Modern Modeling Methods Conference

May 19–20, 2015

UCONN
UNIVERSITY OF CONNECTICUT

2015 Modern Modeling Methods Conference

University of Connecticut

Sponsored by the ***Educational Psychology Department*** and the ***Neag School of Education***, University of Connecticut

Welcome to the 5th annual Modern Modeling methods conference here at the University of Connecticut. Special thanks go to Dr. Del Siegle, chair of the Educational Psychology department, and Dr. Richard Schwab, Dean of the Neag School of Education for their support.

In addition, many thanks to Joanne Roberge, Cheryl Lowe, Robbin Haboian-Demircan, Steve Rugens, Dani Yomtov, Kate Copeland, and conference services for providing administrative and logistical support for the conference.

Also, thank you to all of the keynote speakers and concurrent presenters for making this wonderful program possible.

Finally, thank you to all of the 2015 Modern Modeling Methods conference attendees for coming and being a part of the fifth annual M^3 conference! I hope to see you all back in Storrs in May 2016 for the sixth annual Modern Modeling Methods conference. Confirmed keynote speakers for 2016 include Dr. Bengt Muthèn (UCLA/Mplus) and Dr. Andrew Gelman (Columbia). Bengt will also offer a full day pre-conference workshop.

Proposals for concurrent sessions will be due February 1st, 2016, and can be submitted online at our website: www.modeling.uconn.edu

D. Betsy McCoach, Ph.D.
2015 Chair, Modern Modeling Methods Conference
Professor and Program Coordinator, Measurement, Evaluation, and Assessment Program, Educational Psychology Department
Neag School of Education, UCONN

Monday, May 18th: Pre-conference Workshop

New Features in LISREL 9

Karl Jöreskog

A major new feature in LISREL 9 is the ability to start with raw data and formulate the model directly with SIMPLIS or LISREL syntax or by using the graphical user interface (GUI). This workshop will follow the principle of learning from examples. Several instructive examples will be demonstrated and discussed and, in particular, the connection between the statistical model and the syntax file will be covered. Participants are encouraged to bring their own laptops with LISREL 9 installed (a free student version is available online) and follow the examples.

9:00 am - 5:00 pm in Laurel Hall 102

The following topics will be emphasized:

- Creating the LISREL system data file (LSF file)
- Exploratory and confirmatory factor analysis
- Estimating latent mean differences
- Estimating latent growth curves from data with missing values
- Latent variable scores
- Multilevel analysis

Karl G. Jöreskog is Professor Emeritus at Uppsala University in Sweden. He received a PhD in Statistics at Uppsala University in 1963. He was a Research Statistician at Educational Testing Service and a Visiting Professor at Princeton University from 1964 to 1971. During these years he published several papers in Psychometrika on the method of maximum likelihood applied to exploratory and confirmatory factor analysis, covariance structure analysis, and multiple group factor analysis. These papers laid the foundation for the LISREL model. In 1971 Jöreskog returned to Sweden to become Professor of Statistics at Uppsala University. In 1984 he was appointed a Research Professor of Multivariate Statistical Analysis, a position he held until his retirement in 2000.

Jöreskog received the Arnberg Prize by the Swedish Royal Academy of Sciences in 1971, the Ubbo Emmius Medal by the University of Groningen, Netherlands in 1983, the ETS Award for Distinguished Service to Measurement by Educational Testing Service in 1987, the Sells Award by the Society of Multivariate Experimental Psychology in 2000, and the Olaus Rudbeck Medal by Uppsala University in 2005.

In 2014 he was awarded Jubilee Doctor by Uppsala University at the occasion of the 50th anniversary of his promotion to PhD. Jöreskog has authored several books and numerous journal articles on factor analysis and its extensions and on structural equation modeling. Together with Dag Sörbom he developed the LISREL computer program.

Tuesday, May 19th

8:00 am – 9:00 am in Laurel Hall Atrium

REGISTRATION AND BREAKFAST

9:00 am – 10:30 am in Laurel Hall 102

Keynote Address

Karl Jöreskog, Ph.D.

50 Years of SEM in 50 Minutes??

SESSION #1 10:45 am – 12:15 pm

Session 1.1 (Laurel Hall 201)

An Empirical Test of Accountability Policy: A-F School Grades.
Mwarumba Mwavita & Curt Adams.

Longitudinal Models for the Early Development of Hand Preferences.
Richard A. Faldowski, George F. Michel, Iryna Babik, Julie Campbell, & Emily Marcinowski

Piecewise-Linear Multilevel Models of Sociometric Nominations over the Transition to Middle School.
Richard A. Faldowski, Madelynnn D. Shell, & Heidi Gazelle

Session 1.2 (Laurel Hall 202)

Bayesian Model Averaging Over Directed Acyclic Graphs with Implications for Prediction in Structural Equation Modeling.
David Kaplan & Chansoon Lee

Bayesian Factor Analysis with Variable Selection Techniques.
Zhaohua Lu, Sy-miin Chow, & Eric Loken

Session 1.3 (Laurel Hall 205)

Unipolar Item Response Models.
Joseph F. Lucke

Cross Classified Modeling of Dual Local Item Dependence.
Chao Xie & Hong Jiao

Session 1.4 (Laurel Hall 206)

An Empirical Comparison of Multiple Imputation Approaches for Treating Missing Data in Propensity Score Analyses.
Jessica Montgomery, Eun Sook Kim, Jeffrey D. Kromrey, Rheta E. Lanehart, Patricia Rodriguez de Gil, Yan Wang, & Reginald Lee

Missing Covariates in Causal Inference Matching: Statistical Modeling Using Machine Learning and Evolutionary Search Algorithms.
Landon Hurley

Session 1.5 (Laurel Hall 301)

Simulating Data for Mixture Model Studies: Considering Measures of Data Overlap
Jeffrey Harring & Junhui Liu

A Simple Simulation Technique for Non-Normal Data with Pre-Specified Kurtosis and Covariance Matrix.
Ulf Henning Olsson & Njål Foldnes

Non-normal Data Simulation Using Regular Vines.
Njål Foldnes & Steffen Grønneberg

Session 1.6 (Laurel Hall 302)

Methodological Illustration of Multiple Group Multilevel SEM with LSA Data.
Agnes Stancel-Piatak

An Evaluation of the Alignment Method for Detecting Measurement Non-invariance in Noncognitive Scales.
Jessica Kay Flake & Betsy McCoach

Exploring Noninvariance in Classroom Behavior Trajectories Using Growth Mixture Modeling.
Janice Kooken, D. Betsy McCoach, Sandra Chafouleas, Faith G. Miller, Megan Welsh, T. Chris Riley-Tilman, & & Noel Card

12:15 pm – 1:30 pm in the Student Union Ballroom (SU 330/331): LUNCH

SESSION #2 1:30 pm – 3:00 pm

Session 2.1- Symposium (Laurel Hall 201)

Innovative Developments and Applications in Latent Class Analysis.
Chair: Jay Magidson

Paper 1: Goodness-of-fit of Multilevel Latent Class Models for Categorical Data.
Erwin Nagelkerke, D. L. Oberski, & Jeroen Vermunt

Paper 2: Micro-macro Multilevel Analysis for Discrete Data.
Margot Sijssens-Bennink, M.A. Croon, & Jeroen Vermunt

Paper 3: Divisive Latent Class Analysis Applied to Social Capital.
Mattis van den Bergh, Verena Schmittmann, & Jeroen Vermunt

Paper 4: Resampling Methods for Assessing Latent Class Model Fit.
Geert van Kolenburg, Joris Mulder, & Jeroen Vermunt

Session 2.2 (Laurel Hall 202)

Assessing Associations and Patterns in Multi-Member Multi-Group Data.
Thomas Ledermann, Myriam Rudaz, & Alexander Grob

Using Multiple Group Modeling to Test Moderators in Meta-Analysis.
Alexander M. Schoemann

Using Moderated Nonlinear Factor Analysis (MNLFA) to Develop a Commensurate Measure of Alcohol Use Across Four Independent Studies.
Jennifer L. Walsh, Lance Weinhardt, Seth Kalichman, & Michael Carey

Session 2.3 (Laurel Hall 205)

Small Sample Robust Model Fit Criteria in Latent Growth Models with Non-Informative Dropout.
Dan McNeish & Jeff Harring

Growth Modeling with Selection and Missing Data: A Shared-parameter Model for Predicting College Readiness with Interim Assessment Results.
Yeow Meng Thum & Tyler Matta

Session 2.4 (Laurel Hall 206)

Comparing Aspects of Data Collection to Improve Statistical Power.
Andrew L. Moskowitz, Jennifer L. Krull, K. Alex Trickey, & Bruce F. Chorpita

Partially Nested Randomized Control Trials in Educational Research: Applications to a Summer Learning Program.
Jonathan Schweig & John Pane

Predicting Group-Level Outcome Variables: An Empirical Comparison of Analysis Strategies.
Jeffrey D. Kromrey & V. Lynn Foster-Johnson

Session 2.5 (Laurel Hall 301)

A Study of Classroom Learning with a Mixed Model for Ordinal Variables and Special Emphasis on Individual Differences.
Robert Cudeck

A Mode of Zero: Strategies for Education Data with Cases of Zero.
Lauren Porter

3:00 pm – 3:15 pm

Break

3:15 pm – 4:45 pm in Laurel Hall 102

Keynote Address

Donald Hedeker, Ph.D.

Modeling Between and Within-Subject Variances Using Mixed Effects Location Scale Models for Intensive Longitudinal Data

5:00 pm – 7:00 pm in the Student Union Ballroom (SU 330/331)

POSTER SESSION AND RECEPTION

Please join us in the Student Union Ballroom to visit the posters. The reception includes appetizers and an open bar.

