The multilevel model with contextual effect is often of particular interest in the fields of education and psychology in that the group-mean of the predictor at the individual level reflects the situational influence. In practice and technical consideration of multilevel modeling, a strong intra-class correlation (ICC) at the measure of outcome is the basic requirement. However, in terms of the definition of the contextual effect, both the ICC of outcome (ICCy) and the ICC of the predictor (ICCx) need to be considered simultaneously in a multilevel model. This issue has not been mentioned in the literature. In the present study, a Monte Carlo simulation along with an empirical data containing

38 companies and 1200 employees are adapted to explore the influences of both ICCx and ICCy on the estimation of contextual effect. In particular, the Bayesian estimation is applied to compare with the traditional maximum likelihood method. Results of simulation study reveal that, in the cases of small sample size, a smaller ICCx combined with a higher ICCy is more efficient for the parameter estimation; in contrast, a higher ICCx combined with a lower ICCy has poor performance. The performance of the Bayesian estimation is similar to the maximum likelihood method. However, the Bayesian estimation shows the superiority of predicting the true value of the parameters, especially when the Ncluster is low, revealing that the Bayesian method is a good alternative to the maximum likelihood method for estimating the contextual effects in the multilevel models. This paper shows the importance of appropriately matching ICCx with ICCy in the multilevel model. The limitations and suggestions for further study are discussed at the end of the paper.

Session 5B: Analysis of multivariate longitudinal data in Stan

Tyler Matta, Quinn Lathrop

Abstract: While many research questions related change or event occurrence may be answered using univariate longitudinal models, there are certain research questions that require the joint analysis of multiple outcomes simultaneously. The random-effects framework has many appealing features when it comes to multiple outcomes including a) univariate interpretation of parameter estimates, b) accommodation of different outcomes types, and d) no assumption of balancedness. Implementing a random effects model for multivariate outcomes, however, is not as intuitive. Therefore, this methodological illustration provides a guide to fit two common multivariate models – a model for two continuous repeated measures outcomes and a model for one repeated measures outcome with a time-to-event outcome – in the probabilistic programming language Stan.

Session 5C: Dealing with missing data in propensity score analyses

Evaluating Multiple Imputation Techniques for Treating Missing Data in Propensity Score Analyses

Seang-hwane Joo, Jessica Montgomery, Yan Wang, Patricia Rodriguez de Gil, Eun Sook Kim, Jeffrey D. Kromrey

Abstract: This study compared the performance of four approaches of multiple imputation techniques for missing data in propensity score analyses. A Monte Carlo simulation study was conducted to empirically assess each method in terms of bias and variability in parameter estimates, Type I error rates, and statistical power. Several factors were manipulated: sample size (500, 1000), treatment effect magnitude (0, .05, .10, .15), correlation between covariates (0, .5), proportion of missing observations (.20, .40, .60), proportion of missing covariates (.20, .40, .60), the number of covariates (15, 30), PS methods (ANCOVA, Matching, Stratification), and missing data mechanisms (MCAR, MAR, MNAR). The missing data treatments serve as a within group factor and all samples were analyzed before missing data were imposed to provide a reference condition for the evaluation of MI effectiveness. Results for the multiple imputation approaches and results generated using listwise deletion are presented. Recommendations and implications for practice are also provided.

A comparison of methods for imputation of missing covariate data prior to propensity score analysis

Walter Leite, Burak Aydin, Sungur Gurel, Duygu Cetin-Berber

Abstract: Propensity score analysis has become a popular method for quasi-experimental evaluation of educational interventions, because it can remove confounding due to a potentially large number of covariates. However, these covariates may have missing data that have to be dealt with before propensity score analysis. This paper presents a Monte Carlo simulation study comparing imputation strategies for covariates, including single and multiple imputation, imputation with Bayesian linear regression and predictive mean matching, and with propensity score models that include dummy-indicators of missing values. The study focuses on propensity score weighting and evaluates the effects of missing data method choice on occurrence of extreme weights, covariate balance and percent bias reduction.

Session 5D: What's for dynr: A package for linear and nonlinear dynamic modeling in R

Lu Ou, Michael Hunter, Sy-Miin Chow

Abstract: The past several decades have seen the rise of intensive longitudinal data (e.g. via ecological momentary assessments) and the resulting dynamic modeling methods in social and behavioral sciences. To make the estimation of these models more accessible to researchers, we have created an R package that is based on a novel and efficient state-space estimation algorithm in C. We present a talk in two parts. The first part discusses the mathematical and computational basis used by the dynr R package. The second part uses examples to present the user interface that allows for linear or non-linear discrete- or continuous-time models of multivariate observations with Gaussian

measurement errors. Thus, the dynr package utilizes computationally efficient algorithms for a broad class of models that are increasing in use while maintaining a simple and easy-to-learn interface. Dynamic models can be formed in both discrete and continuous time. In a discrete-time model, time is indicated by integers and changes in the processes are characterized by differences observed in equally spaced data. On the contrary, continuous-time models represent time by real numbers and typically describe the temporal characteristics of the processes via differential equation models, and may be fit to equally- or irregularly-spaced data. Parameter estimation in the dynr package is handled by optimizing the prediction error decomposition function (Schweppe, 1965) – a log-likelihood function computed using one-step-ahead prediction errors from: (1) the extended Kalman filtering (EKF) method for discrete-time models; and (2) a continuous-discrete variant of the EKF (Kulikov & Kulikova, 2014) that uses numerical integration routines to handle the fitting of continuous-time models.

Session 5E: Speed Dating with Regression Methods

David J Corliss

Abstract: With so many regression procedures available for different situations, it can be difficult to know the breadth of available methods and how to select the ones to apply to a given problem. This instructional presentation offers a (very quick!) description of 23 regression-based methods. In a little less than two minute for each, key points of the methods and sample output are presented, including the names of applicable procedures and packages in R and SAS. A flowchart decision tree is provided to assist in selecting the most useful regression procedures for a given context. This is not intended to provide extensive details for using specific procedures but to illustrate which procedures could be considered in a given situation.

Session 5F: Bayesian Multivariate Multilevel Models for Pretest Posttest Data

Bayesian Multivariate Multilevel Models for Pretest Posttest Data

Jean-Paul Fox

Abstract: There is a lot of discussion about the statistical analysis for designs with pretest and one or more posttest measurements. In pretest-posttest designs, the same outcome variable is measured on both occasions, and interest is focused on the relationship between explanatory variables and the posttest while controlling for the pretest. Two popular methods control in a different way for the pretest measurement but both methods have serious deficiencies. A Bayesian pretest-posttest multivariate model is proposed which exceeds both methods, and is easily generalized to more complex pretest-posttest settings. The model supports estimating effects of explanatory variables on the pretest and on the posttest at different modeling levels, while controlling for measurement errors at both time points. The multivariate modeling of pretest and posttest scores makes it also possible to compute occasion-specific effects of explanatory variables, while accounting for occasion-specific error distributions. The developed Bayesian estimation method (MCMC) enables the joint estimation of all model parameters including the differential effects of explanatory variables at the pretest and posttest. The multivariate modeling approach is illustrated with a large-scale data-based decision-making intervention to improve teaching quality.

Concurrent paper session 6 10:30 am – 12:00 pm

Session 6A: New Directions in Mediation

A novel effect size measure for the indirect effect in mediation analysis

Mark Lachowicz, Kristopher Preacher, Ken Kelley

Abstract: Researchers in many academic fields have become increasingly aware of the importance of effect sizes when reporting results. However, consensus has not been reached regarding the most appropriate effect size measure for the indirect effect in mediation analysis. Although several measures of effect size exist for mediation analysis, properties of these measures limit their use in many common mediation models. The goal of the current study was to develop an effect size measure for mediation analysis that addressed the limitations of the currently available measures. We show how modification of a currently existing effect size measure results in a novel effect size measure with properties suitable for use in three-variable mediation models. We present results of a Monte Carlo simulation study that was conducted to examine the finite sampling properties of the effect size measure.

Independence Tests to Determine the Direction of Effects in Mediation Models
Wolfgang Wiedermann
Abstract: In non-experimental studies, the assumption of causal ordering of predictor (X), mediator (M), and outcome (Y) must be addressed by a priori theoretical considerations. However, alternative theories (e.g., X→M→Y versus Y→M→X) may exist. Because exploratory approaches (testing various models) do not provide information on which to base a judgment regarding the plausibility of a tentative model, the paper introduces significance tests for deciding on the directionality of competing mediation models in observational studies. Results of a Monte-Carlo simulation of the performance of the tests under various data scenarios are presented.

Is mediation still evolving? A review of recent advances in understanding and uncovering indirect effects
Emil Coman, Victor Villagra, Maria Coman, Suzanne L. Suggs
Abstract: We provide a review of several recent methodological innovations in mediation understanding and applied modeling, with an emphasis on intuitive & graphical depiction of the issues. We provide a brief historical overview of the concept, and then capture novel contributions gathered from recent modeling advances, including: graphical displays of the indirect, direct, and total effects, visualizing potential outcomes (POs), expanding the mediation model to depict missingness mechanisms and POs, and several relevant contributions from Pearl's causal modeling.

Session 6B: Bayesian Item Response Theory

Bayesian Analysis of Joint Modeling of Response Times with Dynamic Latent Ability
Abhisek Saha, Xiaojing Wang, Dipak K. Dey
Abstract: In educational testing, inferences on ability have been mainly based on item responses while the time taken to complete an item are often ignored. With the advent of computerized testing, information on response time of each item can be obtained without additional cost. To better infer latent ability, a new class of state space models, conjointly modeling response time with time series of dichotomous responses, is put forward. The proposed models can entertain longitudinal observations at individually-varying and irregularly-spaced time points and can accommodate changes in ability and other complications, such as local dependence and randomized item difficulty. Simulations for the proposed models demonstrate that the biases of ability estimation are reduced and their precisions are improved. In applying the models to a large collection of reading test data from MetaMetrics company, we further investigated two competitive relationship in modeling response times with the distance of ability and item difficulty (i.e.monotone or inverted U-shape). The model comparison result supports the inverted U-shape relationship better captures examinees behaviors and psychology in exams.

Bayesian Multilevel Modeling with Small Cluster Sizes
Jiaqi Zhang
Abstract: Multilevel modeling (MLM), which allows appropriate analysis of nested or clustered data, has gained wide acceptance over decades in many fields, including education. Several simulation studies were conducted to examine the effect of small sample sizes at different levels. Results indicate that for small samples, parameter estimates are usually biased, especially at the higher level. Few studies of Bayesian MLM focus on the impact of small cluster size. In this present study, we examine the consequences of the sparseness of level-2 units on the estimation of fixed and random effects in terms of model convergence and both point and interval estimation from Bayesian perspective. We obviate the normality assumption by specifying a prior distribution for unknown parameters, integrate unknowns to obtain a marginal posterior distributions, and compare the Bayesian results with the frequentist perspective. Our results show that the Bayesian approaches outperformed the MLE-based estimates, though under extreme conditions (high ICC + low level-2 cluster size) both estimation approaches are biased. Compared with frequentist perspective, our results suggest that researchers who analyze multilevel data should be encouraged to work from the Bayesian perspective.

Session 6C: Symposium- Recent developments in latent variable modeling
Paper 1: Evaluating the Performance of Latent Class Modeling vs. K-Means and Extended K-Means Clustering by *Jay Magidson*

Paper 2: Assigning New Cases to the Appropriate Latent Classes Using 1-step Scoring in Latent GOLD 5.1 by *Jay Magidson, Jeroen Vermunt*
Paper 3: Prior Sensitivity Analysis in Default Bayesian Structural Equation Modeling by *Sara van Erp, Joris Mulder and Daniel Oberski*

This symposium is organized around recent developments in the field of latent variable modeling aiming to introduce some key methodological innovations in latent class modeling (LCM) and Bayesian structural equation modeling (BSEM). The first three talks focus on LCM. The first talk evaluates the performance of LCM as compared to K-Means and Extended K-means clustering, providing insights into the reasons that LCM would be expected to outperform K-means. The second talk discusses the importance of using local fit measures in latent Markov Modeling. The third talk evaluates the different stepwise-step LCM approaches introduced in recent years and proposes a new two-step estimator that can address some of the known problems with the currently used estimators. The last talk investigates the sensitivity of the solutions of BSEM to different default prior distributions by comparing the performance of vague proper priors to improper priors and empirical Bayes priors, which have not yet been investigated in the context of BSEM.

Session 6D: Issues in Multilevel Modeling

What You Don't Know Can Hurt You: Adjusting for Endogeneity in Multilevel Modeling
Lynn Foster-Johnson, Jeffrey D. Kromrey
Abstract: The performance of three modeling approaches, fixed-effects, random-effect, and Mundalk's within-between (WB) in controlling the effects of omitted variable bias in multilevel data is examined. Overall, the WB approach (Mundlak, 1978) is the superior to either the fixed- or random-effects approaches. Over most conditions, the WB approach resulted in lower RMSE values and tighter confidence interval bands. Further, the data suggest that for large samples, the random effect model produces the best outcomes. For small samples, the random effects models were problematic and fixed-effects methods provided better estimates. Precision was lower and bias larger in conditions where regressor reliability was low, inter-regressor correlation high, and intra-class correlations large. Smaller group- and sample-sizes were linked with greater bias and less accuracy. Over all conditions, as endogeneity increased, results became less precise and more biased. The WB approach does the best job in controlling bias under conditions of high endogeneity.