Wednesday, May 20th

7:30 am – 8:00 am in the Laurel Hall Atrium

BREAKFAST AND REGISTRATION

SESSION #3 8:00 am – 9:00 am

Session 3.1 – Symposium (Laurel Hall 301)

Methods for Analyzing Secondary Outcomes in Public Health Case-Control Studies.
Ofer Harel (Chair), Elizabeth D. Schifano, & Haim Bar

Session 3.2 (Laurel Hall 305)

Introducing N-Level Structural Equations Modeling: Framework, Software and Applications.
Paras D. Mehta

Session 3.3 (Laurel Hall 306)

Comparison of Advanced Methods for Data Imputation in the Context of Item Response Theory: A Monte Carlo Simulation.
Julianne M. Edwards & W. Holmes Finch

Multidimensional Item Calibration and Plausible Value Imputation in Large-Scale Educational Assessments using the Metropolis-Hastings Robbins-Monro Algorithm.
Lauren Harrell & Li Cai

Session 3.4 (Laurel Hall 106)

A Methological Illustration for Regression Mixture Models: Current Issues Of Estimation Problems.
Minjung Kim, Andrea Lamont, & M. Lee Van Horn

Session 3.5 (Laurel Hall 107)

Reconciling Factor-Based and Composite-Based Approaches to Structural Equation Modeling.
Edward E. Rigdon

SESSION #4 9:15 am – 10:45 am

Session 4.1- Symposium (Laurel Hall 301)

The "What", "Why", and "How" of Partial Approximate Measurement Invariance.
Chair: Katherine Masyn & Julia Higdon

Paper 1: Navigating the Full-No Measurement Invariance Passage with Partial Approximate Measurement Invariance.

Paper 2: Partial Approximate Measurement Invariance in Action: Measuring Intergroup Attitudes in Europe.

Session 4.2 (Laurel Hall 106)

Row Fit Derivative Clustering for Heterogeneity Analysis.
Timothy R. Brick

The Method of State Space Mixures.
Michael D. Hunter

Session 4.3 (Laurel Hall 107)

Using Autoregressive Fractional Integrated Moving Average (ARFIMA) Models to Analyze Daily High School Attendance over the Long Term.
Matthijs Koopmans

Mediational Processes in Latent Growth Curve Modeling: Investigation of the Longitudinal Effect of Technology-Based Substance Use Treatment on Drug Abstinence.
Sunny Jung Kim, Lisa A. Marsch, & Haiyi Xie

Analyzing Long-duration and High-frequency Data Using the Time-varying Effect Model.
Haiyi Xie, Robert E. Drake, Sunny Jung Kim, & Gregory J. McHugo

Session 4.4 (Laurel Hall 108)

Detailed Effect Analysis Using Structural Equation Modeling.
Axel Mayer, Lisa Dietzfelbinger, Yves Rosseel, & Rolf Steyer

Structural Equation Models for Comparing Dependent Means and Proportions.
Jason T. Newsom

Session 4.5: Latent Transition Analysis (Laurel Hall 305)

A Guide to the Application of Multilevel Structural Equation Modeling: Bayesian and Frequentist Implementations.
James P. Clifton & Sarah Depaoli

Misspecification of the Random Effect Structure: Implications for the Linear Mixed Model.
Brandon LeBeau

Robust Bayesian Methods in Growth Curve Modeling.
Xin Tong & Zhiyong Zhang

Session 4.6 (Laurel Hall 109)

Comparing the Performance of the Mean- and Variance-Adjusted ML Chi-Square Test Statistic with and without Satterthwaite df Correction.
Jonathan M. Lehrfeld & Heining Cham

Robust Joint Modelling: Questioning the Distributional Assumptions.
Lisa McCrink, Adele Marshall, Karen Cairns, & Damian Fogarty

Two F Approximations to the Distribution of Test Statistics in SEM.
Hao Wu & Johnny Lin

Session 4.7 (Laurel Hall 306)

Errors-in-variables System Identification using Structural Equation Modeling.
David Kreiberg

New Variable Selection Criteria in Model Selection.
Ji hoon Ryoo, Snigdhansu Chatterjee, & Dingjing Shi

Session #5 11:00 am – 12:00 pm

Single Session 5.1 (Laurel Hall 305)

Confirmatory Composite Analysis – Making Structural Equation Modeling Fit for Design Research.
Jörg Henseler

Single Session 5.2 (Laurel Hall 301)

Context Questionnaire Rotation and Imputation with Implications for Estimation of Plausible Values in Large-Scale Assessments.
David Kaplan & Dan Su

Group Session 5.3 (Laurel Hall 106)

Estimating Latent Variable Interactions with Incomplete Exogenous Items.
Heining Cham & Evgeniya Reshetnyak

Performances of Mixture Latent Moderated Structural Equations Approach.
Evgeniya Reshetnyak & Heining Cham

Group Session 5.4 (Laurel Hall 306)

Using a Scale-Adjusted Latent Class Model to Establish Measurement Equivalence in Cross-Cultural Surveys: An Application with the Myers-Briggs Personality Type Indicator (MBTI).
Jay Magidson

Simultaneous Two-Way Fuzzy Clustering of Multiple Correspondence Analysis.
Sunmee Kim & Heungsun Hwang

Session 5.5 (Laurel Hall 107)

Are Indirect Effects Better Captured by Multiple Group Analyses? Benefits of Multiple-Group Structural Modeling in Testing Causal Mediation.
Emil Coman, Judith Fifield, & Monique Davis-Smith

Comparing Recent Advances in Causal Mediation: How Medical Researchers and Practitioners Can Better Understand Causal Mediation and use it for Personalized Medicine.
Emil Coman & Judith Fifield

12:00 pm – 1:00 pm in the Student Union Food Court

Lunch (voucher provided)

Session #6 1:00 pm – 2:30 pm

Session 6.1 - Symposium (Laurel Hall 106)

Applications of Advanced Latent Variable Modeling in the Study of Reading and Motivation.
Chair: Paulina Kulesz

Paper 1: Evaluating the Impact of Common Method Variance in a SEM Model of Reading Comprehension.
Yusra Ahmed, David J. Francis, Jack M. Fletcher, Mary York, & Marcia A. Barnes

Paper 2: Measuring Motivation to Read Through Self-Report: New Insights Through the Application of Polytomous Item Response Models.
Paulina A. Kulesz, David J. Francis, Mary York, and Chris A. Wolters

Paper 3: An Application of Explanatory Item Response Models to Study Developmental Changes in Reading and Visual Processing Skills in Grades K-2.
Shiva Khalaf, Paulina A. Kulesz, Kristi L. Santi, & David J. Francis

Session 6.2 - Symposium (Laurel Hall 305)

Validating Methods for Predicting Individual Treatment Effects.
Chair: Andrea Lamont

Paper 1: An imputation-based approach for predicting individual treatment effects.
Andrea Lamont, Mike Lyons, & Lee Van Horn

Paper 2: A random forest approach to predicting individual treatment effects.
Mike Lyons, Andrea Lamont, & Lee Van Horn

Paper 3: Sample size requirements for imputation-based PITE.
Kathleen Jocoy, Andrea Lamont, & Lee Van Horn

Session 6.3 (Laurel Hall 301)

Because it Might not Make a Big Dif: Improved Differential Test Functioning Measures.
David B. Flora, R. Philip Chalmers, & Alyssa Counsell

Methods for the Comparison of DIF across Assessments.
Holmes Finch, Maria Finch, & Brian French

Permutation Randomization Methods for Testing Measurement Equivalence and Detecting Differential Item Functioning.
Terrence D. Jorgensen, Benjamin A. Kite, Po-Yi Chen, & Stephen D. Short

Session 6.4 (Laurel Hall 107)

Checking Robustness of Longitudinal Results Across Two Types of Gain Scores.
Robert E. Larzelere, Mwarumba Mwavita, Taren Swindel, Ronald Cox, Jr., & Isaac Washburn

Three Steps Toward Improving Causally Relevant Conclusions from Longitudinal Studies.
Robert E. Larzelere, Ronald B. Cox, Jr., & Taren Swindle

Session 6.5 (Laurel Hall 306)

"To Parcel or Not to Parcel" 2.0: Parceling Indicators in Latent Class Models.
Katherine Masyn & Todd Little

A Unified Approach to Functional Principal Component Analysis and Functional Multiple-Set Canonical Correlation.
Ji Yeh Choi & Heungsun Hwang

Feasible Sample Size Determination in Confirmatory Factor Analysis
Jennifer Koran

Session 6.6 (Laurel Hall 108)

Confidence Sets and Exchangeable Weights in Multiple Linear Regression.
Jolynn Pek & R Philip Chalmers

OLS and HCSE Estimation in Linear Models: An Investigation of Non-normality, Heteroscedasticity, and Measurement Error.
Lynn Foster-Johnson & Jeffrey D. Kromrey

2:45 pm – 4:15 pm in Laurel Hall 102

Keynote Address

Thomas Cook, Ph.D.

Systematic Empirical Evidence on Quasi-Experiments That Often Reproduce Causal Estimates From an Experiment Sharing the Same Treatment Group

Thursday, May 21st: Post-conference Workshop

Multilevel Structural Equation Modeling using xxM

Paras Mehta

University of Houston

N-level structural equation modeling is a superset of linear mixed-effects (LME) and structural equation models (SEM). The framework accommodates conventional multilevel models (e.g., HLM, MLM) with random slopes as well as LISREL-like structural equation models for *any* number of levels. A level is defined as any factor with multiple exchangeable units with observed and/or latent variable. With this definition of a level, a SEM model is defined within each level. Observed and latent variables at any level may influence variables at a lower level. A complete NL-SEM model is therefore a directed graph or network of SEM sub-models. The notion of a network of SEM models with influence across sub-models makes the task of specifying complex dependencies with complicated data-structures (e.g., multivariate and longitudinal outcomes with complex cross-classification at multiple levels) rather easy. An R-package called xxM provides an implementation of the NL-SEM framework. xxM (http://xxm.times.uh.edu) is very easy to learn and use. The workshop will provide a practical and hands-on introduction to using xxM for estimating NL-SEM models. xxM uses a LEGO-like approach to building models. In other words, once the user learns how to specify a SEM model for a two-level data-structure, they should be able to specify a model with any number of levels with complex dependency across multiple levels. NL-SEM models have an intuitive graphical representation that is easy to understand. There is a one-to-one correspondence between the graphical model and the xxM script. There are only five simple xxM commands that the user needs to master. The workshop will walk the participants through concrete examples to help them learn the syntax of xxM. More importantly, the examples are designed to help the user to understand how to conceptualize n-Level SEM models with complex data-structures. The datasets, annotated R-script for estimating each model, and annotated output will be made available. The workshop assumes only a basic understanding of multilevel and structural equation models.