The Effect of Multicollinearity on a Multilevel Structural Regression Model
Seda Can, Oya Somer
Abstract: Because variables may be highly correlated in the social and behavioral sciences, multicollinearity might be problematic. This study investigates the effect of multicollinearity manipulated in within and between levels of a two-level structural regression model by using Monte Carlo simulation. Furthermore, the influence of the size of the intraclass correlation (ICC) on the model is investigated. The variables of interest are rate of convergence, inadmissible solutions and the relative parameter and standard error bias on the within and between level parameters of the model. The results show that nonconvergent and inadmissible solutions are obtained when there is between level multicollinearity and the ICC is medium (0.05). Especially the rates of inadmissible solutions increase as the multicollinearity increases. The results show no effect of the experimental conditions on the fit statistics and indices of the model. The within part of the model poses no problems. When the multicollinearity is manipulated in the within level, the measurement model is distorted in the medium ICC, but no bias is found in the high ICC condition. In the 0.85 and 0.90 between level multicollinearty conditions, one of the between level regression coefficients is biased in the medium ICC condition. The results of the standard errors of the parameters are similar to those obtained for parameter estimates in the within level multicollinearity conditions, showing the distortion of the measurement model. Finally, when the between group correlations are 0.80 and 0.85, the standard error estimates of the coefficients and the variance of endogenous latent variable are biased only for the medium ICC condition; however, when the correlation is 0.90, these standard errors are biased for both two ICC levels.

Data Sparseness and Two-Level Logistic Multilevel Models: Impact of Cluster Size on Fixed and Random Effects
Bethany A. Bell, Jason A. Schoeneberger, Zhaoxia Guo, Jeffrey D. Kromrey, John M. Ferron
Abstract: Methodological research continues to address the important issue of sample size at each level when estimating multilevel models. However, to date, these investigations have primarily focused on multilevel linear models, with normally distributed continuous outcomes. Fewer studies have examined the impact that sample sizes have on the

statistical properties of multilevel models with binary outcomes. Moreover, even less attention has been given to the consequences of estimating multilevel logistic models with sparse level-2 units. To help address this gap in the literature, this Monte Carlo study focuses on the consequences of level-2 sparseness on the estimation of fixed and random effects in terms of point and interval estimates as a function of the level-1 sample size, number of level-2 units, proportion of singletons (level-2 units with one observation), collinearity, outcome prevalence, intraclass correlation, and model complexity. SAS IML was used to simulate 1000 data sets across 900 conditions. Results are presented in terms of statistical bias, confidence interval coverage, Type I error control, and statistical power.

Session 6E: Mixture Modeling Applications

Does hearing fluctuate over time in Meniere's Disease? A growth mixture modeling approach for categorizing patterns of change over time.

Laurel Fisher, Michael Hoa

Abstract: The Purpose of the study is to determine underlying sub-groups in changes in hearing in Meniere's Disease patients. Meniere's disease consists of a constellation of 4 symptoms: episodic vertigo, hearing loss, tinnitus, and aural fullness. Typically, the symptoms show a sudden onset and affect one side or ear (unilateral presentation), but may affect both ears (bilateral presentation). Cochlear hydrops is considered to be a variant within Meniere's, in which patients experience hearing loss, tinnitus, and aural fullness, but lack episodic vertigo. Hearing data are multidimensional, with 7 sound frequency thresholds and a word recognition score per ear evaluated over time. The full set of hearing data will be used to test hypotheses that the disease affects part of the cochlea (supporting the histopathological findings) over time or that it affects the entire cochlea. Mixture modeling techniques allow for examination of change by all frequencies, word recognition score, ear, and evaluation, to disambiguate both audiometric status and change over time in Meniere's disease. Audiological data prospectively gathered from patients with a diagnosis of Meniere's disease (n=488) were retrospectively analyzed. The study was IRB-approved and consisted of 228 males and 259 females, with an average age of 51.2 years at presentation. Most of the sample were diagnosed with unilateral disease (n=344) (AAO-HNS criteria), 47 with bilateral disease, and 97 patients with cochlear hydrops. The data were collected in two waves and the second wave had at most 3 repeated evaluations, whereas the first wave had up to 87 separate hearing tests per patient (total 3,340 hearing tests). Significant hearing fluctuation from one hearing test to another was defined at the 4-frequency PTA, speech discrimination score, and at individual frequency thresholds. Growth mixture model analyses using Mplus 7.4 modeled affected ear audiometric threshold data and covariates. The first wave dataset was used to determine hearing change classes. The second wave dataset was used in an attempt to validate the hearing change classes. The analysis indicated that a 4-class model fit the auditory measures across evaluation reasonably well. The latent classes could be described as "low frequency changer", "stable", "all frequency changer", and "speech measure only changer". The 4-class model was only moderately validated in the second wave dataset. Patients clustered into clear categories of changes in hearing, potentially indicating that different types of treatment would be best suited for a patient. Hearing data and change in hearing can be used to categorize patients into mild through severe disease, thus, providing better targets for treatment studies.

Growth Mixture Modeling in Self-Rated Health and Depression

Joseph W. Jones, Thomas Ledermann, Ginger Lockhart

Abstract: Self-Rated Health (SRH) and Depression are versatilely used in Health Psychology. Both identify as risk factors and show great ability at predicting health outcomes of disease and mortality. SRH and Depression are also both highly associated with one another. However, only a few studies have assessed individuals' developmental trajectories and similarities of these two variables over time. Using Growth Mixture Modeling, we sought to observe developmental trajectories of SRH and Depression and identify latent groups differing in their levels and trajectories. Data were obtained from the first six waves of the Australian Longitudinal Study of Aging, which consisted of 2,087 (1,031 Females) subjects aged 65 years and older. Results showed that both SRH and Depression developed three different levels and longitudinal trajectories. All Depression trajectories were positive and significant, while SRH had one latent group with improvement of SRH over time. Analysis of the distribution of Latent Group assignment indicated off-diagonal membership. For example, individuals with higher scores of SRH also showed increased scores of Depression. Given the off-diagonal membership observed, SRH and Depression may not be as strongly connected as once thought and there may exist possible protective factors for one variable and not the other.

Modelling menstrual cycle length and variability at the approach of menopause by using hierarchical change point models

Michael Elliott, Xiaobi Huang, Sioban Harlow

Abstract: As women approach menopause, the patterns of their menstrual cycle lengths change. Our proposed hierarchical model incorporates separate mean and variance change points for each woman and a hierarchical model to link them together, along with regression components to include predictors of menopausal onset such as age at menarche and parity. Additional complexity arises from the fact that the calendar data have substantial missingness due to hormone use, surgery and failure to report. We integrate multiple imputation and time-to-event modelling in a Bayesian estimation framework to deal with different forms of the missingness. Our method successfully models patterns of women's menstrual cycle trajectories throughout their late reproductive life and identifies change points for mean and variability of segment length, providing insight into the menopausal process. More generally, our model points the way towards increasing use of joint mean–variance models to predict health outcomes and to understand disease processes better.

Session 6F: Innovations in large data techniques

Evaluating Monte Carlo Simulation Through Data Mining Approaches

Oscar Gonzalez

Abstract: A lot of the research used to evaluate the properties of statistical models involve Monte Carlo simulation methods. After obtaining the outcomes from a Monte Carlo simulation, researchers investigate if the factors that varied across simulation conditions (sample size, effect size of parameters, etc.) influence power, bias, coverage or Type 1 errors in parameter estimates. Monte Carlo results are either described in tables, plots, or the outcomes are used as dependent variables in ANOVA's or logistic regression with simulation conditions as predictor variables. A potential drawback to using the linear models is that the factors might lead to higher-level interactions that are very difficult to probe and conceptualize in the previous approaches. This paper proposes to use a Classification and Regression Tree approach (CART) to describe and analyze Monte Carlo simulation data. A description of the CART approach and code in the R statistical environment to carry out the CART approach out are presented. The previously four approaches would be compared by answering the same question with simulated data from a statistical mediation model. Potential benefits and drawbacks from each of the four approaches are discussed.

Functional Parallel Factor Analysis for Functions of One- and Two-dimensional Arguments

Ji Yeh Choi

Abstract: Parallel factor analysis (PARAFAC) is a useful multivariate method for decomposing three-way data that consist of three different types of entities simultaneously. This method estimates trilinear components, each of which is a low-dimensional representation of a set of entities, often called a mode, to explain the maximum variance of the data. Functional PARAFAC permits the entities in different modes to be smooth functions or curves, varying over a continuum, rather than a collection of unconnected responses. The existing functional PARAFAC methods handle functions of a one-dimensional argument (e.g., time) only. In this paper, we propose a new extension of functional PARAFAC for handling three-way data whose responses are sequenced along both a two-dimensional domain (e.g., a plane with x- and y-axis coordinates) and a one-dimensional argument. Technically, the proposed method combines PARAFAC with basis function expansion approximations, using a set of piecewise quadratic finite element basis functions for estimating two-dimensional smooth functions and a set of one-dimensional basis functions for estimating one-dimensional smooth functions. We apply the method to simulated and real data to demonstrate its empirical usefulness.

Asymptotic efficiency of the pseudo-maximum likelihood estimator in multi-group factor models with pooled data

Fan Yang Wallentin

Abstract: A multi-group factor model is suitable for data originating from different strata. However, it often requires a relatively large sample size to avoid numerical issues such as nonconvergence and non-positive definite covariance matrices. An alternative is to pool data from different groups in which a single-group factor model is fitted to the pooled data using maximum likelihood. In this paper, properties of pseudo-maximum likelihood (PML) estimators for pooled data are studied. The pooled data are assumed to be normally distributed from a single group. The resulting asymptotic efficiency of the PML estimators of factor loadings is compared with that of the multi-group maximum likelihood

estimators. The effect of pooling is investigated through a two-group factor model. The variances of factor loadings for the pooled data are underestimated under the normal theory when error variances in the smaller group are larger. Underestimation is due to dependence between the pooled factors and pooled error terms. Small-sample properties of the PML estimators are also investigated using a Monte Carlo study.

Concurrent paper session 7 **2:45 pm – 4:15 pm**

Session 7A: Modeling Innovations in Social Science

Structural Equation Modeling in Archival Accounting Research: An Application to Disclosure and Cost of Capital
Steven Utke, Lisa Hinson
Abstract: Structural equation modeling (SEM), a methodology currently underutilized by archival accounting researchers, enables examination of paths linking constructs in complex models. In this paper, we first discuss SEM, including basics, underlying assumptions, advantages and disadvantages, properties of the estimates, and best practices. We then apply SEM to an accounting topic by using common factor analysis to create latent variables (constructs or factors) for earnings quality, voluntary disclosure, information asymmetry, and cost of equity capital and then examining the paths among those constructs. We find that higher earnings and voluntary disclosure quality are negatively associated with cost of capital both directly and indirectly through information asymmetry. The paths from voluntary disclosure quality to cost of capital are significant even after controlling for the paths from earnings quality to cost of capital. SEM allows us to perform our analysis, which would not be possible with other techniques, and potentially offers fruitful avenues for future research in accounting.

Illustration of the novel Mplus' SKEWT option for testing for randomized intervention effects when outcomes are skewed. How much did the eConsult trial reduced costs of care in treated patients?
Emil Coman, Victor Villagra, Daren Anderson, Ianita Zlateva
Abstract: We illustrate a novel feature implemented recently in Mplus for modeling outcomes that are skewed and 'fat-tailed' in their distribution. The data comes from eConsults, a randomized trial testing the efficacy of alternative electronic (non-face-to-face FTF) consultations between a PCP and a cardiology specialist, which utilizes secure messaging to exchange information. Mplus SKEWT models provide 2 additional estimated parameters in its output for each outcome (besides the mean & variance): skew and DF, all tested for significance. We provide details about the models, emphasizing the benefits of 2-group (by Tx condition) models over the 1-group Y-on-Tx regression type model. All models besides the SKEWT failed to find a significant effect of treatment on costs; the SKEWT option found 'average' savings of $627 ($233), p = .007 (estimated change in skewt-costs of -$4,535 vs. -$3,908).

Models for detecting nuanced health disparities (HD). Assessing the size of HDs and when and how they are reduced or not
Emil Coman, Victor Villagra, Daren Anderson, Ianita Zlateva
Abstract: We present a modeling method of investigating Health Disparity (HD) effects that moves beyond assessing the size of raw differences, like differences in means, to reveal differences in causal mechanisms between racial/ethnic (R/E) groups. We use as illustration data gathered as part of a randomized intervention to reduce unnecessary medical utilization and hence costs, eConsult, a randomized trial testing the efficacy of alternative electronic (non-face-to-face FTF) consultations between a PCP and a cardiology specialist. Every parameter in multiple R/E groups structural models can be directly compared against its mirror counterpart from the other group, and their difference tested for significance. For a 4 group (Tx-by-R/E) model, a richer picture emerges: the intercept of the outcome cost, for example, will have four estimates, two for W (White&English): Wo (controls) and W1 (treated), and two other for N (Non-White or Non-English), No (controls) and N1 (treated). We exemplify these options with a 4-group model with an 'All costs' changes outcome, regressed on its baseline variant, then move into detailing mediation-type models with Δ(Cardio costs)->Δ(All-but-Cardio) costs run in the 4 Tx-by-HD groups, paint an increasingly complex picture.