8:30 am – 5:00 pm in Laurel Hall 102

Morning topics:

The first part of the workshop focuses on understanding the structure of core multilevel SEM models and the corresponding syntax in xxM:

- Conventional single Level Structural Equation Modeling
- Conventional two and three level/cross-classified multilevel models
- Latent Growth Curve Models with nested data

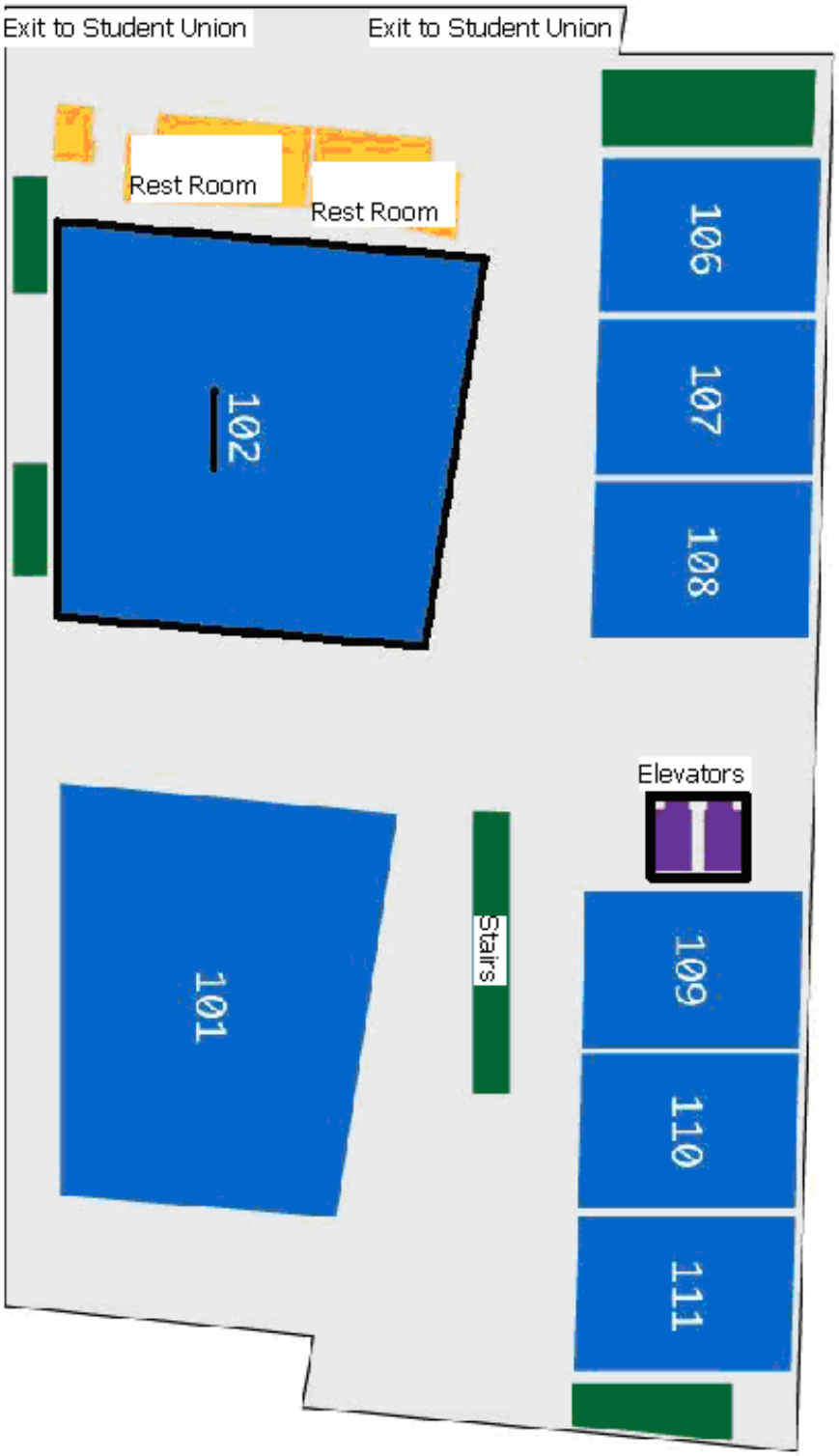
Afternoon topics:

The second part of the workshop focuses on the application of xxM for estimating models with complex dependencies:

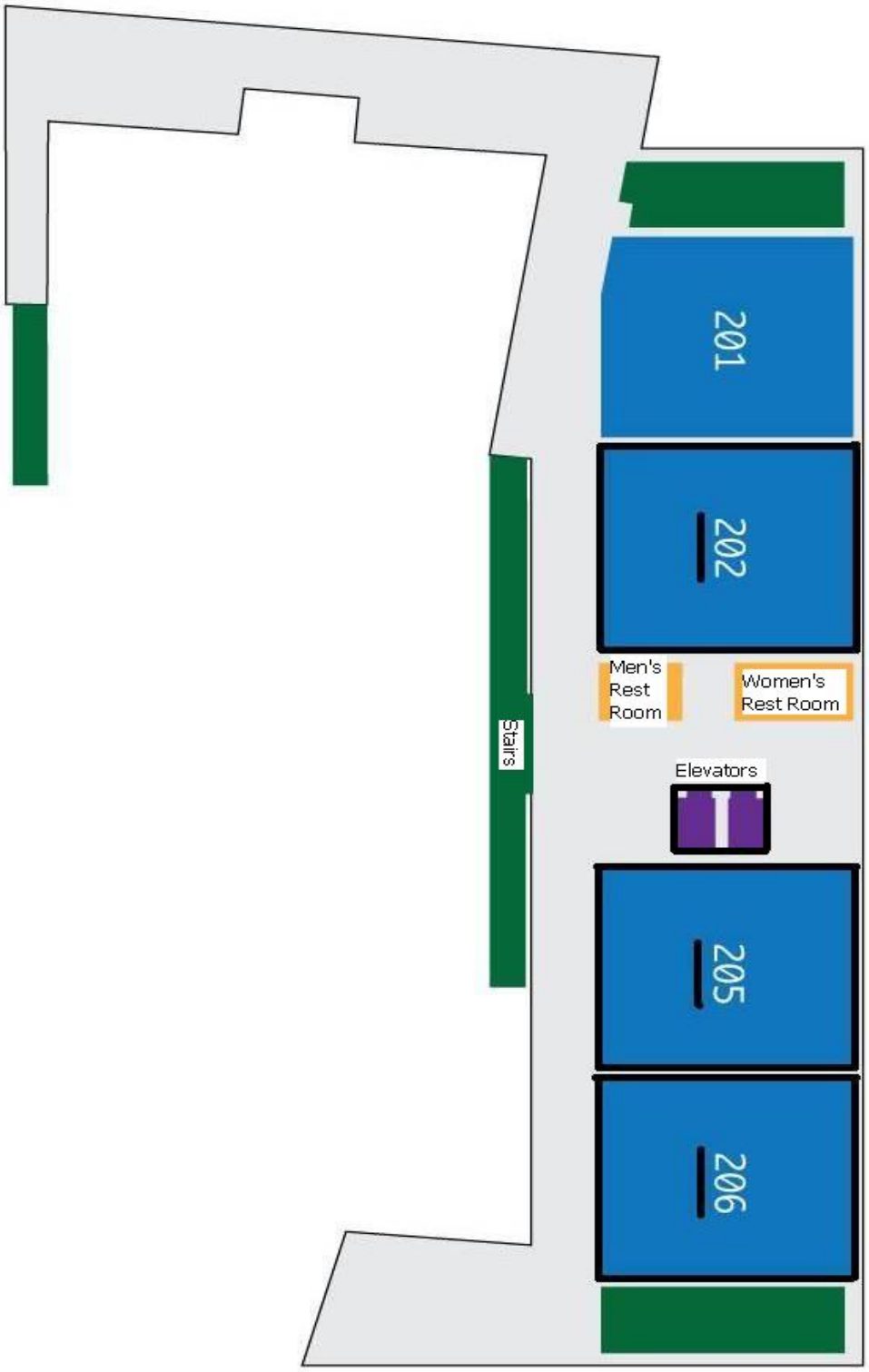
- Longitudinal data with changing classification
- Social Relations Model: Reciprocal ratings obtained in round-robin design

Paras D. Mehta is an associate professor of clinical and industrial/organizational psychology at the Texas Institute for Measurement, Evaluation and Statistics at the University of Houston. His research interests are in multi-level structural equation modeling, growth-curve modeling, and the application of these techniques to educational and organizational research. He is a current member of the Society for Multivariate Experimental Psychology (SMEP).

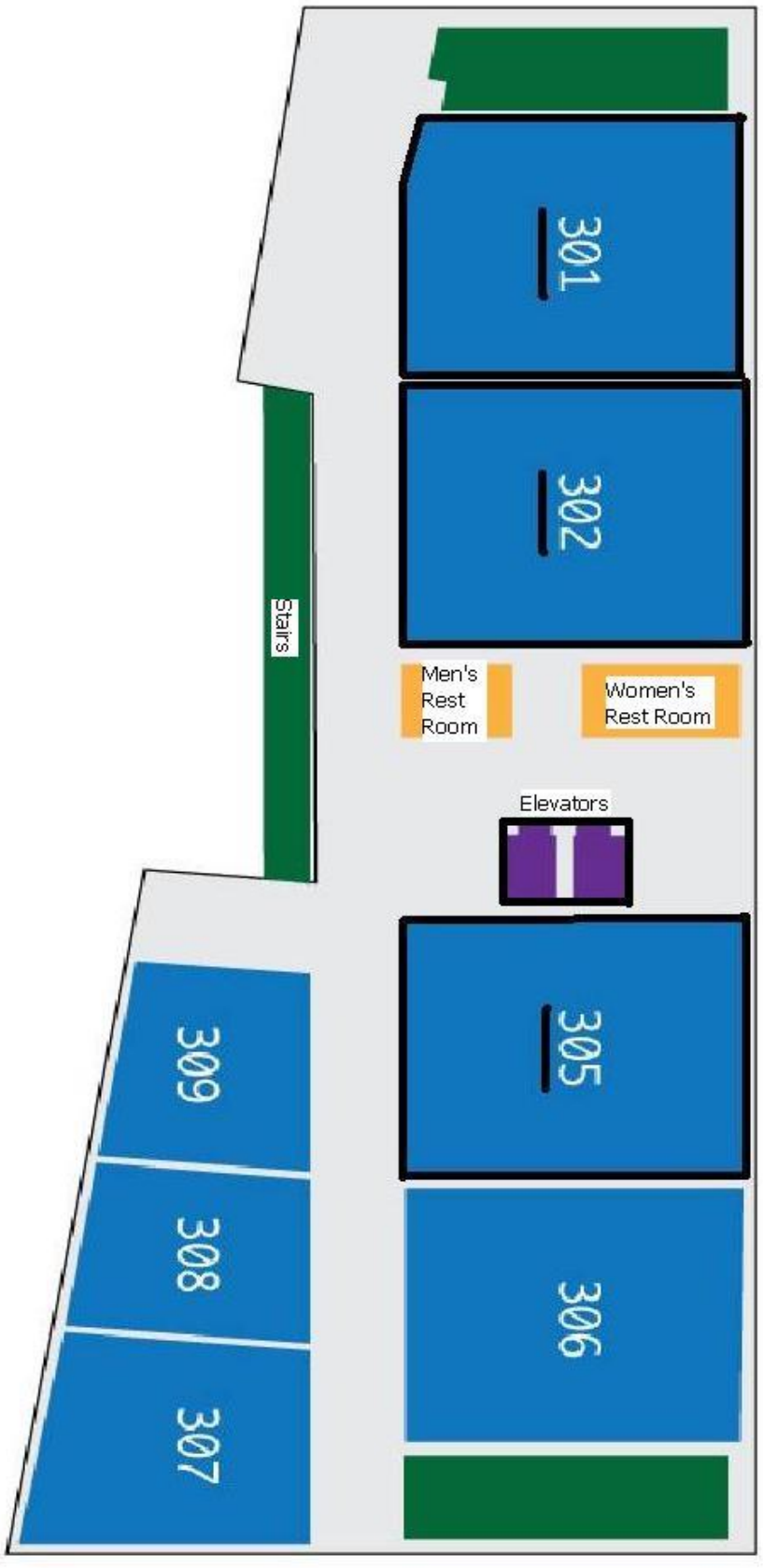
Laurel Hall (Classroom Building) Floor Plans
1st Floor



Laurel Hall (Classroom Building) Floor Plans
2nd Floor



Laurel Hall (Classroom Building) Floor Plans
3rd Floor



Tuesday May 19th - Keynote Address
9:00 am – 10:30 am
Laurel Hall Room 102

Dr. Karl Jöreskog, Uppsala University

50 Years of SEM in 50 Minutes??

Tuesday May 19th - Session 1
10:45 am – 12:15 pm (90 minute sessions)

Session 1.1 (Laurel Hall 201)

An Empirical Test of Accountabilty Policy: A-F School Grades.
Mwarumba Mwavita & Curt Adams.

As states implement new accountability systems under the NCLB legislation, there is growing concern that attention to achievement gaps and the performance of marginalized children has faded. Many approved accountability plans no longer report achievement by student subgroups or include subgroup performance in the calculation of accountability indicators. We evaluated the informational significance of Oklahoma A-F school grades by analyzing reading and math test scores from over 25,000 students in 81 elementary and middle schools. HLM results showed that A-F grades do not differentiate schools by effectiveness levels. Free and Reduced Lunch and minority students average about one standard deviation lower in math and reading than their peers. Test score gaps varied across A-F school grades with the largest gaps existing in "A" and "B" schools. Our evidence suggests that a composite letter grade provides very little meaningful information about achievement differences.

Longitudinal Models for the Early Development of Hand Preferences.
Richard A. Faldowski, George F. Michel, Iryna Babik, Julie Campbell, & Emily Marcinowski

The goal of the current presentation is to address long-standing methodological issues in longitudinal modeling of infant/child hand preferences. It will compare previously proposed 95% confidence interval, latent class analysis, and latent growth trajectory class modeling approaches, highlighting limitations of the methods for understanding the development of hand use preferences. We propose a new solution, which we call a feedforward cumulative model, based on a cumulative longitudinal trinomial logit model. This technique recognizes three primary handedness groups among young children (right, left, and indeterminate), reflects increasing amounts of information over time (children's ages), and decreasing degrees of uncertainty of assignment to hand-use preference groups over time. This modeling approach may be more broadly applicable to problems where evidence is expected to accumulate over time (or repeated assessments).

Tuesday May 19th - Session 1
10:45 am – 12:15 pm (90 minute sessions)

Piecewise-Linear Multilevel Models of Sociometric Nominations Over the Transition to Middle School.

Richard A. Faldowski, Madelynnn D. Shell, & Heidi Gazelle

Longitudinal sociometric nomination data, which is based on counts of the number of classroom peers who think a particular characteristic applies to their classmate, often have distributional properties that violate standard assumptions of linear multilevel longitudinal statistical models. A particularly difficult analytic challenge arises when longitudinal peer nomination data are collected over the transition to middle school. At this discrete transition point, children's nomination trajectories are expected to exhibit abrupt, qualitative changes in their trajectory patterns. In this presentation, we propose multilevel piecewise-linear longitudinal Poisson and negative binomial models as more appropriate alternatives that can easily accommodate the transition point. Using data from a project that tracked the development of socially anxious children over the middle school transition, we illustrate key practical, methodological, and interpretational issues that must be addressed when using the techniques.

Session 1.2 (Laurel Hall 202)

Bayesian Factor Analysis with Variable Selection Techniques.

Zhaohua Lu, Sy-miin Chow, & Eric Loken

In factor analysis, determining the structure of loading matrix is important for modeling the covariance structure of observed indicators and the interpretation of factors. Recently, a Bayesian structural equation modeling (BSEM) approach (Muthén and Asparouhov, 2012) was proposed as a way to explore the presence of cross-loadings in CFA models. We show that the issue of determining factor loading patterns may be formulated as a Bayesian variable selection problem in which Muthén and Asparouhov's approach can be regarded as a BSEM with ridge regression prior (BSEM-RP). We propose an alternative Bayesian structural equation modeling with spike and slab prior (BSEM-SSP). We review the theoretical advantages and disadvantages of the approaches and evaluate the empirical performance of BSEM-SSP, BSEM-RP, forward-backward stepwise modification indices (FBS-MI), and model evaluation criteria such as BIC, DIC, Bayes factor, and Lv measure. A teacher stress scale data set (Byrne, 2012) is used as demonstration.

Bayesian Model Averaging Over Directed Acyclic Graphs with Implications for Prediction in Structural Equation Modeling.

David Kaplan & Chansoon Lee

This paper examines Bayesian model averaging as a means of improving the predictive performance of structural equation models. Parameter uncertainty in Bayesian SEM is handled through the specification of prior distributions on all model parameters. It is recognized, however, that there is uncertainty in the choice of models themselves insofar as a particular model is chosen based on prior knowledge of the problem at hand. This form of uncertainty is presently not accounted for in frequentist or Bayesian SEM. An approach to addressing the problem of model uncertainty lies in the method of Bayesian model averaging. We consider a structural equation model as a special case of a directed acyclic graph. We provide an algorithm that searches the model space for sub-models and obtains a weighted average of the sub-models using posterior model probabilities as weights. Our simulation studies indicate that the model-averaged model provides better posterior predictive performance compared to the estimation of the initially specified model.

Tuesday May 19th - Session 1
10:45 am – 12:15 pm (90 minute sessions)

Session 1.3 (Laurel Hall 205)

Cross Classified Modeling of Dual Local Item Dependence.
Chao Xie & Hong Jiao

Previous studies have mainly focused on investigating one source of local item dependence (LID). However, in some cases, such as scenario-based science assessments, LID might be caused by two possible sources simultaneously. In this study, such kind of LID that is caused by two factors simultaneously is named as dual local item dependence (DLID). This study proposed a cross-classified model to account for DLID. A simulation study was conducted with the primary purpose of evaluating the performance of the proposed cross-classified model. Results of the simulation study indicated that the proposed cross-classified model yielded more accurate parameter recovery, including item difficulty, persons' ability, and random effects' standard deviation. The illustration using the real data generally supported model performance observed in the simulation study.

Unipolar Item Response Models.
Joseph F. Lucke

IRT models typically posit that the latent trait can take on real values, an assumption I call the bipolar trait assumption. Here I propose and develop a unipolar IRT model for which the trait can take on nonnegative values. The unipolar model appears more applicable to clinical and health traits (depression, addiction, pain) that have a lower bound (no depression, no addiction, no pain) than is the standard IRT model. The model is developed from the original Rasch model and extended by the principles of allometric scaling. The model is demonstrated by a Bayesian analysis of a survey on gambling disorders. The Bayesian analysis illustrates the new MCMC program Stan.

Session 1.4 (Laurel Hall 206)

An Empirical Comparison of Multiple Imputation Approaches for Treating Missing Data in Propensity Score Analyses.
Jessica Montgomery, Eun Sook Kim, Jeffrey D. Kromrey, Rheta E. Lanehart, Patricia Rodriguez de Gil, Yan Wang, & Reginald Lee

This study examines the efficacy of four multiple imputation techniques utilized to address missing data in propensity score analyses. Simulated data are used 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), 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.

Missing Covariates in Causal Inference Matching: Statistical Modelling Using Machine Learning and Evolutionary Search Algorithms.
Landon Hurley

Causal inference is the long sought after goal in the social sciences, oft complicated by the 'fundamental problem of causal inference'. Rosenbaum and Rubin (1983) introduced a principaled approach to establishing exchangeability across treatment conditions, and Rosenbaum and Rubin

(1985) later optimised balancing with Mahalanobis distance. These approaches do not maintain unbiasedness in the presence of missing data, and more advanced methodologies, such as multiple imputation, are required to produce unbiased, but not necessarily accurate, causal inference. In this manuscript, we review applied and theoretical issues in handling missing data, and conduct sensitivity analyses on the effects of these different approaches. Our results indicate less than ideal results using traditional techniques, and our recommended approach consists of a two stage methodology: data imputation using machine learning approaches that prioritise accuracy, followed by optimal covariate matching on both propensity score and imputed covariates.

Session 1.5 (Laurel Hall 301)

A Simple Simulation Technique for Non-Normal Data With Pre-Specified Kurtosis and Covariance Matrix.

Ulf Henning Olsson & Njål Foldnes

We investigate a particularly simple way to generate non-normal data using linear combinations of independent variables. The obtained data has a pre-specified covariance matrix and pre-specified kurtosis for each univariate marginal. In contrast to the widely used Vale-Maurelli transform, the obtained data is shown to have a non-Gaussian copula. The simulation method is implemented in the R environment. We demonstrate by Monte Carlo simulations how the normal-theory based likelihood ratio statistic behaves diversely for data obtained by the Vale-Maurelli transform and the simple proposed transform. Algebraic conditions for asymptotic robustness of the normal theory likelihood ratio are established.