Session 7B: Tutorial: Conducting propensity score analysis with structural equation models using MPLUS

Walter Leite, Laura Stapleton

Abstract: The objective of this presentation is to provide a tutorial on combining propensity score analysis (PSA) with structural equation modeling (SEM) using the MPLUS software. There are three contributions that SEM can provide to PSA: SEM can be used to include latent variables in the propensity score model, removing adverse effects of measurement error in the estimation of propensity scores. Second, in the analysis stage, SEM can be used to estimate treatment effects on latent variables, adjusting for unreliability of measurement. Third, SEM can provide validity evidence for the measurement of latent variables that are hypothesized to be affected by the intervention of interest. This tutorial will follow the steps of propensity score analysis in the context of estimating the effect of new teachers being assigned a mentor on their perception of workload manageability, using data from the School and Staffing Survey.

Session 7C: Issues in Multilevel Modeling

Efficient estimation of variance components in nonparametric mixed-effects models with large samples

Nathaniel E. Helwig

Abstract: Linear mixed-effects (LME) regression models are a popular approach for analyzing correlated data. Nonparametric extensions of the LME regression model have been proposed, but the heavy computational cost makes these extensions impractical for analyzing large samples. In particular, simultaneous estimation of the variance components and smoothing parameters poses a computational challenge when working with large samples. To overcome this computational burden, we propose a two-stage estimation procedure for fitting nonparametric mixed-effects regression models. Our results reveal that, compared to currently popular approaches, our two-stage approach produces more accurate estimates that can be computed in a fraction of the time.

Bootstrapping for Two-Level Synthesis of Single-Case Experimental Data

Mariola Moeyaert

Abstract: This paper presents four different types of bootstrap methods (i.e., the effect size bootstrap, the raw data bootstrap, the error bootstrap and the cases bootstrap) to bias correct the variance components estimates in context of multilevel modeling of standardized regression based effect sizes. A computer intensive simulation study is being conducted to evaluate parameter recovery under the four different bootstrap estimation methods.

The package does matter: A comparison of multilevel modeling in 5 common software packages

D. Betsy McCoach, Graham Rifenbark, Aarti Bellara, Xioran Li, Sarah Newton, Janice Kooken, Anthony Gambino, & Dani Yomtov

Abstract: The purpose of this study is to guide applied and methodological researchers in selecting multilevel software that provides the best match for the requirements of their research. The different software choices bring a wide range of functionality and flexibility. Little is known to help researchers identify which software has features that best align with the research question and optimizes the accuracy and efficiency of the modeling step. This paper reports the results of a simulation study comparing HLM, Mplus, SAS, R, and Stata, in the recovery of fixed and random effects when varying the size of random effects from small to moderate. Combining these results with a comparison of model run times, estimation techniques, output formats and program flexibility, we will provide a guide to the selection of multilevel software.

Session 7D: Effect Sizes and Hypothesis Tests

Bayes Factor null hypothesis tests are still null hypothesis tests

Matt N. Williams

Abstract: Many researchers are aware that null hypothesis significance tests (NHST) have significant problems. One alternative strategy for simple research questions is the use of Bayes Factor null hypothesis tests. I will argue that while Bayes Factor null hypothesis tests are an improvement over NHST, they are unsuitable as a default analytic strategy in the social sciences. Specifically, Bayes Factor null hypothesis tests are only useful for producing conclusions about the posterior probability of hypotheses if our prior knowledge takes a very specific form: a "spike and slab" distribution, with a substantial (ideally 50%) probability that the parameter is exactly zero, and the rest of the prior density spread over a wide range of values. Such a prior is both highly informative and also rather unlikely to be a reasonable representation of the actual prior knowledge held about a particular effect or relationship in the social sciences. I argue that we need to stop

worrying about testing the null hypothesis, which is almost always false. Instead, Bayesian estimation is a more useful default analytic strategy—perhaps with a particular focus on the probability that a particular parameter is positive or negative. However, to allow Bayesian estimation to become widely used by everyday researchers, we as methodologists need to produce sensible informative default priors, and implement them in easy-to-use software. I will produce some suggestions for how this might be accomplished.

Asymptotic efficiency of the pseudo-maximum likelihood estimator in multi-group factor models with pooled data
Fan Yang Wallentin
Abstract: A multi-group factor model is suitable for data originating from different strata. However, it often requires a relatively large sample size to avoid numerical issues such as nonconvergence and non-positive definite covariance matrices. An alternative is to pool data from different groups in which a single-group factor model is fitted to the pooled data using maximum likelihood. In this paper, properties of pseudo-maximum likelihood (PML) estimators for pooled data are studied. The pooled data are assumed to be normally distributed from a single group. The resulting asymptotic efficiency of the PML estimators of factor loadings is compared with that of the multi-group maximum likelihood estimators. The effect of pooling is investigated through a two-group factor model. The variances of factor loadings for the pooled data are underestimated under the normal theory when error variances in the smaller group are larger. Underestimation is due to dependence between the pooled factors and pooled error terms. Small-sample properties of the PML estimators are also investigated using a Monte Carlo study.

Poster Abstracts

1. A Lens Model Analysis of Individual Nutrition Judgments

Kristina A. Carter

Consumer judgment of the nutritional value of food products is increasingly pertinent as obesity and nutritionally-related illnesses remain at record levels in developed nations. The U.S. 1990 Nutritional Labeling and Education Act standardized a nutrition label but the degree to which consumers use this Nutrition Fact Panel label to judge the nutritional quality of food is warrants further investigation. In two studies, Lens Model Analysis was used to examine what nutrient information influenced participants nutritional judgments and how well participant nutritional judgment corresponded to a gold standard criterion for nutritional quality. The gold standard used in these studies was the product's NuVal® score, an algorithmically computed measure of nutritional quality ranging from 1-100. In the first study, 196 18+ individuals living in the United States rated the nutritional quality of either cereals or snacks on a scale of 0 (not at all nutritious) to 100 (extremely nutritious) based on an image of the front of the product as well as its ingredients list and Nutrition Fact Panel. In a follow-up study, 300 participants recruited through Ohio University's psychology participant pool viewed and rated 74 cereals. In the first condition, participants viewed the NFP, ingredients, and cereal box as in the previous study but two additional conditions were added in which participants viewed a Front of Package (FOP) label highlighting nutrients either strongly or marginally related to overall nutrition. In both studies, differences in judgment accuracy according to demographic differences such as gender, age, race, BMI, educational level, health status, dieting status, nutritional knowledge, health beliefs and behaviors, and numeric literacy were also examined. Variation in accuracy of nutritional judgment and consistency in cue usage according to individual, domain, and label format differences are reported.

2. Impact of South Carolina's TANF Program on Earnings of New Entrants Before and During the Recent Economic Recession

Qiduan Liu, Cynthia Flynn

This study employs multiple-group, piecewise, latent curve models to assess the impact of South Carolina's Temporary Assistance for Needy Families (TANF) program on the longitudinal earnings of three cohorts of approved applicants that entered the study before the recent recession, at the beginning of the recession, and at the height of the recession. Applicants screened out by intake staff as ineligible and then statistically matched to approved applicants by family characteristics and local economic conditions served as the comparison group. The findings show the state's TANF work support program had a positive impact on the earning trajectories of TANF clients when the economy was healthy. The effect became weaker during the state's period of rising unemployment, and disappeared at the height of recession.

3. Understand the International Difference on Math Achievement from a Cognitive Perspective

Xiaoran Li

Comparing the educational performance of the United States with other countries has been studied widely; however, limited research focuses on cognitive ability. This study employs a cognitive diagnostic model using TIMSS 2011 database in order to better understand the math achievement differences reflected in cognitive ability between countries.

4. Researching Administrative Records Usage in the Nonresponse Followup Operation of the 2020 Census

Andrew D. Keller, Scott M. Konicki

The Census Bureau is researching how to use administrative record information from government and other sources in place of field visits during the Nonresponse Follow-up (NRFU) operation. This poster shows an approach for identifying vacant and occupied housing units to be enumerated using administrative records during the NRFU operation. While the approach allows flexibility in balancing cost and quality, we show results that were utilized in determining the 2020 Census Operational plan.

5. Confirmatory Factor Analysis of the Spanish version of CAPES Intensity in Mplus and R: Illustration of the method for estimating the significance of chi-square test when using Multiple Imputations with WLSMV estimator

Ania Filus, Anilena Mejia, Rachel Calam, Alina Morawska

Confirmatory Factor Analysis is a valuable procedure for examining factor structure of a new language version of an established measure. Mplus diverges from most other SEM software packages in its ability to fit latent variable models to data files that contain ordinal or dichotomous outcome variables. The default estimator for models that contain categorical outcomes is the WLSMV estimator (Muthén et al., 1997). If missing data are present on categorical outcomes, Multiple Imputations (MIs) provide efficient method to handle missingness, if data are at least MAR (Enders, 2010). Under WLSMV with MIs in Mplus, the WLSMV $\chi 2$ and the fit indices are averaged across imputations; yet the p value for the $\chi 2$ is not provided. This paper illustrates the method for estimating the p value of chi-square test pooled across multiple imputations by combining modeling in Mplus and R. As an example we describe the evaluation of the factor structure of the Spanish version of CAPES Intensity Scale (Morawska et al., 2014) among 190 parents from Panama. The CAPES is a 27-item measure of child emotional and behavioral problems with each indicator evaluated on the 4-point Likert type scale. We used CFA in Mplus v. 7.3 with WLSMV estimator and MIs (using 50 imputations) to evaluate the factor structure of CAPES. The fit indices were averaged over imputed data sets and the $\chi 2$ test statistics from each imputation were combined to yield an F statistic and a p value for the chi-square test using a method developed by Li and colleagues (1991) that was utilized in miPoolChi macro in semTools package in R. The CFA supported a two-factor structure of CAPES with two scales Behavioral and Emotional Problems and Child's Competencies.

6. Evaluation of Fit Statistics for Longitudinal Factorial Invariance Studies in Structural Equation Modeling

Ayotunde Akinleye

This study examines the performance of a modified CFI, the chi-square difference test and other fit indexes for assessing longitudinal measurement invariance in SEM. Establishment of measurement invariance (across time, and across groups) is a required condition for meaningful interpretation of assessment tools, and theoretical models in longitudinal research, and many researchers in social science disciplines have adopted the structural equation modeling methodology as a viable tool for addressing substantive questions of change, and growth that require them to assess longitudinal measurement invariance. The process of testing measurement invariance under the SEM framework entails a series of model comparisons whereby the fit of a hypothesized model (typically assumed to be noninvariant, or less invariant) is compared to a nested model (representing the invariance condition) in which all (or a subset) of a class of parameters (i.e. loading pattern, factor loadings, factor intercepts, unique factor) are constrained to be equal across the levels of a particular dimension of interest (e.g. time for longitudinal measurement invariance, and/or group). Measurement invariance hypotheses in SEM can be evaluated with the chi-square difference test due to nesting relationships of the models compared, however alternative fit indices are required (i.e. in addition to the chi-square difference tests) for evaluating measurement invariance hypothesis because the chi-square difference test is known to be influenced by the model fitting sample size, and by the size of the estimated model. The CFI is one of the most commonly used alternative fit indices in applied research, and it is particularly useful for tests of measurement invariance because it is an incremental fit index which is defined in terms of the proportional reduction in the noncentrality parameter of a substantive model of interest over that of a baseline model that is nested within it. Many research studies have been conducted to evaluate the performance of fit indices for evaluating model fit in SEM (Hu & Bentler, 1998, 1999; Fan & Sivo, 2005) including their performance for evaluating growth curve models (Wu & West, 2010; Leite & Stapleton, 2011), and measurement invariance across groups (Lai & Yoon, 2015), however to our knowledge, no studies to date have evaluated the appropriateness of Hu and Bentler's (1998) guidelines for evaluating longitudinal measurement invariance within the SEM framework. Although the procedures for conducting tests longitudinal invariance are identical to those used for testing measurement invariance in cross-sectional research, Widaman & Thompson (2003) have identified a need for redefining the baseline models that are used to derive incremental fit indices in longitudinal SEM analysis including. They note that the standard independence null model used as the default model for computing incremental fit index values in many SEM software packages is not nested within configural invariance hypothesis model and within many substantive models that are of interest to applied researchers. Accordingly Widaman & Thompson (2003) consider an "intercept only" model to be an acceptable baseline model. Likewise, Lai & Yoon (2015) suggest that specification of an appropriate null model for longitudinal measurement invariance hypotheses would take dependencies that result from repeated measurements on the same individuals into account. We conduct a simulation study to examine the specification of an appropriate baseline model for longitudinal measurement invariance and evaluate the effect of this modification on the performance of the

CFI, and other fit statistics (the chi-square difference test and alternative fit indices) for conducting tests of longitudinal measurement invariance.

7. Validation of the Genderism and Transphobia Scale and the Collective Self-Esteem Scale in a Mainland Chinese Sample

Bing Chen

Social Identity Theory (SIT) can be used to explain prejudiced attitudes and discriminatory behaviors by moving away from an individual or interpersonal approach to a focus on social groups. Social identity consists of one's self-concept which derives from his or her membership in one or more groups (e.g., gender and ethnicity). The terms collective identity and collective self-esteem are similar to social identity. Gender self-esteem is one type of collective self-esteem and refers to the importance of one's gender to one's self-concept. Some studies that used SIT to help understand transprejudice (prejudice against transgender people) in samples from Western countries have found that stronger gender self-esteem is associated with more transprejudice in Western countries, or individualist societies. People in collectivistic societies may be more likely than people in individualistic societies to use their social positions, such as social status or roles, as a way to distinguish themselves from others. However, in determining verbal or physical aggression, people's ingroup/outgroup status is more important than whether they come from an individualistic or a collectivistic society. Ingroup/outgroup status may be important in both collectivistic and individualistic societies for predicting aggression. Gender self-esteem as a predictor of transprejudice has not been studied in a collectivistic society. Is gender self-esteem, which may contribute to transprejudice in individualistic societies, also a significant contributor to transprejudice in mainland China, a collectivistic society? Because measures to assess transprejudice - Genderism and Transphobia Scale (GTS; Hill & Willoughby, 2005) and gender self-esteem - Collective Self-Esteem Scale (CSE; Luhtanen & Crocker, 1992) were developed in Western countries and have never been used in a sample from mainland China, this study validates the two measures in a Chinese sample of college students. The data were gathered from an online survey, and participants were a snowball sample consisting of 148 college students from mainland China. EFA using principal axis factoring with promax rotation was used to shorten the length of the scales, explore new measure structures, and make the tested models smaller. Items with loadings greater than or equal to 0.5 but less than 0.9 were retained. The shortened CSE had two factors. Factor 1 consisted of 3 items and measured one's positive judgments of one's gender group. Factor 2 consisted of 4 items and measured the degree to which one reports he or she is a worthy member of his or her gender group. The shortened GTS included two factors. Factor 1 consisted of 3 items and measured violence towards transpeople. Factor 2 consisted of 6 items and measured general negative attitudes toward transpeople. CFA was conducted using EQS. The models tested for both shortened CSE and GTS included a single-factor model, a two-factor uncorrelated model, and a two-factor correlated model. Fit indices (small chi-square to degrees of freedom ratio, CFI values above .90, and RMSEA values below .10) were used to establish the best-fitting model. Results showed that for CSE, the two-factor correlated model fitted the data best. However, for GTS, none of the three models fitted the data very well, though results from nested comparisons showed that the two-factor correlated model fit better than other two models.

8. The relationship between grieving process and depression: A cross-lagged analysis

Bo YAN, Amy CHOW Y.M.

This panel study applied a cross-lagged analysis to explore the relationship between grieving process and depression among Hong Kong Chinese bereavement population. 106 participants (30 male, 76 female, average age 49) were recruited from a local hospice and a hospital-based palliative care unit. Two waves of data in the panel study were collected around one month (Time 1, T1) and four months (Time 2, T2) after participants' family member death (spouse, children, sibling, parents, or other). Guided by the Dual Process Model (DPM), we conceptualized grieving process as Loss Orientation (LO) and Restoration Orientation (RO). The reciprocal relationship among LO, RO, and Depression were examined in the modified cross-lagged model. Overall, the fit index shows that the model adequately fits the data. The findings suggest that LO at T1 and Depression at T1 predict LO at T2. Meanwhile, RO at T1 and Depression at T1 predict Depression at T2. Interestingly, RO at T2 is not predicted by any of T1 key variables, LO, RO, or Depression. Implications for further research and practice issues are discussed.

9. Estimating reliability in intensive longitudinal data with ordinal outcomes

Cara Arizmendi, Laura Castro-Schilo

With an increased focus on collecting psychological data with more frequently repeated measures, creating measures intended explicitly for intensive longitudinal data (ILD) or adapting measures intended for panel data is needed (Shrout

and Lane, 2012). Current methods of intensive longitudinal studies typically involve using measures created for panel data, not daily diaries; however, we do not always know if these measures have enough sensitivity to demonstrate within person change. In order to determine if a given measure can reliably detect change in individuals, methodologists have turned to estimating the reliability of change. However, differing methods utilize differing ideas of what change means, and the term may be a misnomer in some cases. After reviewing potential methodological issues specific to ILD, I briefly review current methods of estimating the reliability of change, provide empirical examples of one of the multilevel methods, and extend the method by treating the outcome as ordinal instead of continuous.

10. Fitting the growth curve model to changes in perceived love in the Bayesian framework

Chelsea Muth, Zita Oravecz

The growth curve model (GCM) is one of the most frequently used classes of longitudinal models in psychological research. GCMs, also referred to as latent trajectory or latent curve models, allow us to model individual trajectories over time, and to compare these trajectories across individuals and groups. The GCM is flexible in handling varying growth functions (linear, curvilinear, nonlinear), as well as unbalanced designs, meaning, study participants may be measured at different occasions and do not have to be excluded from analysis if some of their measurements are missing. We propose using the Bayesian statistical framework for fitting growth curve models. Bayesian methods provide flexible tools to fit growth curve models with various levels of complexity. For example, several studies have also demonstrated the robust nature of Bayesian GCMs to handle nonnormal residual variance without adding error to inference. Our poster introduces a Bayesian growth curve model, provides a detailed description and example with computer code for fitting it, and gives guidelines on how to incorporate further extensions for particular research questions. We start by explaining and comparing the classical and Bayesian formulations of the growth curve model. Next we provide a step-by-step guide on how to analyze a sample dataset – longitudinal measures of perceived marital love across the transition to parenthood – with a Bayesian GCM. Then we consider several extensions that can easily be incorporated to extend the introduced GCM. Lastly, we recap the potential applications and benefits of using this modeling framework. Keywords: longitudinal modeling, growth curve model, Bayesian statistical framework

11. CI of Fit Indices through the Inversion of a Bootstrapping Sampling Distribution

Chuchu Cheng

In this paper we propose a new method of constructing confidence intervals (CIs) of fit indices in SEM. The most recent development in this area (Zhang & Savalei, 2015) uses Yuan, Hayashi and Yanagihara (2007, YHY) bootstrap and has been shown to perform better than approaches based on traditional naïve bootstrap and Bollen-Stine bootstrap. However, based on the quantiles of a bootstrapping sampling distribution at a single fixed level of misspecification, this method is similar to a Wald type CI. Because a profile-likelihood-based approach to CI is in most cases superior to a Wald type CI, we construct CIs by inverting a bootstrapping sampling distribution of a statistic and expect our method to perform better than the previous approaches.

12. Examining the Effects of Including Differently Correlated Auxiliary Variables in Multiple Imputation Models

Jiaqi Zhang

Multiple imputation (MI) is one modern methods for handling missing data when missing is assumed missing at random (MAR). The missing data literature suggests the efficiency f MI can be improved with the inclusion of auxiliary variables (AV), which are selected to aid the imputation process, and are unrelated to the analysis. However, a review of the AV literature found inconsistencies among the guidelines for including the different types of AVs in the MI analysis, although this review included simulations of varying complexity. In this simulation study, we examined the effects of including these different types of potential auxiliary variables. Unlike other similar study, we examined the improvement in MI performance through root-mean-square-error, bias and standardized bias, and the standard error in mean and coefficient estimates by including different types of auxiliary variables with varying correlations between missingness indicators and the outcome within the MI model. We examined these effects along four primary dimensions: correlation strength (strong, moderate, low), missingness mechanism (MAR vs. MNAR), proportion of missingness (15%, 30%, 50%), and normally and non-normally distributed data. Results suggest that including auxiliary variables can improve the MI model under both MAR and MNAR assumptions, however, the contribution of single AV is trivial when other AVs are included in the MI model. Strongly correlated auxiliary variables play important roles, including auxiliary variables with low correlations is not useful. The performance of AVs under non-normal distribution was worse than with normally distributed data.

13. An Exploration of Two Artifact Correction Methods for Correcting Correlated Reliabilities in Meta-analysis
Lei Zhao
In current meta-analysis practice, artifact corrections for sampling error and measurement error have been well recognized and popularly applied (Hunter & Schmidt, 2004; Aguinis, Dalton, Bosco, Pierce & Dalton, 2011; Murphy, 2003). As information accessibility for sample size and reliability estimates has improved across social science studies, correcting sampling errors and measurement errors has become a well-accepted convention for most meta-analyses (Murphy, 2003). As a result, an appropriate use of reliabilities for correcting measurement errors in meta-analysis has been attracting more attention. Hunter, Schmidt and Le (2006) explicitly addressed the consequences of using inappropriate reliability estimates for meta-analysis studies. Yet discussions exclusively around the incomplete corrections for measurement error due to the use of mixed types of reliabilities across sample studies are still limited (Hunter & Schmidt, 2004, chapter 3). In addition, the practical implications for the assumption of reliability independency require further investigation (Kohler Cortina Kurtessis & Golz, 2015). The formulae for correcting measurement errors in current meta-analysis practice assume independence between the reliability of independent variable and the reliability of dependent variable (Hunter & Schmidt, 2004). Kohler et al. (2015) tested the tenability of this assumption, through an empirical examination of the correlation between reliabilities of two variables across individual studies: they found substantial correlations between the reliabilities of the perceived organizational support (POS) and the reliabilities of its outcome variables, such as organizational citizenship behaviors and job performance. This relationship varies between different types of reliability estimates for independent and dependent variables. This interdependent relationship between reliability estimates of the two variables would potentially result in biased corrections in meta-analysis. The individual correction procedures developed by Raju, Burke, Normand, and Langlois in 1991 (RBNL), produce meta-analytic estimates for a bivariate correlation between two true scores. RBNL was originally developed to specifically hold account of sampling errors of the artifacts themselves, considering the available artifact values across studies always suffer from sampling errors. Strengths of this method are demonstrated in improving the accuracy of estimation when compared to other methods (Raju et al., 1991). Pragmatically speaking, this method requires fewer assumptions and addresses various scenarios of missing values for artifacts. Most importantly, RBNL is stated as a suitable artifact correction method regardless the types of reliabilities. "Instead of offering a separate sampling variance formula for each definition of reliability, this development propose a single formula for the sampling variance of corrected corrections, which may be used with any definition of reliability" (Raju & Brand, 2003). However, RBNL has not been applied as popular as Hunter & Schmidt method (1991, 2004). Only 3.1% meta-analytic studies from 1982 to 2009 adopted RBNL out of more than five thousands of correlation effect sizes that were reviewed by Aguinis et al. (2011). We examine the impact of correcting reliability artifacts in meta-analytic estimates with inter-correlated reliabilities. Although a Monte-Carlo study by Raju and his colleagues (Raju, Anselmi Goodman & Thomas, 1998) has theoretically examined this topic, an investigation with a real-world case has not been done. Expressly, through an examination of data from published studies using Hunter-Schmidt's (1990, 2004) and RBNL (Raju et al, 1991; Raju & Brand, 2003) methods, this study will not only provide empirical evidence for the impact of correlated reliabilities on meta-analytic estimates, but also further explore how this impact is contributed by the inter-correlations amongst different types of reliability estimates, and the sensitivity of the two analysis methods to the violation of this artifact independence. Specifically, this proposed study explores the differences in meta-analytic parameter estimations, such as estimates of observed variance, credibility intervals, and confidence intervals, under different degree of reliability interdependency, when using Hunter & Schmidt method and RBNL method. The results of these two methods are compared in terms of variations among the parameter estimations under reliability interdependency.

14. Interaction effects of neighborhood disadvantage and individual social support on frequency of alcohol use in youth living with HIV
Leslie Ann D Brick, Nicole R Nugent, Shoshana Y Kahana, Douglas Bruce, Mary R Tanney, M Isabel Fernandez, Jose A Bauermeister
Neighborhood context is associated with individuals' health-risk behaviors, including substance use. To date, no large scale US studies have examined neighborhood and individual level influences on alcohol use in youth living with HIV (YLH). Although the broader literature on the effects of neighborhood socioeconomic disadvantage has been somewhat inconsistent, we hypothesized that adolescents living in disadvantaged neighborhoods would report more alcohol use and that social support would mitigate the negative impact of residence in socioeconomic disadvantaged neighborhoods. The Adolescent Medicine Trials Network for HIV/AIDS Interventions (ATN) recruited a total of 2225 behaviorally and perinatally infected adolescents and young adults to participate in a cross sectional survey between December 2009 and

June 2010. Participants in the study were linked to care at clinics or units associated with the ATN that were broadly distributed across the continental United States. Individual-level Data. Participant-level demographic information (e.g. birth sex, HIV diagnosis in the past 12 months, race, sexual orientation, age in years, mode of HIV transmission) and social support for alcohol avoidance were assessed. Measures of alcohol and other drug use were determined using the Alcohol, Smoking, and Substance Involvement Screening Test (ASSIST). Neighborhood specific data was compiled from the 2010 U.S. Census Bureau and matched with individual-level data. Given the nested structure of individuals within geographic areas, multilevel modeling was employed to acknowledge that both neighborhood-level and individual-level effects might contribute to variation in frequency of alcohol use. Analyses proceeded via a stepwise examination of a series of four models: Model 1 (null model) to examine the variation in alcohol frequency across neighborhoods; Model 2 (random intercepts, fixed slopes) to examine the effects of all predictors; Model 3 (random intercepts, random slopes) to examine random variation in the relationship between social support and alcohol use across neighborhoods; Model 4 (cross-level interaction) to examine the interaction of social support and neighborhood disadvantage. The sample included 1194 youths living with HIV across 538 neighborhoods in the United States who acquired the disease behaviorally and who endorsed having tried alcohol in their life. Examination of results from the Model 1 demonstrated significant variation in frequency of past three-month alcohol use across neighborhoods (0.08 [95% Confidence Interval (CI): 0.01, 0.15]) and an ICC of 0.07. Consequently, individual-level predictors and neighborhood-level economic disadvantage were added (Models 2 and 3). Individual-level effects were consistent across all models such that other drug use, being male by birth, and age were associated with increased frequency of alcohol use. Being non-Black was associated with increased frequency of alcohol use. New HIV diagnosis status in the past 12 months and sexual orientation were not significant. Neighborhood-level effects were also consistent across all models, such that higher levels of socioeconomic disadvantage were associated with lower frequency of alcohol use. A negative cross-level interaction in Model 4 revealed that the random slope of social support on alcohol use frequency varied as a function of economic disadvantage (γ_{11} = 0.06 [95% CI: 0.01, 0.12]) such that economic disadvantage attenuated the relationship between social support and alcohol use frequency. Contrary to the hypothesis that adolescents living in disadvantaged neighborhoods would report more alcohol use, findings from this study indicated that neighborhood disadvantage was associated with less frequency of alcohol use. However, findings also revealed that as social support increased, frequency of alcohol decreased but that this effect became weaker in more economically disadvantaged regions.