Non-normal Data Simulation Using Regular Vines.

Njål Foldnes & Steffen Grønneberg

We propose a new non Gaussian simulation method for covariance models. The method generalizes the so-called *Normal to anything method*, but starts with a a regular vine and not a Gaussian random vector. With this method we can freely choose marginal distributions for observed variables, together with a target covariance matrix. In addition we can choose from a variety of bivariate copula families in the vine structure. If this combination of marginals, covariance matrix and vine structure is feasible, our method calculates the $p(p+1)/2$ copula parameters in the vine structure.This yields a large class of non-normal distributions new to the SEM literature. We present some preliminary simulation studies on the performance of test statistics under this new class, and compare it with the Vale-Maurelli simulation approach. It is shown that our new simulation method generate distributions that differ from the Vale-Maurelli approach in significant ways, although the marginals and covariance matrix can be chosen to be identical.

Simulating Data for Mixture Model Studies: Considering Measures of Data Overlap.

Jeffrey Harring & Junhui Liu

Various studies appearing in the literature of mixture models have concluded that the estimation and classification accuracy of latent variable mixture models are largely affected by how well the data of subgroups are separated from one another. The initial step to conduct a study related to mixture models is to find a measure of mixture distribution separation (or overlap). This research study investigates different statistical indices measuring mean-structure and variance-covariance structure differences and propose a new measure of difference between variance/covariance matrices. A simulation study is conducted to demonstrate how mean structure and variance/covariance structure interactively affect the overlap of mixture distributions.

Tuesday May 19th - Session 1
10:45 am – 12:15 pm (90 minute sessions)

Session 1.6 (Laurel Hall 302)

An Evaluation of the Alignment Method for Detecting Measurement Non-invariance in Noncognitive Scales.

Jessica Kay Flake & Betsy McCoach

In recent years a new methodology, the alignment (Asparouhov & Muthén, 2014), has surfaced for detecting measurement non-invariance (i.e., DIF) across many groups. The purpose of the current study is to investigate the alignment method when testing for measurement invariance of non-cognitive scales across groups of students from different educational contexts (e.g., schools). Asparouhov and Muthén (2014) have investigated the method with continuous and binary item scales. This work extends previous research by using simulation techniques to evaluate the method with polytomous items, which are often used to measure noncognitive constructs. We also evaluate existing research by focusing on evaluating the new tests of non-invariance produced by the alignment than has been seen in previous research.

Exploring Noninvariance in Classroom Behavior Trajectories Using Growth Mixture Modeling.

Janice Kooken, D. Betsy McCoach, Sandra Chafouleas, Faith G. Miller, Megan Welsh, T. Chris Riley-Tilman, & & Noel Card

Growth mixture modeling (GMM) and latent class growth analysis (LCGA) represent person centered modeling techniques that are used to study variations in change over time (Jung & Wickrama,2007). Current best practices for model specification emphasize the importance of the parameterization, number of classes and testing invariance assumptions (Nylund, Asparouhov, & Muthen, 2007). Although variability in the variance of intercepts and slopes is addressed in methodological literature, examples of substantive interpretation of these variances are uncommon in the literature. Methodological research addresses variability in the residual variances (Muthén, 2004), yet this issue is rarely addressed in applied GMM literature. As a case study, changes in student classroom behavior over one year will be analyzed using multilevel GMM. This research will provide baseline descriptive growth models of classroom behavior reflecting the characteristics of heterogeneous groupings and characteristics of change.

Methodological Illustration of Multiple Group Multilevel SEM With LSA Data.

Agnes Stancel-Piatak

Current research within the field of educational effectiveness aims at explaining phenomena related to learning processes. Although these attempts may help to surpass simple input-output correlations, the empirical implementation remains challenging for educational researchers. In large scale educational assessments (e.g. PIRLS, PISA, TIMSS, PIAAC), the complexity of the design and the hierarchical nature of the data require special analysis techniques such as multilevel modeling, weighting, use of plausible values, etc. This presentation demonstrates how to apply Multiple Group Multilevel Structural Equation Modeling (MG-MSEM) with Large Scale Assessment (LSA) data for cross-country comparisons. Using PIRLS 2011 data, the methodological implementation of multilevel SEM is presented. Specific topics related to the correct implementation of the analysis according to the data design are considered.

Lunch, 12:15 pm –1:30 pm
Student Union Ballroom (SU 330/331)

Tuesday May 19th - Session 2
1:30 pm – 3:00 pm (90 minute sessions)

Session 2.1- Symposium (Laurel Hall 201)

Innovative Developments and Applications in Latent Class Analysis.
Chair: Jay Magidson

Latent Class Analysis (LCA) is a well known clustering technique that can be used to classify respondents into (latent) subgroups based on manifest variables. The latent classification can be the main goal of the analysis but can also be used in subsequent analysis. Recent developments showed that LCA can also be used beyond the simple clustering of respondents. In this symposium, four of these innovative developments are presented. In the first talk, divisive latent class analysis is introduced to estimate a hierarchical structure of latent classes and the practical use of this is shown in an empirical application. In the second talk, a very fast resampling scheme to assess model fit is presented since current resampling methods can be very time consuming. Additional complexities arise when model fit is evaluated in multilevel Latent Class Models. Therefore, in the third talk two local fit statistics are proposed, BVRgroup and BVRpair, to improve the understanding of a multilevel Latent Class Model, allow individual testing of the local independence assumptions, and inspect the fit of the higher level of the model. In the final talk of this symposium, multilevel LCA is applied to the field of micro-macro analysis in which group-level variables are explained by individual-level predictors. All research in this symposium was supervised by J.K. Vermunt and can be performed with Latent GOLD 5.0 software.

Paper 1: Divisive Latent Class Analysis Applied to Social Capital
Mattis van den Bergh, Verena Schmittmann, & Jeroen Vermunt

Latent class analysis is used to identify substantively meaningful clusters. This talk introduces divisive latent class analysis to estimate a hierarchical structure of latent classes. This gives a clear insight in the link between retrieved classes and facilitates interpretation. Divisive latent class analysis is a recursive procedure starting with a single class. In each step a class is split in two classes, or not, depending on model selection criteria. The procedure stops when none of the classes needs to be split up anymore. The result is a hierarchical tree of classes. An example with data on social capital from Owen and Videras (2009) shows the practical use of the application.

Paper 2: Resampling Methods for Assessing Latent Class Model Fit
Geert van Kolenburg, Joris Mulder, & Jeroen Vermunt

The assessment of model fit is an important part of statistical analysis. The researchers' interest may lie with specific aspects of a model, or in the global fit. Asymptotic p-values are not available for every conceivable statistic and even when they are available they may not be valid when sample sizes are not very large. To get more reliable p-values, researchers may resort to resampling methods. Some of these methods are time consuming, while others may provide p-values which are not uniform under the null-hypothesis. This talk will illustrate the most common (resampling) methods to test Latent Class model fit. The presentation will discuss a recently proposed calibration the posterior predictive p-value. Finally a very fast resampling scheme is proposed where the statistics are based on data only, which requires that each model of interest is estimated only once.

Paper 3: Goodness-of-fit of Multilevel Latent Class Models for Categorical Data
Erwin Nagelkerke, D. L. Oberski, & Jeroen Vermunt

In the context of multilevel latent class models, the goodness-of-fit depends on multiple aspects, amongst which two local independence assumptions. However, due to a lack of local fit statistics, the issues related to model fit can only be inspected jointly through global fit statistics. This is especially

true for the fit of the higher level of the model, for which there currently are no local fit statistics. As a result the search for model improvements is hindered, since it cannot be determined where misfit originates and which of many possible model adjustments may improve the fit. Also, when solely relying on global fit statistics assumption violations may be obscured and go undetected, leading to wrong substantive results. In our paper, two local fit statistics are proposed to improve the understanding of the model, allow individual testing of the local independence assumptions, and inspect the fit of the higher level of the model. Through an application in which the local fit statistics BVRgroup and BVRpair are used as guidance it is shown that they pinpoint misfit, enhance the search for model improvements, provide substantive insight, and lead to a model with different substantive conclusions, which would likely not have been found when relying on global information criteria. A simulation study shows that the residual statistics have adequate power and Type I error rates.

Paper 4: Micro-macro Multilevel Analysis for Discrete Data
Margot Sijssens-Bennink, M.A. Croon, & Jeroen Vermunt

This talk gives an overview of the use of multilevel Latent Class Analysis in the field of Micro-Macro Analysis, in which a group-level outcome is explained by means of individual- and group-level predictors. Traditionally, the micro-level variable is first aggregated to the macro-level using a manifest group mean or mode and then a group-level analysis is performed in which the aggregated measures are related to the remaining macro-level predictors and outcome. To be able to incorporate measurement error in the aggregated scores, a latent variable is used for the aggregation, and in case of a discrete latent variable, the model is a latent class model. First, this latent class model is presented and compared to the more traditional manifest approaches. Second, it is discussed how to incorporate micro-level predictors that are measured with multiple items. Third, it is explored how to estimate the proposed latent class model in a stepwise manner by separating the latent aggregation from the group-level analysis, as is in fact also done when aggregating with a manifest variable. All developments are illustrated with empirical data examples from different research fields such as personal network analysis, educational measurement, and business analysis.

Session 2.2 (Laurel Hall 202)

Assessing Associations and Patterns in Multi-Member Multi-Group Data.
Thomas Ledermann, Myriam Rudaz, & Alexander Grob

Collecting data from groups, we often have what can be called multi-member multi-group (MMMG) data. Examples are two parents with a child (three members, two groups), same-sex and opposite-sex peers of different status (two members, four groups), or gay, lesbian, and heterosexual couples (two members, three groups). This talk presents a flexible framework enabling the testing of specific hypotheses and patterns in MMMG data by taking group composition into account. We call it the multi-member multi-group Actor-Partner Interdependence Model or MMMG APIM and discuss three models in detail: The three-member two-group (3M2G) APIM for triads representing two different groups, the two-member four-group (2M4G) APIM for dyads representing four different groups, and the two-member three-group (2M3G) APIM for dyads representing three groups. We also propose a general classification scheme for APIM results. Cross-sectional and longitudinal data are used to illustrate the models and computer code is provided.

Tuesday May 19th - Session 2
1:30 pm – 3:00 pm (90 minute sessions)

Using Moderated Nonlinear Factor Analysis (MNLFA) to Develop a Commensurate Measure of Alcohol Use Across Four Independent Studies.

Jennifer L. Walsh, Lance Weinhardt, Seth Kalichman, & Michael Carey

The moderated nonlinear factor analysis (MNLFA) model allows for the development of commensurate measures when studies are combined in integrative data analysis. MNLFA was used to create alcohol use scale scores across 4 studies accounting for differences in both the latent factor and individual items as a function of study, sex, and time. Across studies, 3,752 participants completed 13,200 person-time observations. Alcohol use measures included drinking frequency, heavy episodic drinking, drinks per drinking day, drinks per week, and frequency of intoxication. Using a calibration sample (1 observation per participant) and accounting for differences in factor means and variances, we tested for item intercept and loading differences on an item-by-item basis; combined significant moderators in a final model; and derived scale scores. Items functioned differently across study and sex; however, studies were linked by some items functioning identically. Application of the MNLFA model can aid in pooling data across multiple studies.

Using Multiple Group Modeling to Test Moderators in Meta-Analysis.

Alexander M. Schoemann

Meta-analysis is a popular and flexible analysis that can be fit in many modeling frameworks. Two methods of fitting meta-analyses that are growing in popularity are Structural Equation Modeling (SEM) and Multilevel Modeling (MLM). By using SEM or MLM to fit a meta-analysis, researchers have access to powerful techniques associated with SEM and MLM. This paper details how to use one such technique, multiple group analysis, to test categorical moderators in meta-analysis. Using multiple groups to test for moderators is especially relevant in random-effects meta-analysis where both the mean and the between studies variance of the effect size may be compared across groups. A simulation study and the analysis of a real data set is used to illustrate multiple group modeling with both SEM and MLM. Issues related to multiple group meta-analysis and future directions for research are discussed.

Session 2.3 (Laurel Hall 205)

Growth Modeling with Selection and Missing Data: A Shared-parameter Model for Predicting College Readiness with Interim Assessment Results.

Yeow Meng Thum & Tyler Matta

Joint modeling of longitudinal interim test scores and high-school college entrance examination results, subject to missingness and selection, is presented as a solution for obtaining predictive benchmarks for college readiness on the test score scale of the interim test. We demonstrate the approach with data for multiple student cohorts from four mid-sized to large school districts. An inference framework is developed for evaluating the likelihood that a 5th or 6th grader may be off-track for college and defined and evaluated. We further illustrate how the choice of benchmarks is informed by results of cross-validations, one from model-based predictions and another from a set-aside sample.

Small Sample Robust Model Fit Criteria in Latent Growth Models with Non-Informative Dropout.

Dan McNeish & Jeff Harring

Recent studies have addressed small sample properties of methods for longitudinal data such as mixed effects models and generalized estimating equations; however, no explicit small sample studies have been conducted for latent growth models (LGMs). Consequently, global model fit

criteria, an advantage of LGMs over other methods, have also not been investigated with LGMs and small samples. This study investigates the small sample performance of standard model fit criteria and three small sample corrections is investigated with LGMs and missing data. A missing data adjustment to small sample corrections by Bartlett, Swain, and Yuan is proposed and examined in a simulation with linear growth and latent basis models. The adjusted small sample corrections were found to perform quite well across conditions and improve upon the standard small sample corrections in the presence of missing data.

Session 2.4 (Laurel Hall 206)

Comparing Aspects of Data Collection to Improve Statistical Power.
Andrew L. Moskowitz, Jennifer L. Krull, K. Alex Trickey, & Bruce F. Chorpita

When determining how to collect data, researchers must often find a balance between the number of participants they assess and how often such assessments are made. As researchers strive to find this balance, they are required to consider not only what is methodologically best for their study, but also the ability of their participants to continuously provide quality information. The current study provides researchers with a way to evaluate how changes to an initial study design may impact their power to detect a treatment effect over time. Monte Carlo methods were used to generate random data over a variety of sample sizes, measurement reliabilities, assessment frequencies, intraclass correlations, and effect sizes. Equations were derived to estimate the necessary change in one data collection factor that has a comparable effect on power as another change in study characteristics. A substantive example is provided to illustrate the applications of equations.

Partially Nested Randomized Control Trials in Educational Research: Applications to a Summer Learning Program.
Jonathan Schweig & John Pane

Partially nested randomized controlled trial (PN-RCT) designs are commonly found in educational research, where one experimental arm is clustered, and the other is unclustered. However, they are rarely analyzed with methods that yield valid statistical inferences. One promising analytic method, originally developed to study teacher contributions to student learning, is particularly well-suited to the study of PN-RCTs. We illustrate the utility and flexibility of this novel method on a real world data set, taken from a multi-year, multi-district study of a program designed to decrease summer learning loss. We show that this method is flexible enough to use under real world experimental conditions, and can accommodate cross-overs (students who are assigned to control, but who experience treatment anyway), imperfect information about treatment clusters (i.e., missing data about group membership), and cross-classified random effects.

Predicting Group-Level Outcome Variables: An Empirical Comparison of Analysis Strategies.
Jeffrey D. Kromrey & V. Lynn Foster-Johnson

Methods for analyzing multilevel data with group-level outcome variables were compared in a simulation study. The analytical methods included OLS analyses of group means, a two-step approach suggested by Croon and van Veldhoven (2007), and a Full Information Maximum Likelihood Latent variable technique proposed by Lüdtke et al. (2008). Type I error control, power, bias, standard errors, and RMSE in parameter estimates were compared across design conditions that included number of predictor variables, level of correlation between predictors, level of intraclass correlation, predictor reliability, effect size, and sample size. Results suggested that an OLS analysis of group means, with White's heteroscedasticity adjustment, provided more power for tests of group-level predictors but less power for tests of individual-level predictors. Further, this simple analysis

avoided the extreme bias in parameter estimates and inadmissible solutions that were encountered with other strategies. Results were interpreted in terms of recommended analytical methods for applied researchers.

Session 2.5 (Laurel Hall 301)

A Study of Classroom Learning with a Mixed Model for Ordinal Variables and Special Emphasis on Individual Differences.

Robert Cudeck

A mixed model is applied to data in a study with kindergarten and first grade children who learned to match the musical pitch of a greeting sung by their teacher. Like many age-appropriate academic tasks, pitch matching is not easy for young children and most require several practice sessions to master it. The response variable was a three point ordinal rating scale, scored by the teacher as 0, 1, or 2 for unsuccessful, partly correct, completely correct. There are two kinds of individual learning curves in the repeated measures on this ordinal variable which are informative about the way each child performs. The estimated fixed effects from the model are interesting. The subject-specific results are diverse, complex, and fascinating. One definitely does not want to neglect these provocative and impressive individual differences.

A Mode of Zero: Strategies for Education Data with Cases of Zero.

Lauren Porter

Originating in the field of econometrics, the hurdle model is appropriate for use in count data. The approach is appropriate to model count data with two processes at play: one process that defines a case as a zero of non-zero, the other assessing the degree or dosage once a case has been identified as a non-zero in the model. Hurdle models are primarily seen in economics literature and to a lesser degree in medical and addiction literature. This talk will focus on the potential applications of the hurdle model for use in educational research.

Tuesday May 19th - Keynote Address
3:15 pm – 4:45 pm
Laurel Hall Room 102

Dr. Donald Hedeker, University of Chicago

Modeling Between and Within-Subject Variances Using Mixed Effects Location Scale Models for Intensive Longitudinal Data

Intensive longitudinal data are increasingly encountered in many research areas. For example, ecological momentary assessment and/or experience sampling methods are often used to study subjective experiences within changing environmental contexts. In these studies, up to 30 or 40 observations are usually obtained for each subject over a period of a week or so. Because there are so many measurements per subject, one can characterize a subject's mean and variance and can specify models for both. In this presentation, we focus on an adolescent smoking study using ecological momentary assessment where interest is on characterizing changes in mood variation. We describe how covariates can influence the mood variances and also extend the statistical model by adding a subject-level random effect to the within-subject variance specification. This permits subjects to have influence on the mean, or location, and variability, or (square of the) scale, of their mood responses. Models for both continuous and ordinal outcomes are described and will be illustrated with examples. These mixed-effects location scale models have useful applications in many research areas where interest centers on the joint modeling of the mean and variance structure.

Donald Hedeker, Ph.D., is a Professor of Biostatistics in the Department of Health Studies at the University of Chicago. Previously, from 1993 to 2014, Don was a faculty member of the School of Public Health, University of Illinois at Chicago. He received his Ph.D. in quantitative psychology from the University of Chicago. Don's main expertise is in the development and use of advanced statistical methods for clustered and longitudinal data, with particular emphasis on mixed-effects models. He is the primary author of several freeware computer programs for mixed-effects analysis: MIXREG for normal-theory models, MIXOR for dichotomous and ordinal outcomes, MIXNO for nominal outcomes, and MIXPREG for counts. In 2008, these programs were restructured into the Supermix software program distributed by Scientific Software, Inc. With Robert Gibbons, Don is the author of the text "Longitudinal Data Analysis," published by Wiley in 2006. More recently, Don has developed methods for intensive longitudinal data, resulting in the freeware MIXREGLS program. In 2000, Don was named a Fellow of the American Statistical Association, and he is an Associate Editor for Statistics in Medicine and Journal of Statistical Software.

For more information about Dr. Hedeker, please visit his website: health.bsd.uchicago.edu/People/3129

Tuesday May 19th - Poster Session and Reception
5:00 pm – 7:00 pm
Student Union Ballroom (SU 330/331)

1. **A Rasch Analysis of Math Anxiety Scale across Cultures.**
   Menglin Xu

This study aims to investigate the psychometric property of Math Anxiety Scale using Rasch model. All the sample and data were drawn from PISA 2012 student questionnaire. Data from Finland (n = 5688), Hong Kong (n = 3230), Korea (n = 3355) and USA (n = 3258) were selected to represent western and Asian cultures. Winsteps results showed that, the partial credit model fits the data well in all the cultures, and the items are displaying similar ranking of item measure. A slight mis-matching between person and item distribution was detected in Asian cultures, suggestions of item redundancy, item wording, and differential item functioning across gender was made in each culture.