15. Time and Other Considerations in Mediation Design

Meghan Cain

This project serves as a practical guide to mediation design and analysis by evaluating mediation models in a power analysis using limited data under varying conditions and misspecifications. The cross-sectional mediation model, which has been shown to be biased when the mediation is happening over time (Maxwell & Cole, 2007; Maxwell, Cole & Mitchell, 2011), is compared to longitudinal models: sequential, dynamic, and cross-lagged panel. These longitudinal mediation models take this time effect into account but bring many problems of their own, such as choosing measurement intervals and number of measurement occasions. Furthermore, researchers with limited resources often cannot collect enough data to fit an appropriate longitudinal mediation model. These issues were addressed using a simulation study. Data were simulated under multilevel mediation models and results were compared across the aforementioned mediation models, each using the same amount of data but with differing numbers of people and time points, against data with varying characteristics that may be incorrectly specified in the model. Models were evaluated using power and type I error rates in detecting a significant indirect path. Sequential mediation and to a lesser extent cross-sectional mediation analysis were found to have the best performance, though multilevel longitudinal mediation is still necessary under some conditions.

16. An Evaluation of FIML and MI Methods for Confirmatory Factor Analysis with MAR

Menglin Xu

This study seeks to compare the performance of four imputation approaches for CFA model with MAR and 30% missing rate. A Monte Carlo simulation study was conducted. Twenty-four conditions were considered (3 sample sizes * 2 factor structures * 4 imputation approaches). A two-factor CFA model was used as data generation model. Mplus 7.3 was used to impute and analyze the CFA models. Results showed that congenial imputation model (Ho_informative) with informative priors for factor loadings performed uniquely well when sample size is extremely small (n = 50) and factor loading is moderate (0.5) as compared to other imputation approaches.

17. Utilizing Latent Profile Analysis to Determine the Appropriate Mentor Relationship for Community College STEM Students

Micaela V. C. Morgan

Science, technology, engineering and mathematics (STEM) majors tend to have higher major attrition than non-STEM students. This study explored the type of mentorship community college STEM students' desire versus what they need to be successful (i.e., transfer to a 4-year university and graduate with a STEM major). A latent profile analysis (LPA) was conducted utilizing results from 343 students at two community colleges who took the STEM Engagement and Transfer Success (SETS) survey. The LPA results revealed that students reside in one of four classes of mentorship: High Mentor Characteristics and High Mentorship Products, Medium Mentor Characteristics and High Mentorship Products, Low Mentor Characteristics and High Mentorship Products and Low Mentor Characteristics and Medium Mentorship Products. It was discovered that STEM majors were 12.69 times more likely to reside in the Low Mentor Characteristics and High Mentorship Products class, which was also the class with the lowest cumulative GPA. This study suggests that STEM students could benefit from more robust mentorship and these results could inform how faculty and institutions provide mentorship to community college STEM majors in the future.

18. Iterative Target Rotation with a Suboptimal Number of Factors

Nicole Zelinsky

Exploratory factor analysis is an important analytic tool which helps researchers develop scales, generate theory, and inform structure for a confirmatory factor analysis. When a model is unknown, the researcher might use a target rotation. Target rotations partially specify the structure of the final solution; however, determining what the partially specified matrix contains can be subjective. One type of target rotation that requires less user input is called iterative target rotation (ITR) in which the researcher specifies the target matrix based on information from a previous rotation. Moore (2013) tested this rotation method through simulation and later expanded this work in Moore et al. (2015). This body of work illustrated that ITR had better-estimated factor loadings compared to the single rotation traditionally used in EFAs. However, these simulations only focused on conditions with the optimal number of factors. Therefore, little is known about how ITR functions when suboptimal factors are present compared to a traditional single rotation. Because it is impossible to tell whether the number of factors in a simulation is optimal or not, it is important to know if ITR estimates are also more accurate when a suboptimal number of factors are extracted. We tackle this point by implementing exploratory factor analysis using LibQUSL+ data (Thompson, 2004) on library quality with both a single factor rotation and an ITR. Then we examine the differences between true parameters and parameter estimates from a single factor and ITR to explore the impact different factor solutions has on final model results.

19. Controlling for Type I & II Error Using the False Discovery Rate in Large-Scale Test Security Screenings

Patrice Renee Cobb

When statistical tests are conducted repeatedly to detect test fraud (e.g., copying) the overall false-positive rate should be controlled. Three approaches to adjusting significance levels were investigated with simulated and real data. A procedure for controlling the false discovery rate by Benjamini and Hochberg (1995) yielded the best results.

20. Assessing the effect of proportion of variance in the principal-component-as-auxiliary approach

Pavel Panko, Elizabeth Green, Kyle Lang, Todd Little

In their 2015 paper, Howard, Rhemtulla, & Little established the "principal component auxiliary" (PC-AUX) approach which showed that including auxiliary variables in the form of one or more principal components allowed for maximization of the all-inclusive strategy for inserting auxiliary variables. This approach helped to condense the imputation model by providing a reasonable approximation for missing values by extracting as much (linear and nonlinear) variance from the original data as possible. However, their work included a limited range for the proportion of the variance explained by the principal components. The current simulation study will build on Howard's (2015) work by evaluating the effect of the proportion of variance explained on the bias in parameters estimated from the PC-AUX imputed data. As well, this study will determine the necessity of including non-linear components in the imputation model by examining conditions where the missing observations are related to non-linear patterns in the data.

21. Structural Equation Modeling In Diagnosis of the Marketing Research Effectiveness in the Enterprises – Problems and Solutions for the SEM Models with Many Indicators and Latent Variables

Piotr Tarka, PhD

In the article, author presents and characterizes the relationships between two general conditions (i.e., organizational and methodological), which affect the level of the marketing research effectiveness in the enterprises. The diagnosis of the researched phenomenon was based on the data collected from the sample (N = 391) of randomly selected companies located in Poland. In the analysis, a modified variant of the structural equation model (SEM) was taken into account. The former above-mentioned condition, was defined as the organizational aspect of the marketing research effectiveness, and was formed on the basis of the multidimensional construct. The structure of this construct, in its scope, included the following the common factors: 1) the importance of the marketing research for the company; 2) attitudes of the managers in relation to the problems of the marketing research projects; 3) internal environmental stimuli as well as the analytic organizational culture; 4) market orientation as the necessary condition for obtaining the effectiveness in marketing research; 5) the appropriate diffusion of the market information through internal communication channels and processes existing in the company, and 6) appropriate decision-making style. The latter condition, that referred to the methodological aspects of the marketing research effectiveness in the enterprises, in its structure contained the following common factors, such as: 7) scientific standards and principles used by companies in the implementation of the marketing research; 8) adequacy of particular types of the research; 9) selection of the appropriate research methods; 10) the proper construction of measurement instruments and the organization of the field research, 11) the adequacy and selection of the appropriate methods for data collection; 12) the appropriateness of the application of statistical methods in the marketing data analysis. The main obstacle that has appeared in the construction of the measurement models (CFA) and in the process of the structural modeling (SEM), concerned the large number of indicators falling into the respective common factors. Having conducted the exploratory factor analysis (EFA), there were generated 12 factors, due to inclusion in the analysis of 56 indicators. Moreover, all these factors were intended to compose a single SEM model. However in practice, most of so constructed meta-models (with too many variables, included in a one set) provide the unacceptable level of the model fit, as well as the misleading values of the parameter estimates, standard errors, etc. In short, due to fact that SEM models are too complex and overloaded with many indicators and latent variables, they may yield unreliable and invalid results. To avoid this problem, a solution was proposed, which included the hierarchical approach to nesting of the indicators comprising the respective factor (latent variable). Also, for the correct identification of the SEM model and further assessment of the model fit to the data, as well as the evaluation of the (indicators-factor) configuration suitability, there was initiated an iterative procedure.

22. Can survey participation affect respondents' behavior? Estimating panel conditioning using respondents' admin data

Yi-Jhen Wu, Insu Paek

Panel conditioning refers to changes over time caused by participation in a panel survey. Changes may occur in respondents' behavior (for example, participants in an election panel may be more likely to vote) and/or in their reporting of behavior (for example, respondents become more trusting of interviewers and more accurately report socially undesirable behavior). Using administrative data linked to a large panel survey, we address the challenges posed to previous researchers and analyze changes in respondents' behavior due to repeated participation in the panel, specifically whether responding in a labor market survey leads to increases in the take-up of federal labor market programs. We use propensity score weighting to estimate the average treatment effect of participation in several waves of the panel survey on participation in federal labor market programs. Results show that panel respondents participate in more labor market programs than those who were also eligible for participation but were not selected. These results suggest that panel conditioning not only affects the reporting of behavior (as previous studies have demonstrated), but can also alter respondents' actual behavior. Thus, researchers using panel data should be aware of this (often ignored) source of panel specific error.

23. Investigating the Performance of a New Approximate Measurement Invariance Approach for Dichotomous Variables

Shanshan Wang

Measurement invariance (MI) is an important pre-requisite for comparing latent means across many groups. This paper synthetizes current methodological development on MI and investigates the performance of approximate MI (i.e., Fox's random items effect model), full MI and partial MI under five different conditions, followed by an empirical illustration

where the test for full MI fails, but where allowing for approximate MI results in a well-fitting model. / The preliminary results of the simulation study show that approximate MI outperforms full and partial MI in recovering the true item parameters under varying conditions. The full MI also performs well when MI is present or when minor measurement non-invariance exists. The approximate MI is robust to different patterns of measurement invariance whereas the average correction bias is slightly larger when the item parameters are not drawn from a normal distribution. The PISA 2009 reading test data supports the hypothesis that approximate MI is more useful to estimate group-specific factor means and variances without requiring exact measurement invariance.

24. Personality Testing and the Terman Study

Valerie Ryan

The purpose of this study was to evaluate Lewis Terman's six factor model of personality with data from the Terman Study of the Gifted. Teachers rated study participants on a 25-item personality trait questionnaire (N = 1,252) during the first phase of data collection in 1922. Terman proposed that these items be split into six factors: intellectual, volitional, physical, moral, social, and emotional. To evaluate Terman's model, exploratory factor analyses (EFA), confirmatory factor analyses (CFA), and multiple sample confirmatory factor analyses (MSCFA) were conducted. / First, five data sets were imputed to account for missing data using the mice package in R. Next, each of the five datasets was randomly split in half; EFAs were conducted on half of each of the datasets in R. It was found that six factors explained the data quite well when using a Promax rotation. Five items were removed from the analysis due to complex factor loadings. CFAs were run on the other half of the datasets using EQS 6.3. The correlated model demonstrated the best fit ($\chi2(155) = 917.08$, $p < .001$, CFI = .816, RMSEA = .089) over a single factor model ($\chi2(170) = 1032.17$, $p < .001$, CFI = .791, RMSEA = .090), $\Delta\chi2$ (15) = 115.09, $p < .005$, and an orthogonal model ($\chi2(170) = 1806.98$, $p < .001$, CFI = .604, RMSEA = .124), $\Delta\chi2$ (15) = 889.90, $p < .005$. / Results from preliminary MSCFAs suggest that configural, metric, and parallel forms invariance hold when comparing trait ratings for male and female participants.

25. Optimal Number and Allocation of Data Collection Points for Linear Spline Growth Curve

Wei Wu, Fan Jia, Richard Kinai, Todd D. Little

Spline growth modelling (SGM) is a popular tool to model change processes with distinct phases in longitudinal studies. Although the analysis issues of SGM have received thorough treatment in the past, little attention has been paid to the design issues of SGM. An important design issue of SGM is how to collect longitudinal data in a way to maximize the efficiency of estimating target parameters in a spline growth model. To address this question, the optimal number and allocation of repeated measures need to be determined with possible resource constraints taken into account. Focusing on linear spline growth models with two phases and a fixed change point (the transition point from one phase to the other), this study evaluates the efficiencies of numerous data collection designs, each represents a unique combination of number and allocation of repeated measures, using a Monte Carlo Method. Two resource constraints are considered: budget and sample size constraints. The result suggests that the optimal number of repeated measures should be no more than six (four or five in most conditions). The optimal allocation of the repeated measures is dependent on the location of the change point, the kind of resource constraint, and the amount of certainty on the change point. Efficient designs are recommended for the cases where the exact location of the change point is known (complete certainty), the cases where only the interval that contains the change point is known (partial certainty), and the cases where no prior knowledge on the location of the change point is available (zero certainty).