2. **A Bayesian method to determine intrinsic item response functions in cognitive diagnostic models.**
   Diego A. Luna-Bazaldua & Lawrence DeCarlo

Considering the general framework of the Cognitive Diagnostic Models, a new Bayesian approach is introduced to determine the response function that an item might present when two competing models are compared. For this purpose, several models formerly proposed in the CDM literature (e.g., DINA model, DINO model, and NIDA model) can be reparameterized using the logit function; such reparameterizations can be further extended to test the probability that an item follows a specific response function. Results using simulated data indicate that, when more than one skill is measured by an item, the method correctly identifies the underlying response function between 80 and 90 percent of the time under different simulation conditions. Implications of this methodology are underscored in the context of educational testing, where researchers, policy makers, teachers, and students can obtain more information about the underlying skills measured by the test.

3. **A Comparison of Different Methods of Zero-inflated Data Analysis.**
   Si Yang, Lisa Harlow, Gavino Puggioni, & Colleen Redding

Count data with excessive zeros (i.e., zero-inflation) are prevalent in many disciplines. This study evaluated the performance of several regression models under different conditions of zero-inflation and over-dispersion (i.e., high variability). Model comparison was demonstrated through computer simulation and an empirical data analysis. Akaike Information Criterion values and Vuong tests were used to evaluate relative quality of the regression models. Results showed that the special zero-inflated models performed better than the other models and there was a tendency for the worse models to have smaller standard errors. In conclusion, this study suggests using special zero-inflated models like zero-inflated negative binomial when the data have both excessive zeros, and skewness in the non-zero portion of the data.

4. **A Typology of Cyber-victimization and Traditional Victimization: A 3-Step Latent Profile Analysis.**
   Diana Mindrila, Pamela Davis, & Lori Moore

The study aimed to develop a typology of victimization based on the extent to which students experienced traditional and/or cyber-victimization and some of their psychosocial consequences. The sample consisted of 497 adolescents (ages 12-18) who took the 2011 School Crime Supplement (SCS) of the National Crime Victimization Survey (NCVS) and had at least one cyber-victimization experience. Latent profile analysis (LPA) with the 3-step estimation procedure prosed by Asparouhov & Muthen (2012) was employed, using school behavior management as a covariate and weapon carrying as a distal outcome. LPA yielded three latent profiles: a) Average (N = 441), b) Traditional &

Cyber-victims (*N* = 33), and c) Traditional victims (*N* = 23). As behavior management effectiveness increased, the likelihood of being assigned to groups with higher victimization levels decreased. Further, individuals in the Average group were significantly less likely to carry a weapon than the individuals in the other two victimization groups.

5. Techniques for Small-Sample, Longitudinal Research: Comparing Linear Mixed Effects Models and Generalized Estimating Equations.

Chelsea Muth, Karen L. Bales, Katie Hinde, Nicole Maninger, Sally P. Mendoza, & Emilio Ferrer

Unavoidable sample size issues beset psychological research that involves scarce populations or costly laboratory procedures. When incorporating longitudinal designs these samples are further reduced by traditional modeling techniques, which perform listwise deletion for any instance of missing data. Researchers require sound quantitative methods to work with limited but valuable measures without degrading their datasets. This poster provides a brief tutorial and exploration of two alternative longitudinal modeling techniques, linear mixed effects (LME) models and generalized estimating equations (GEEs), as applied to a repeated measures study (*n* = 12) of pairmate attachment and social stress in primates. Both techniques provide comparable results, but each model offers unique information that can be helpful when deciding the right analytic tool.

6. An Illustration of the Advantages of the Visual Analog Scale.

Pavel Panko, Brittany K. Gorrall, Todd D. Little, Jacob D. Curtis, & Esteban Montenegro

The use of Visual Analog Scale (VAS) has been limited in the past. Despite the contextual advantages of the VAS, the Likert scale is still the most common contemporary response format (Moullec et al., 2011). The VAS requires fewer items to reach the same level of validity and reliability (Davey et al. 2007, de Boer et al. 2004, Williams, Morlock & Feltner, 2010), is more intuitive for select populations (Pautex, 2005) and provides the respondent a truly continuous response format. This poster presents the advantages of the VAS based on extant literature and data. If completed, this demonstration project will present new data from a study that compares participant responses with both the VAS and Likert on the Big-5 personality characteristics as well as on the Depression, Anxiety and Stress Scales (DASS; Hekimoglu, 2012). The results will be analyzed using structural equation modeling. The questions to be addressed are: does the VAS require fewer participants and/or items to be just as reliable and valid as the Likert scale?

7. Application of Multiple-Groups Confirmatory Factor Analysis to Test for Psychometric Measurement Invariance.

Samara L. Rice, Mark A. Prince, Robert C. Schlauch, Joseph F. Lucke, & Gerard J. Connors

The assessment of measurement invariance is a necessary precursor before investigating differences in clinically useful outcomes between groups. Confirmatory Factor Analysis (CFA) was utilized to examine measurement invariance by gender when developing a measure of Felt Ambivalence about reducing drinking. Items measuring Felt Ambivalence were administered to 279 at-risk drinking college students (64.5% female) via an online survey. A subset of five well-performing items were chosen (alpha = .93) and then subjected to a Multiple-Groups CFA. Nested models which allowed factor loadings and/or intercepts to vary by gender were tested using the chi-square difference test. Results supported the invariance of loadings and intercepts (strong factorial invariance). Findings suggest that it is appropriate to use the Felt Ambivalence scale to compare ambivalence levels between female and male at-risk college drinkers. Future research will also assess longitudinal measurement invariance to address the suitability of this instrument for assessing change over time.

Tuesday May 19th - Poster Session and Reception
5:00 pm – 7:00 pm
Student Union Ballroom (SU 330/331)

8. Assessing the Performance of Single Item Longitudinal Models over Varying Conditions.
Ruben Castaneda

Longitudinal item response theory (IRT) has traditionally been used for a series of exams given to multiple groups of people who respond to shared items (i.e. linking). However, longitudinal IRT models may be useful when assessing a person's latent trait over time using a single item. However, optimal conditions for the application of longitudinal IRT models with single items are not well known. This study compares the performance of longitudinal IRT across various conditions and examines possible inferences it may provide. For this, we propose the following conditions: studies featuring 2, 3, 10, and 30 time points and using 50, 100, and 200 subjects. Additionally, we simulate 3 ability levels, one where the ability level is stable and normally distributed and another where the ability level changes across time and a mixture of two abilities. Finally, item discrimination is set to 1 for all items.

9. Autoregressive Latent Growth Modeling: A Bayesian Approach.
Yuzhu Yang & Sarah Depaoli

In this study, we extended the autoregressive latent trajectory (Bollen & Curran, 2004) model to examine the estimation accuracy through simulation of a series of linear and non-linear growth curve models and growth mixture models with autoregression specified between the repeated measures data. We examined the impact of different forms of prior to assess whether estimation of the autoregression components and other parameters would be improved through the use of the Bayesian framework. The main findings indicated that the Bayesian estimation conditions performed better than the frequentist framework in terms of model convergence and estimate bias. Findings of a follow up sensitivity analysis suggested that the accuracy of the parameters estimates was improved when a specific setting of informative priors was placed on both variance-covariance matrices and the residual variances. We conclude this study with an applied example using Bayesian methods to illustrate changes in math achievement throughout elementary school.

10. Characterizing Conceptual Change with Latent Transition Analysis.
Glen Davenport

Research in science education shows that students enter instruction with incorrect, naïve conceptions that act as obstacles to learning. These misconceptions, and the processes that change them into correct conceptions, are identified through vigorous qualitative research. To explore the conceptions and their process of change across populations of students, researchers developed surveys of conceptual knowledge, such as the Force and Motion Conceptual Evaluation (FMCE). Unfortunately, most quantitative studies of conceptual change rely on subscale scores which have low resolution, obscuring the details of the process. Latent Transition Analysis (LTA), however, can identify conceptions at pretest, at posttest, and estimate probabilities of moving from one conception to another across instruction. This study applies LTA to a large sample, n > 2200, of matched FMCE responses and, by using responses as nominal indicators, explores states that exist between the 'common incorrect' and 'correct' conceptions.

Tuesday May 19th - Poster Session and Reception
5:00 pm – 7:00 pm
Student Union Ballroom (SU 330/331)

11. Comparing forensic DNA testing outcomes for biological evidence from stranger versus non-stranger rapes: Results from the Detroit Sexual Assault Kit Action Research Project.

Steven J. Pierce, Dhruv B. Sharma, & Rebecca Campbell

Several US cities have discovered substantial collections of sexual assault kits (SAKs) in police storage that were never submitted for forensic testing, which can yield offender DNA profiles that may be entered into the FBI's Combined DNA Index System (CODIS) and compared to profiles from other crimes. Matches (hits) may reveal or confirm the offender's identity and may also reveal that an offender is a serial rapist. Responding to backlogged SAKs raises policy issues about which SAKs to test when resources are scarce; one debate focuses on the utility of testing SAKs from non-stranger rapes. The testing is a sequential filtering process: SAKs pass on from each stage only if they meet specific criteria. We used a continuation-ratio model to demonstrate that CODIS entry rates, hit rates, and serial sexual assault rates did not differ enough between SAKs from stranger versus non-stranger rapes to justify selection based on victim-offender relationship.

12. Comparing the Gamma Generalized Linear Model, Log Transformation, and Yuen-Welch for Power and Generalizability.

Victoria Ng & Rob Cribbie

Alternatives for analyzing positively skewed and heteroscedastic data include the robust Yuen-Welch, log (power) transformations, and the generalized linear model (GzLM) with gamma distribution and log link. The Yuen-Welch uses a 20% trimmed mean estimator, which allows for adequate power but reduced generalizability; in contrast, transformations and GzLMs allow for the inclusion of all relevant data. Little work has compared power for detecting group differences with these parametric estimators, and certainly less has compared robust approaches to parametric models – particularly when both power and interpretability for population means are of interest. Results suggest that the Yuen-Welch is relatively consistent with power, though it can deviate from population means. When inference to the population mean is desired beyond power, the GzLM is suggested for its unbiasedness and its relative power. Researchers are encouraged to explore parametric alternatives before mechanially applying robust alternatives.