26. Longitudinal School District Effectiveness Identification in Ohio and Texas: Using Hierarchical Growth Model Comparisons to Promote Cross-District Improvement

Xinyu Ni; Alex J. Bowers; Jennifer Esswein

The purpose of this study is to apply recent district effectiveness research (DER) site identification models across states, applying hierarchical growth model in Ohio and Texas. We identify groups of districts from the entire population of a state that are significantly outperforming peer districts, and match them for central office and state administrator professional development opportunities to promote state-level cross-district information sharing and improvement cycles. Identifying unusually effective school districts from the entire population of districts throughout a state is an ongoing problematic issue for research, policy and practice. In the present study, our central research questions are: 1) To what extent can school districts be identified from all of the districts in Ohio and Texas that significantly outperform or underperform long-term performance trends across multiple indicators? 2) To what extent does this identification system compare to each state's current rating system? 3) To what extent do the district financial variables play a role on the growth trajectories of

school districts for both Texas and Ohio. This study relies on longitudinal Ohio and Texas public data matched to the U.S. Department of Education National Center for Education Statistics Common Core of Data and District Finance Survey. Specifically, we focus on administrative panel data for the years 2006-2013 for all Ohio districts (n=608) and 2005-2011 for all Texas districts (n=1038). The dependent variable in the model is a district-wide Performance Index (PI) Score, which captures the yearly performance of students in a district across all schools and subjects tested, from grades 3 through 12 and including mathematics, reading, social studies and science. The Ohio Department of Education calculates the PI score for districts. Lacking a single PI score district performance metric in Texas, the full final paper details our efforts to apply PI score policies from Ohio and similar states to Texas. The independent variables in the models include district demographic and context variables as well as finance expenditures. As an advance over Value Added Models (Hallinger, Heck, & Murphy, 2014; Heck & Hallinger, 2014), we extend the recent research in applying hierarchical linear growth modeling to longitudinal district performance trend analysis (Bowers, 2010, 2015) to replicate the model with additional years in Ohio and test the model in Texas for the first time. The HLM growth model framework replicated in Ohio with additional years and finance variables works well in Texas, identifying 32 districts that outperformed their context and demographic variables in comparison with peer districts in the state. Consistent with the previous research results (Bowers, 2015), there were multiple significant variables on the estimates of the intercepts- the average Performance Index score for the first year-for the models of both Texas and Ohio. Additionally, the different types of financial expenditure variables did play significant roles in the estimates of both intercepts and slopes for the two states. By including the financial variables, the intraclass correlations coefficients of models for both Texas and Ohio decreased and variance explained between groups increased. Importantly, we also pilot visualizations to compare the HLM growth model results to each state's current district accountability rating systems through a quadrant plot. Our aim is not to replace state systems, but to augment the conversations in states around district effectiveness, providing additional means to identify effective or less effective districts. Our goal is to help districts bring together central office personnel for professional development, in which district contexts can be matched across the two rating systems, such as the difference in if a district is rated highly by the state and the HLM growth model, low in one high in the other or low in both, as the HLM takes advantage of long-term growth while the state system is used for AYP status.

27. Accuracy of person latent classification in the Mixture Rasch model under Maximum likelihood estimation

Yi-Jhen Wu, Insu Paek

Mixture IRT modeling, which combines advantages of both item response theory (IRT) and latent class analysis (LCA) I, has been used extensively in education and psychology fields. Most studies have investigated the performance of mixture IRT in terms of item profile recovery, model selection and the estimation methods. However, there has been much less emphasis on the accuracy of classification of person latent groups and the factors affecting the accuracy. This simulation study focused on the accuracy of classification of person latent groups under maximum likelihood estimation (MLE) and examined what factor(s) and their interactions are influential on the accuracy. Results showed that the number of items in a test and its interaction with the number of latent groups are major influential effects on the accuracy.

28. Characterizing Latent Classes using Auxiliary Variables in Mixture IRT Models: A Simulation Study of the Influence of Different Response Scales on Measurement Invariance

Zachary K. Collier, Walter L. Leite

The issue of bias in rating scales, failure to correct for different tendencies in responding to items, has been well discussed in previous literature. The Polytomous Mixture Rasch Model (PMRM) has been proposed as a tool to detect measurement invariance and characterization of response tendencies using external variables. This study simulated continuous and categorical external variables while investigating how well the PMRM recovered model parameters under various sample sizes, response scales, and number of latent classes. Additionally, the effect of different ordered response categories on measurement invariance was tested. Item responses that characterized extreme response style (ERS), middle-category response style (MRS), acquiescent response style (ARS), and ordinary response style (ORS) on a 3-,4-, and 5-category Likert scale were generated using probabilities from previous literature. According to Cho's (2013) study, the PMRM is flexible enough to account for response styles on rating scales. Preliminary implications indicate misspecification impacts the overall accuracy of person trait estimation. This model provides insight into the variety of ways people answer test items based on their background.

29. A comparison of recently introduced point and standard error estimation procedures for multidimensional IRT modeling in a R package "mirt" and a commercial software flexMIRT

Zhongtian Lin, Insu Paek

Multidimensional item response theory (MIRT) model applications are increasing. To improve the accuracy and efficiency of MIRT model estimation and to cope with high-dimensional applications, recently in the IRT literature, Metropolis–Hastings Robbins–Monro (MHRM) algorithm and Quasi-Monte Carlo estimation (both for point estimation), as well as sandwich estimator and supplemental EM (both for standard error estimation), have been introduced and implemented into both commercial IRT programs and free R IRT packages. However, no study have focused on evaluating and comparing systematically the performances of these methods on both platforms. This simulation study investigates the performance of a well-received R package "mirt", relative to a popular commercial package "flexMIRT" under various multidimensional data conditions, in terms of parameter recovery and standard error estimation. Results from this study will generate evaluative information of the R mirt software compared with flexMIRT, especially on those recently introduced estimation methods mentioned above, so that researchers and practitioners have an opportunity to examine the quality of both software when they consider those methods for their research or other practical purposes. In addition practical information such as run time will be generated and compared as well.

30. Exploring Ethnic Differences in a Revised Mathematics Achievement Model: Applications of SEM and Multi-group Analysis

Leigh M. Harrell-Williams, James Ford, Christian Mueller, Caroline Hart

Research has shown that academic identity, interest, value, and efficacy can predict students' desire to engage in academic tasks and impact academic achievement. The present study sought to 1) propose a variation of Middelton's (2013) model of academic achievement by using a more closely-aligned Eccles (2005) inspired model of motivation and engagement and 2) to employ Multigroup Structural Equation Modeling with Black, Hispanic, and White students using our proposed model to further tease apart potential motivational differences among these groups. Similarities and differences among the proposed model and the Middleton model are discussed. Results of the multi-group analysis provide some evidence that the impact of math self-efficacy on math identity and the impact of math interest on math value varies across Black, Hispanic, and White students.

31. Association between Relationship Education and Depressive Symptoms in Fathers with Young Children: A Complier Average Causal Effect Estimate

Jaipaul L. Roopnarine, Elif Dede Yildirim

Objective: Drawing on stress and coping theory and family science models that emphasize preventative science, we used Complier Average Causal Effect (CACE) Estimates to determine whether relationship skills education moderated the associations between fathers' avoidance of destructive conflict behavior, constructive conflict behavior, support and affection, and depressive symptoms assessed when their children were 15-months old and depressive symptoms when their children were 36 months old. Method: The sample consisted of low-income Hispanic American, European American, and African American fathers (N=2,540) from the Building Strong Families Study. Fathers from eight sites across the US were randomly assigned to either a treatment group who received relationship skills education or a control group. Fathers depressive symptoms were assessed prior to and one year after relationship skills education classes by using the CES-D. Results: Paternal age and family residential stability predicted whether fathers utilized relationship skills education classes. Relationship skills education moderated the associations between depressive symptoms and deconstructive conflict behaviors at 15 months and depressive symptoms at 36 months. The impact of dosage of relationship skills education on depressive symptoms was inconsistent under different missing data assumptions and dependent on compliance. Conclusions: Data are interpreted in the context of the efficacy of intervention programs for tempering mental health difficulties in low-income fathers with young children. Public Health Significance: This study underlines the significance of relationship skills education in lessening depressive symptoms in low-income fathers with young children.

32. Applying the two-stage meta-analytic structural equation modeling approach to test the theory of planned behavior of alcohol consumption-Poster

Ivan Jacob Agaloos Pesigan, Shu Fai Cheung

Meta-analytic structural equation modeling (MASEM) combines the techniques of MA and SEM in order to synthesize correlation or covariance matrices from primary studies and fit structural equation models using the pooled matrix. In this study, we employed the two-step MASEM approach proposed by Cheung and Chan (2005) to test the theory of planned

behavior (TPB) framework proposed by Ajzen (1985) in the context of alcohol consumption. The theory of planned behavior links beliefs and behavior. We tested a mediation model where intention to consume alcohol mediated the effect from attitude toward alcohol consumption, subjective norm, and perceived behavioral control to actual alcohol consumption. We also specified a direct path from perceived behavioral control to alcohol consumption and controlled for the effect of past behavior. In the first stage we pooled 16 correlation matrices (n=2,888). The correlations varied from study to study, so we accounted for heterogeneity in effect sizes by employing random effects model in synthesizing the correlation matrices from the primary studies. We then used the pooled correlation matrix from stage one to fit the mediation model in stage two. A correlation structure was fitted with the weighted least squares (WLS) estimation by using the inverse of the sampling covariance matrix of the average correlation vector as the weight. The pooled matrix fit the model adequately with $\chi 2(3) = 13.809$, $p = .003$, CFI = .988, and RMSEA = .035. Statistical significance of the parameter estimates was tested using 95% likelihood based confidence interval. Attitude and subjective norm positively predicted intention to consume alcohol with $\beta = .64$ (.53 to .66) and $\beta = .19$ (.06 to .30) respectively. Perceived behavioral control negatively predicted intention to consume alcohol with $\beta = -.25$ (-.41 to -.10). The direct effect of perceived behavioral control to alcohol consumption was nonsignificant $\beta = -.04$ (-.22 to .11). The indirect effects of attitude, subjective norm, and perceived behavioral control on alcohol consumption via intention to consume alcohol were all significant with $\beta = .22$ (.15 to .30), $\beta = .06$ (.02 to .11) and $\beta = -.09$ (-.15 to -.04) respectively. For the most part, the results confirm the predictive validity of TPB in predicting health compromising behavior particularly in the context of alcohol consumption. By employing MASEM, the study provides a more generalizable conclusion regarding the nature of the effects of TPB variables on alcohol consumption with attitude having the strongest indirect effect followed by perceived behavioral control and subjective norm. The sign of the overall effect of perceived behavioral control on behavioral was, however, inconsistent with the literature. The nature of the construct should be further analyzed. Possible moderators like the nature of the behavior being studies (e.g., light episodic drinking vs. heavy drinking) should be examined to account for the inconsistency. The results also suggest that beliefs continue to have an effect on behavior even after controlling for past behavior suggesting that beliefs can affect behaviors over and above past behavioral patterns. These results can inform intervention efforts to reduce alcohol consumption by targeting attitudes, subjective norms, and perceived behavioral control in order to reduce intention to consume alcohol ultimately reducing alcohol consumption.

33. The Impact of Prior Distributions in Latent Class Analysis Models: Simulation and Application

James P Clifton and Sarah Depaoli

Latent class models have been shown to exhibit poor parameter recovery and low convergence rates under the traditional frequentist estimation approach, namely, maximum likelihood via the EM algorithm (see e.g., Tueller & Lubke, 2010). The Bayesian estimation framework may be a viable alternative for estimating latent class models--especially when categorical items are present and priors can be placed directly on the categorical item-thresholds. The current poster presents two studies involving Bayesian latent class analysis (LCA) with categorical items. The first study is a Monte Carlo investigation demonstrating that the frequentist framework and the Bayesian framework with diffuse (noninformative) priors were unable to properly recover the latent class proportions and many of the other model parameters (e.g., latent class item-thresholds); a substantive interpretation of the obtained results would lead to improper conclusions under these estimation conditions. However, specifying informative priors and some types of weakly informative priors within the Bayesian framework produced accurate parameter recovery, indicating that this may be a more viable estimation approach for LCA models with categorical indicators. The second study presents an empirical example using the Youth Risk Behavior Surveillance System survey to illustrate the implementation and specification of priors through a substantive LCA example. The paper concludes with a general discussion surrounding the advantages of Bayesian estimation for LCA models. Suggestions for effectively implementing the Bayesian approach are also included.

34. Modeling Age-Related Changes in Emotion Regulation within a Hierarchical Latent Stochastic Differential Equation Framework

Julie Wood, Nilam Ram, Pamela Cole, Sy-Miin Chow

Lifespan theories of aging suggest both improvements and decrements in individuals' motivation or ability to regulate emotions over time. Multiple time-scale study designs provide new opportunities to examine long-term changes in the short-term dynamics of regulatory processes. Stochastic differential equations provide concise operationalizations of those dynamics. Merging theory and method and data, the goal of this paper is to demonstrate the utility of Ornstein-Uhlenbeck (OU) process models for describing specific aspects of short-term regulation and how they change over the long-term. Applying the Bayesian Hierarchical Ornstein-Uhlenbeck Modeling toolbox (Oravecz, Tuerlinckx, &

Vandekerckhove, 2009) to data from the iSAHIB study (Ram et al., 2014), wherein 150 adults aged 18 to 89 years reported on their emotion valence and arousal after 64217 social interactions (average 6.71 per day, SD = 2.70) over the course of one year, we demonstrate how the hierarchical dynamic model can be used to articulate and examine health- and age-related changes in dispersion and attractor strength of individuals' emotion reactivity/regulation, as well as the role specific regulatory strategies (reappraisal, suppression) play in moderating those processes.

35. Understanding Mathematics Performance from Student Attitudes and Behaviors: A Multiple-Group Analysis Using International Student Assessment Data

Kalina Gjicali, Stefan Krumm, Anastasiya A. Lipnevich

To date, there have been limited efforts to using the attitude-behavior relationship in understanding achievement outcomes in mathematics as measured through international achievement performance standards. Using data from PISA 2012, in the proposed research program we plan to close this chasm and answer a series of questions that relate mathematics attitudes to students' behavior and academic performance. By using the theory of planned behavior (TpB) framework, student attitudes, intentions, and behaviors will be used to predict mathematics performance. A multiple-group analysis is proposed, where data from participating countries in the PISA 2012 will be used to understand the TpB model and its relationship to mathematics achievement examined at both the with-country and between-country level.

36. Bayesian Model Comparison for Longitudinal Growth Models

Kathrene Valentine, Phillip K. Wood, Edgar Merkle

In a recent article, Wood, Steinley, and Jackson (2015) outlined several statistical models for longitudinal data which vary in terms of their dimensionality, parsimony, and degree of parametrization and argued that researchers should consider comparison of such models by means of a three-step "right-sizing" approach to model comparison. These candidate growth models included several of the traditionally used growth models based on linear or polynomial effects, but also included free basis growth models, McDonald's Linear Factor model, and parametric sigmoid growth models such as the Schnute, Gompertz, von Bertalanffy, and Richards curves to name a few. Although their article demonstrated the feasibility of the approach using frequentist maximum likelihood approaches, the success of recent Bayesian approaches to growth modeling lead to the question of whether such approaches can be applied to the model comparisons described by Wood et al. In this poster, Bayesian model estimation and comparisons are presented for one real-world data set (of children's vocabulary considered in Wood et al.) and one simulated data set (a Gompertz curve). Model specification for free basis, linear and quadratic growth models can be readily accommodated using the Blavaan package in R (Merkle & Rosseel, 2016). Estimation of parametric growth models is accomplished by adapting from the JAGS scripts provided by Blavaan in order to estimate the growth parameters assumed by the sigmoid curves. Model comparison of the growth curves is evaluated using DIC, BIC, and Bayes Factors. Although some care must be taken to assure that initial values for the MCMC sampler are reasonable and that appropriate thinning is conducted for some models (such as McDonald's Linear Factor model), Bayesian model comparisons involving the Bayes factor appear particularly promising.

37. The Impact of Item Scale Misspecification and Dichotomization on Class and Parameter Recovery in LCA of Count Data

Kathryn S. Macia, Robert E. Wickham

Mixture modeling is increasingly being used in the social and behavioral sciences, and this is particularly the case in substance use research, where observed items are often measured on a count scale. A review of recent articles published in addiction journals which used latent class analysis (LCA) and included at least one substance use indicator measured on a count scale revealed that 65% dichotomized count indicators, and of the articles that did not dichotomize, none had any indication that a count distribution was used. This study used simulated data to examine the impact of dichotomizing and misspecifying count items as normally distributed on class and parameter recovery in LCA. Data were generated according to a four class LCA model with seven Poisson distributed items. Simulation conditions varied by sample size and class separation. Each simulated data set was then analyzed under four item conditions: distributed Poisson, distributed normal, dichotomized by any endorsement (0 vs. 1+), and dichotomized by median split, and results were compared on ability to recover true number of classes, class assignments, and mixture parameters. Assuming items were distributed normal resulted in the most bias, whereas dichotomizing the items generally led to much less bias, with the analyses for the median split condition in particular performing almost as well as the Poisson condition. Overall, findings underscore the deleterious effects of item distribution misspecification in LCA of count data and suggest that dichotomizing based on sample-specific distributional cut points may have negligible impact on class and parameter recovery.

38. Toward a typology of technology-using teachers: A Latent Class Analysis (LCA) of the Fast Response Survey System Teachers' Use of Educational Technology, 2009 (FRSS 95)

Kenneth E. Graves, Alex J. Bowers

The purpose of this research is to investigate the extent to which there is a typology of teachers who use technology using the 2009 Teachers' Use of Educational Technology in U.S. Public Schools Fast Response Survey System (FRSS 95). The amount of prior research on teachers' technology use is comprehensive and vast; however, there are few studies that use national data to describe teachers' technology habits by examining subgroups of individuals. Using Latent Class Analysis (LCA) with a nationally representative dataset (N=2,764), results show that there are four significantly different types of technology-using teachers: Dexterous, Presenters, Assessors, and Evaders. Furthermore, contextual factors, like the percentage of students on free and reduced lunch, school types, hours of professional development, years of experience, and the number of total computers, are significantly associated with the odds of belonging to some of the groups. Implications on technology integration practices, teacher professional development, teacher evaluations, and school technology leadership will be discussed.

39. Handling Sparse and Missing Data in Estimating Growth Trajectories: A Functional Mixed Effects Model Approach

Kimberly Ward, Hye Won Suk

Researchers often encounter an issue in longitudinal analyses: how to deal with sparse and missing data. Sparse data occur when researchers have limited measurements on each individual. When only a few measurements are available per person, it can be difficult to get an accurate picture of the underlying trajectory for each person. Additionally, missing data can occur due to attrition or participants missing measurement days. When data are missing randomly, analyses can be carried out in the usual fashion without concern for any bias. However, when data are missing in a systematic fashion results may be biased if special care isn't taken to account for the missingness during analyses. Moreover missing data can exacerbate sparseness. We approach these issues by treating longitudinal data as functional data, and analyzing it in the functional data analysis (FDA) framework. FDA assumes that functional data arise from infinite-dimensional, smooth functions evaluated at a finite number of time points. FDA is unique in that it analyzes the underlying curves or functions rather than the observed discrete data. It is a powerful tool for estimating complex, nonlinear trajectories. This is due to FDA's non-dependence on parametric models (e.g., linear, quadratic, or exponential) to describe the shape of the trajectories. Among various FDA methods, we specifically focus on a functional mixed effects model (FMEM) that is useful for estimating growth trajectories. A FMEM is an extension of the standard mixed effects model for functional data. As the standard mixed effects model, FMEMs estimate both fixed and random effects. However, the estimated fixed and random effects are curves. The fixed effect curve represents the overall grand mean trajectory and the random effect curves represent the person-specific trajectories deviating from the grand mean trajectory. FMEMs use data from all individuals to estimate the grand mean curve and person specific curves, and thus, are them useful tools for handling sparse and missing data. However, there are no studies available that systematically examine how various degrees of sparseness and different types of missing data potentially affect the accuracy of the mean and individual curves estimated by FMEMs. Thus, we perform a simulation study to examine the performance of FMEMs under various conditions combining the five factors: sparseness level, shape of underlying trajectory, irregularity of time points, type of missingness, and percent missing. The mean square errors of both estimated mean curve and individual curves under each condition will be presented. In addition, suggestions and guidelines will be provided for substantive researchers who want to use FMEMs.

40. A Bayesian Vector Autoregressive (VAR) Model with Nonignorable Missingness in Dependent Variables and Covariates: Development, Evaluation, and Application to Emotion Processes

Meng Chen, Linying Ji, Zhao-Hua Lu, Zita Oravecz, Sy-Miin Chow

One common problem in longitudinal studies is the presence of missing data. Intensive longitudinal data involving repeated assessments of constructs such as emotions are particularly prone to nonignorable missingness, namely, missingness where the missing data mechanism depends on unobserved information. For example, in measuring day-to-day changes in negative emotions, it is possible that the participants may opt not to report their feelings on the days with heightened negative emotions, thereby leading to nonignorable missingness when modeling emotion processes. While processes such as emotions are often modeled using time series models such as vector autoregressive (VAR) models, the consequences of omitting or misspecifying the missing data mechanism in the context of these models are not well understood. In this study, we utilize a Bayesian approach to simultaneously model missingness in the

dependent variables and covariates in the context of the VAR model. A simulation study was conducted to evaluate properties of the proposed Bayesian VAR model under different missing data mechanisms, and when paired with different degrees of model misspecification. Power and sensitivity of commonly adopted Bayesian fit statistics and model comparison criteria (including Deviation Information Criterion; Fung, Wang, & Seneta, 2014; Spiegelhalter, Best, Carlin, & Van Der Linde, 2002; Bayesian Information Criterion; Schwarz, 1978; and Conditional Predictive Ordinate; Geisser & Eddy, 1979; Gelfand, Dey, & Chang, 1992) in detecting misspecification in missing data mechanism were also investigated. The proposed model is illustrated using a set of ecological momentary assessment data from the Affective Dynamics and Individual Differences (ADID; Emotions and Dynamic Systems Laboratory, 2010) study in which the relation between individuals' positive and negative emotions are modeled using a VAR model with stress and instances of positive/negative events as covariates.

41. Bayesian mediation analysis with latent variables in the Bayesian SEM framework

Milica Miocevic, David MacKinnon

Statistical mediation analysis is widely used in social sciences to describe the mechanisms through which the independent variable affects the dependent variable (MacKinnon, 2008). For example, an intervention to reduce smoking might change attitudes toward smoking, which in turn might decrease the number of cigarettes smoked. Mediated effects in models with manifest variables and only one continuous mediator and outcome are often estimated using Ordinary Least Squares (OLS) regression, and more complex models involving several correlated mediators and outcomes are often estimated using Maximum Likelihood (ML) in the Structural Equation Modeling (SEM) framework. Bayesian mediation analysis has been described for models with manifest variables (Yuan & MacKinnon, 2009; Enders, Fairchild, & MacKinnon, 2013), but not for models with latent variables. This project will give an example of how to specify a Bayesian mediation model with a latent mediator and latent outcome in the Bayesian SEM framework, and will highlight important statistical considerations for substantive researchers working with latent mediators and outcomes in the Bayesian framework.

42. Dyadic Analysis using Grid Sequence Methods: Inter-Dyad Differences in Intra-Dyad Dynamics

Miriam Brinberg, Timothy Brick, Nilam Ram, Denis Gerstorf

Spouses are proximal context for and influence each others' behaviors, particularly in old age. Dyads are typically studied either with longitudinal panel approaches that emphasize between-dyad differences or time-series approaches that emphasize within-dyad dynamics. In this paper, we forward an integrated approach that merges state space grid methods used by child developmentalists to study parent-child dyads with sequence analysis methods used by sociologists to study work/family histories into a "grid sequence" method for studying inter-dyad differences in intra-dyad dynamics. Using experience sampling data obtained from six times daily reports obtained from 90 older couples (M = 75.0 years, SD = 3.79) over 7 days, we illustrate how a "grid sequence" analysis can be used to identify a taxonomy of dyads with different activity-emotion dynamics. Results provide a basis for measuring a set of truly dyad-level variables that capture stable equilibrium (attractors), the typical paths to equilibrium (equilibrium dynamics), and the inter-dyad differences therein. Simultaneous examination of dyadic relations at micro- and macro-dynamics is relatively rare. Methodologically, this project expands the toolbox of techniques for analysis of longitudinal dyadic data. Substantively, we identify patterns of dyad-level micro-dynamics that may serve as new markers of risk/protective factors, and potential points for intervention into older adults' proximal context.

43. Handling item non-response in Structural Equation Modelling with ordinal variables

Myrsini Katsikatsou, Irini Moustaki

As maximum likelihood estimation is computationally infeasible for Structural Equation Models (SEM) with ordinal variables, limited-information estimation methods have been developed such as the three-stage least squares (3S-LS), which is the conventional estimation approach for SEM with ordinal variables, and the pairwise likelihood (PL) method, which is recently proposed. The PL estimator is defined as the value of the parameter vector that maximizes the PL function. The PL function for an observation is defined as the product of the bivariate likelihoods over all pairs of variables. PL has been found to exhibit competitive performance to 3S-LS with respect to bias and MSE of the parameter estimates and standard errors when the data are complete. In the presence of missing data, 3S-LS and PL estimators can ignore the missing values when they are missing completely at random (MCAR) but not when they are missing at random (MAR); otherwise, they yield biased estimates. Under 3S-LS, multiple-imputation (MI) is mainly used to deal with MAR. However, MI relies on the assumption that the imputation model is correct. The predictors of the imputation model may need to be selected carefully to avoid multicollinearity which can affect the computational feasibility and efficiency of the approach.

The computational time may increase as the number of ordinal variables that serve as predictors increases or when more general unrestricted models are used in an attempt to minimize the risk of misspecifying the imputation model. We compare three PL estimators under MAR with the MI 3S-LS estimator in terms of bias and MSE and easiness of application. In particular, the complete-pair (CP) PL, available-case (AC) PL, and doubly-robust (DR) PL are considered. The CP-PL and AC-PL draw on the approach of full information maximum likelihood with missing data, while DR-PL draws on the theory of doubly-robust estimators where a predictive model for the missing values given the observed and a model for the probability of different observed missing patterns are considered along with the model for the observed data.

44. Evaluating the Use of Propensity Scores with Multilevel Data

Noah Greifer, Felix Thoemmes

Propensity scores remain a popular tool to reduce bias due to observed confounders. In the context of multilevel data, in which individual units are clustered within groups, using propensity scores requires additional considerations that must be addressed and reported in order for researchers to convincingly eliminate imbalances in their data. These considerations include whether and how to include group-level covariates in modeling the propensity score, whether conditioning on the propensity score should occur within groups or across groups, whether group-level covariates or cross-level interactions need to be considered in balance checking, and how the stable unit treatment value assumption is met when interference due to the clustering is probable. We conducted a systematic review of 16 studies across disciplines that used propensity scores in the context of multilevel data published up to the fall of 2015. For each published paper, we identified choices made to address the aforementioned considerations and evaluated the reporting of the analyses. Based on this review we suggest best practices for the use and the reporting of propensity scores with multilevel data.

45. Modeling and Linking Speeded Tests: An Application with the VASE Vocabulary Tests

Ruben Castaneda, Yang Liu, Nicole A. Zelinsky, Judith A. Scott, Susan L. Flinspach, Jack L. Vevea

In theory, a pure test of speed is composed of a large set of easy items and administered with a time limit, whereas a pure test of power is composed of items with varying difficulty levels and administered without a time limit. In practice, tests often contain both power and speed components, which call for a proper modeling framework to account for the dependency among item responses coming from both sources. This research uses the Vocabulary Assessment Study in Education (VASE) 2014 dataset to illustrate item calibration, scoring, and linking when neither the speed nor the power component of the test is negligible. We first fit an item response model assuming a pure power test, and assess the violation of local independence among items located towards the end of the test. Then, a multidimensional calibration model is specified to capture the residual dependencies caused by speededness. Finally, linking between multiple test forms is performed via calibrated projection. This study extends previous research in three aspects: a) the influence of speededness is empirically evaluated by local dependence diagnostics, b) the speed nature of the test is considered at the item calibration stage, and c) the score conversion between two speeded test forms is obtained by a recently developed linking method.

46. Bayesian multilevel factor analysis of the Emotion Regulation Checklist: Examining cross-level and longitudinal measurement invariance

Steven J. Pierce, KyungSook Lee, Holly Brophy-Herb, Laurie Van Egeren

The Emotion Regulation Checklist (ERC) is a 24-item published survey instrument that asks teachers or parents to rate a preschool-age child's ability to regulate emotions, which is important to children's success in classroom settings. We used the ERC in a multi-site, cluster-randomized longitudinal trial testing the efficacy of a science curriculum in Head Start preschool classrooms (56 teachers, 522 students, 2 longitudinal observations). We assessed children's emotion regulation as part of this study. Our analyses revealed a factor structure to the ERC that differed from previously published results on the measure. Therefore, we present our own Bayesian multilevel structural equation model where two latent factors (regulatory skills and temperament) are measured by 8 ordinal items selected on the basis of content. Exploratory and confirmatory analyses revealed that item-level ICCs ranged from 0.068 to 0.329, good composite reliability for both factors, and that teachers' ERC ratings have factor loadings that are invariant across level (between and within teachers) and over time. Additional analysis with parent ratings suggest that the factor loadings are invariant across rater type.

47. Models for Understanding & Predicting Consumer Perception of Radiance

Supriya Satwah, Anthony Cece

Radiance is a hot topic throughout the cosmetics and personal care industries. We have focused on understanding Perceived Aging, Healthy Skin appearance for both Body and Face, and to a much lesser extent Radiance and or Glow. This research is focused on how to understand and quantify "natural radiance" obtained through photo evaluation via its relationship to quantitative parameters culled from image analysis and color measurements on groups of women ages 18-60 in Thailand and Shanghai. Perceived Radiance/Dullness in both countries appears to be related to color, spots, and texture however the balance of the contribution of each of these appears to be different between the two countries with increased focus on Spots and Texture in China. In both countries L* is a key contributor to the perception of Radiance/Dullness with increasing L* (lighter) related to increased perception of Radiance and decreased perception of dullness. A variety of analysis techniques were employed to understand the relationships between consumer perception and the quantification of attributes. Bayesian nets were used to gain a sense of the overall structure and relationships in the data. Partitioning methods were used to build trees for understanding key factors related to the dependent variables. The primary model building tool used was Partial Least squares analysis to develop and refine the final models.

48. An application of Bayesian weighted linear regression using survey weights: preschool childcare and body mass in Quebec children

Tanya J. Murphy, Nandini Dendukuri, Jay S. Kaufman, Thomas Charters, Seungmi Yang

Research question: Does body mass in early childhood differ by preschool childcare arrangement (CCA) and, if so, how much of the difference is attributable to the CCA? The study population was a representative birth cohort of infants born near-term from Quebec, Canada. The study period focuses on 1999 to 2003—shortly after the introduction of the provincially-funded, low-cost and quality-monitored childcare program. Study subjects were restricted to children who at enrolment (5 months of age) were neither obese nor wasting [and had not started full-time non-parental childcare (n=1500)]. The differences in Body Mass Index z-score (zBMI, WHO standards for Canadians) by CCA was estimated for children aged 2 to 7 years old. CCA types were categorized as 1) regulated centre-based, 2) regulated family-based, 3) unregulated non-parental care, and 4) parental care (reference group). Average zBMI was estimated using hierarchical linear models to regress yearly measures of zBMI for each child on CCA types (plus covariates). We aimed to summarize and balance potential confounders across groups by estimating a propensity score for the multinomial CCA exposure. These analyses were conducted using Bayesian methods with non-informative prior distributions for the parameters. Anticipated results and discussion: The unadjusted and adjusted differences in zBMI between CCA groups, as well as their interactions with time and family socioeconomic position are reported. [Non-inferiority hypotheses will be tested for the pairwise comparisons between CCA types with the a priori cut-off for inferiority defined as a 10% greater [incidence/prevalence] of very overweight or obesity (zBMI > 2.5).] The primary conclusions derive from Bayesian statistical inference, but are compared to classical analyses. Propensity score estimation for sub-groups—in particular two- and lone-parent families—and Bayesian methods for using survey weights are discussed.

49. Application of multiple groups propensity score weighting to measure the effect of science disciplines in beliefs about science

Valentina Giaconi, Patricio Felmer, Pascal Bressoux

This study is about the effect of science courses in science related beliefs of high school students. These courses are done during the summer in a university environment. To enter students have to apply, are selected by their school grades and have to pay a fee. To understand which kind of scientific discipline produces change in beliefs we divided the students in three groups: students in courses about life objects (for example Cellular Biology); students in courses about non-life objects (for example Mathematics); and students which applied but finally didn't enroll to the courses (control group). In the three groups we collected data about science related beliefs with a likert-type questionnaire before and after the summer school. With the function mnps of the package Twang in R, the three groups were balanced using weightings derived from propensity scores calculated with pre-treatment variables. We compared the beliefs post-treatment in the weighted groups. In general, students that took courses about life objects reported to feel more capable of learning sciences and having more sophisticated views about science in comparison with students from courses about non-life objects. In most of the variables, no differences with the control group were found.

50. The Influence of Family Literacy Context on Adolescents' Math Achievement: A Bayesian Hierarchical Regression Analysis with Informative Priors

Zijia Li, Hao Zhou, Hongwei Yang, Ruixue Liu

This paper aims to study the influence of family literacy context on 15-year-old students' mathematics achievement in the US. The primary dataset for Bayesian modeling is derived from the Program for International Student Assessment (PISA) in 2012 with a secondary 2003 PISA dataset serving to provide the information needed to specify informative priors. As is known, the cycles of PISA are implemented and administered under comparable conditions so that the use of results from the prior cycles as informative priors is strongly justified. In the study, the dependent variable is a composite measure of mathematics literacy score, calculated from an exploratory factor analysis of all five PISA 2012 mathematics achievement plausible values for which multiple evidences are found supporting data unidimensionality. Among the independent variables are multiple family literacy context variables, including possessions of poetry, art, dictionary, study space, number books at home, etc. Besides, demographic information of students and parents is also included as controlled variables, such as students' gender, race, and parents' education and job status. The same set of variables has been selected from both PISA 2012 and 2003. Prior to the analysis, the study first addresses the issue of data missingness on key variables involved in the study using multiple imputation. To that end, the study implements an MCMC-based fully conditional specification method to generate one fully imputed data set for each of the two PISA datasets. Then, each full dataset is analyzed in place of the original one to address the research questions posed in the study. Specifically, the study takes a hierarchical approach where predictor variables are entered into the regression model sequentially in the order of theoretical importance. By evaluating each model using available Bayesian model fit statistics, a best-fit and most parsimonious model is selected where Bayesian statistical inference is conducted for the purpose of result interpretation and discussion. Through a variance component analysis of the dependent variable in each data set, the study identifies evidences supporting the use of multilevel modeling (student and school levels) in both cases. Therefore, for each fully imputed PISA dataset, the same hierarchy of multilevel models is specified where the models for the 2012 data are formulated under the Bayesian framework with conjugate priors for all structural and covariance parameters. The 2003 hierarchy of multilevel models is next estimated under the traditional maximum likelihood method to attain the parameter estimates for specification of informative priors for a subset of the parameters of the Bayesian multilevel models of the 2012 data. Finally, the 2012 models with informative conjugate priors are analyzed under the Bayesian framework using Gibbs sampling. As a baseline of comparison, these models are also analyzed using noninformative conjugate priors. The final interpretation and discussion of results is based on a Bayesian statistical inference of the 2012 models under the two types of priors.

51. Methods matter in multilevel mediation analyses

Zuchao Shen, Dr. Ben Kelcey

Studies have found that various inferential tests perform differently (i.e., statistical power and type I error) in detecting single-level mediation effects. Literature is incomplete about the performance of different inferential tests in detecting multilevel mediation. We investigated the performance of five inferential tests (e.g., first-order delta method, second-order delta method, distribution of the product, Monte Carlo confidence interval, and test of joint significance) in detecting multilevel mediation effects through a simulation study. Specifically, we compared the performance of these inferential tests along several primary design dimensions: 1) different multilevel mediation models, 2) magnitude of the mediation effects and direct effects, 3) the variation of intraclass correlation coefficients, 4) sample sizes at different levels, and 5) different designs. In general, concerning the statistical power and type I error rates in detecting 2-level mediation effects, especially for 2-2-1 and 2-1-1 mediation models, the distribution of the production and the Monte Carlo methods perform the best, and thus are recommended. The first order and second order delta methods provide less power and less accurate type I error rates, and thus are not recommended for researcher to use. But for 1-1-1 mediation models, the joint significance method performs the best, this is consistence with studies on simplest mediation effects that the best balance of Type I errors and statistical power across all cases is joint significance.

Modern Modeling Methods Conference
May 23-24 2017
University of Connecticut

Call for Proposals due February 1st, 2017

The Modern Modeling Methods (M^3) conference is an interdisciplinary conference designed to showcase the latest statistical modeling methods and to present research related to these methodologies. The sixth annual M^3 conference will be held at the University of Connecticut, May 22nd to 25th, 2017.

Confirmed keynote speakers for 2017 include **Dr. Steven Boker** (UVA) and **Dr. Kenneth A. Bollen** (UNC). Steven Boker will be offering a full day Pre-Conference workshop on Dynamic SEM; and Ken Bollen will be offering a half day Post-Conference workshop on Model implied instrumental variables using MIIVsem. We hope to add a third keynote speaker to the line-up in the coming months. If you have suggestions, feel free to email them to me at betsy@uconn.edu.

We are currently soliciting both methodological research papers and papers that illustrate methodological techniques in the area of modeling, broadly defined. Papers related to multilevel modeling, structural equation modeling, mixture modeling, longitudinal modeling, and item response theory 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 will be 90 minutes in length.

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. Methodological Research proposals should be no longer than 1000 words and should include purpose, background, methods, results, discussion, and significance. Methodological Illustration papers 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.

There are three different types of presentations: Paper sessions (in which authors submit a paper), Symposia (in which a group of authors submit a set of related talks/papers), and posters. Proposals for symposia should include titles, authors, and brief descriptions/abstracts for all of the paper presentations within the symposium. Symposium proposals may be longer than 1000 words if needed, but they should be less than 2000 words. Graduate students are also encouraged to submit proposals, especially poster sessions. All proposals should be submitted electronically.

Proposals for the 2017 conference are due February 1st, 2017. Notifications of presentation status will be emailed by February 18th, 2017. If you have any questions about the conference, please email D. Betsy McCoach at betsy.mccoach@uconn.edu.

Statistical Innovations

Thinking outside the brackets!™

Exciting Job Opportunity

Reply to Latent GOLD® users' questions on latent class modeling:

Work remotely and choose your hours.

Also, assist in high-profile consulting projects*.

Candidate must be well organized, with familiarity of latent class modeling, strong computer, writing and communication skills.

How to Apply: Email your resume and a cover letter to sabrina@statisticalinnovations.com

Latent GOLD® is #1 program for latent class modeling.

How to get familiar with Latent GOLD:

Download the demo and view the Latent GOLD® tutorials available at SI's website: http://www.statisticalinnovations.com/tutorials2/

Attend M3 Session -- Wednesday, May 25 at 10:30 a.m.
Recent Developments in Latent Variable Modeling – Jay Magidson, founder and president of SI will present a demo of Latent GOLD 5.1 as part of this session.

***Statistical Innovations engages in high-profile consulting projects**:

Latent Markov modeling was used for ***U.S. Census Bureau*** to adjust unemployment rate estimates for measurement error.

Jay Magidson developed a new version of the popular Myers-Briggs Type Indicator (MBTI) based on latent class modeling for ***Consulting Psychologists Press***.

Jay Magidson provided latent class training at ***Genentech*** and assisted them on various projects such as modeling long term survivors, and coauthored publications.

For more info see: http://www.statisticalinnovations.com/consultancy/past-clients/