13. Comparing Traditional and Bayesian Approaches for Testing Mean Equivalence.

Alyssa Counsell, Robert Cribbie, & Victoria Ng

Procedures such as the two-one sided tests (TOST) method have been popular for examining mean equivalence, as it is inappropriate to declare means equivalent through nonrejection of the traditional null hypothesis (Ho: $\mu 1 = \mu 2$). This study evaluated two Bayesian methods for equivalence testing of two group means; Kruschke's (2013) Bayesian estimation method compares a high density interval to the equivalence region, whereas Morey and Rouder's (2011) Bayes Factor method compares an interval- based null hypothesis to the complement of this hypothesis. The current simulation study compared the probabilities of concluding equivalence of the procedures across different conditions. Results demonstrated marked differences between the Bayesian and TOST methods, as well as between the two Bayesian methods. Results highlight that the theoretical differences between the methods may lead researchers to different conclusions regarding equivalence, and it is therefore important for researchers to adopt the method that best suits their research goals.

Tuesday May 19th - Poster Session and Reception
5:00 pm – 7:00 pm
Student Union Ballroom (SU 330/331)

14. Detecting Relations Among Dynamic Processes with Two-Occasion Data.
Corinne Henk & Laura Castro-Schilo

Most novel analytic methods for longitudinal data are applicable to studies spanning three time-points of data at a minimum, whereas methods for two-occasion data have garnered comparatively little attention. Here, we address this limitation by introducing the two-wave latent change score (2W-LCS) model, a technique appropriate for preliminary detection of relations among dynamic processes with two-occasion data. The 2W-LCS model is well suited for the investigation of hypotheses in which changes in a construct are posited as predictors of changes in another construct. In an empirical illustration using data of elderly Hispanics from the Health and Retirement Study, we demonstrate how the 2W-LCS model provides the best match to theories rooted in changes, and highlight the advantages of this approach over other modeling alternatives.

15. Determining Necessary Sample Size for Dynamic Group Models.
Wendy Christensen & Jennifer Krull

Dynamic groups models are a powerful tool for psychologists who wish to model longitudinal multilevel data with group-level change over time, but there are no current guidelines for the sample sizes needed to appropriately model these complex effects. Moreover, it is difficult to theorize about the group-level covariance structure ahead of time, making a post hoc model choice necessary (e.g. compound symmetry, autoregressive, ARMA(1,1), etc.). This study examined the effect of number of groups (15 to 300) and group sizes (5 to 100) on the relative biases of the random and fixed effects for both incorrectly and correctly specified models. Specific sample size recommendations for different effects of interest are provided. For post hoc model choice, AIC and BIC were generally able to detect if a complex model was needed, but were likely to recommend a relatively simpler model when the data's structure was actually more complex.

16. Effects of ARMA Processes, Specification of Correlated Error, and Number of Time Points on Latent Growth Model Fit and Parameter Bias: A Monte Carlo Study.
Daniel M. Smith

This Monte Carlo experiment investigates the effects of ARMA processes, sample size, and number of measurement occasions on model fit and bias in parameter estimates across two variations of an unconditional latent growth model (LGM). These LGMs include one in which correlations between adjacent residuals are specified and one in which the nonindependence is ignored. These models are evaluated across varying numbers of measurement occasions, sample size, and first-order AR and/or MA processes. Using a factorial design, 48 unique combinations of the experimental variables were each simulated 200 times. Descriptive and inferential techniques (MANOVA) and interaction plots are used to assess and describe main and interaction effects of experimental conditions on fit and bias. Specifying correlations between adjacent residuals alleviates some bias and maintains excellent model fit, but is not entirely adequate to address nonindependence due to ARMA processes.

17. Estimating Trends in River Water Temperature at Level 3 (year), Given Variable Occasion Designs, Data Sparseness and Potential Confounders at Levels 1 And 2.
Brian Gray

Temporal trend estimates may be biased given temporal variation in potential confounders. Such bias may arise, for example, when estimating long-term trends in water temperature from measurements obtained at haphazard times and dates. We address this concern using a three-level linear model, where linear time effects at the measurement scale are permitted to vary by date and linear date

effects by year. The method is evaluated using Monte Carlo simulations with imbalance, trends in time and date and sparse data. Results indicate unbiased time, date and trend estimators, with variance terms mostly unbiased. We also demonstrate that bias in the trend estimator induced by confounding with trend in time approaches zero with years, and that the trend estimator from an ecologic model, while unbiased, is susceptible to inefficiency induced by collinearity between year and mean annual time.

18. Estimating Interaction Effects in Multilevel Models: A Simulation Study Examining Power and Type I Error.

Julie Lorah

This study examines four methods for investigating moderation within a multilevel framework. Simulated data allow for comparison of power and Type I error among these methods: (1) t-test of the interaction slope coefficient; (2) comparison of a model with and without an interaction term with an F test; (3) comparison of said models with the chi-square difference test; and (4) comparison of said models with Bayesian information criterion (BIC). Further, the effect of varying the number of groups is assessed for both power and Type I error rates. Results indicate that a simple t-test for the interaction slope coefficient may be the preferred method as it shows high power and Type I error rate just lower than the nominal rate of 0.05. Further, regardless of which method is used, the number of groups does not strongly influence the power or Type I error rate.

19. Evaluating Cost-effectiveness of Community Risk Prevention Programs: Illustration of Simulating What Would Happen in other Communities.

Maria Coman, L. Suzanne Suggs, Gisela Rots, & Emil Coman

Cost-effectiveness analyses of preventive interventions implemented in communities is usually assessed by obtaining the cost per-capita for a certain desired improvement, like a specific drop in alcohol use by teenagers. Such calculations however need to be backed by estimating the cost-to-impact relations across several cases, i.e. interventions in distinct communities, such that variation in cost is then linked to variation in effects across communities. Rich two-level models that link both community level and individual level variables (like individual alcohol use obtained by surveying individual youth from all communities) are however hard to come by datawise, so we illustrate here a cost-effectiveness analysis of one alcohol prevention program conducted in a small town in MA, USA, complemented by simulations to get a sense of what would happen cost-wise in other such communities who would decide to implement variants of it, considering their own specific conditions.

20. Evaluation of a Bayesian Approach to Estimating Nonlinear Mixed-Effects Mixture Models.

Sarfaraz Serang, Zhiyong Zhang, Jonathan Helm,, Joel S. Steele, & Kevin J. Grimm

The growth mixture model has become increasingly popular, given the willingness to acknowledge developmental heterogeneity in populations. Typically, linear growth mixture models based on polynomials or piecewise functions are used in substantive applications simulation studies. Growth mixture models that follow inherently nonlinear trajectories, referred to as nonlinear mixed-effects mixture models, have received comparatively little attention, likely due to estimation complexity. This study proposes and evaluates the use of a Bayesian method to estimate these models, and compares them to a linearization method implemented in the SEM framework. Simulations found that although the Bayesian approach produced estimates with slightly lower bias and smaller standard errors, it encountered much more difficulty with convergence and also required more informative prior information. As such, we recommend the use of the SEM approach to derive initial estimates, which can then be refined using the Bayesian approach to provide greater precision in parameter estimation.

21. How to Find Confidence Interval for CFI and RMSEA using Bootstrap.
Xijuan Zhang & Victoria Savalei

Bootstrapping fit indices in SEM is important because most fit indices do not have known analytic distributions. Model-based bootstrap is inappropriate for obtaining confidence intervals for fit indices because it assumes the null is exactly true. In addition, because sample noncentrality is a biased estimate of the population noncentrality, naïve bootstrap is not expected to work well for those fit indices that are based on the chi-square statistic (e.g., RMSEA and CFI). We studied a novel bootstrap approach due to Yuan, Hayashi, and Yanagihara (2007) that seems ideal for bootstrapping fit indices that are based on the chi-square. This method transforms the data so that the "parent" population has the population noncentrality parameter equal to the estimated noncentrality in the original sample. We conducted a simulation study that compared the performance of the Yuan, Hayashi and Yanagihara (2007) bootstrap and the naïve bootstrap for confidence intervals of CFI and RMSEA.

22. In the Eye of the Beholder: Is Perceived Similarity a Product of the Individual or the Dyad?
Sonya M. Stokes, Bobbie A. Dirr, & Paras Mehta

Self-report data has garnered attention from methodologists for decades who warn of the potential issues of such methods. Common method variance, improperly specified "nuisance" variance, and confounded sources of variance are among the problems pointed out. Using perceptions of similarity as a vehicle, we illustrate the use of the Social Relations Model (SRM; Kenny & Albright, 1987) in tandem with a new R-based program, xxM (Mehta, 2013), as an antidote for many of these problems. A univariate, four level SRM model indicates that rater effects (ψ_1,1^4,4=1.22, 95% CI = .73, 1.89) are significant predictors of ratings, while target effects (ψ_1,1^3,3=.13,ns) were not. In addition, dyadic agreement, or the degree to which individual agree about their similarity, was significant, θ_1,1^1,2 = 1.73, 95% CI:.82, 2.63.

23. Is Multivariate Technique Necessary: Estimating the Effects of Moderated Multiple Regression Under Ols Framework.
Dingjing Shi & Ji Hoon Ryoo

This study is to investigate the effectiveness of moderated multiple regression analysis as a strategy to evaluate the hypothesized moderator effects. A moderated multiple regression is referred as a regression model by adding interaction effects to a given multiple regression model (Preacher, 2006). While applying the moderated multiple regression, there exists a concern that such a strategy, in facing of multicollinearity, may reduce statistical power. Therefore, moderated effect is commonly examined with alternative approaches, such as multivariate analysis technique (Tabachnick & Fidell, 2012) or structural equation modeling. However, is the multivariate analysis under maximum likelihood estimation really necessary? Simulation studies were implemented in the proposal to compare statistical power in detecting the true moderated effect under OLS framework with various levels of multicollinearity, and among various sample sizes. The Early Childhood Longitudinal Study Kindergarten were then used to clarify and confirm the results of the simulation study in this proposal.

24. Items that Hang Together may not Change Together: Exploring Dimensions of Change in Sense of Identity.
Thai Ong & Monica Erbacher

Given sense of identity significantly predicts college success (i.e., GPA), it is critical for higher education institutions to explore students' sense of identity development. The present study

investigated the dimensionality of change in responses on the Sense of Identity Scale (Lounsbury & Gibson, 2004) from college students. Exploratory factor analysis criteria supported a two-factor solution, which was validated using CFA on two independent samples. Change factor scores were tested as predictors of cumulative GPA. Scores on the second factor, representing change in values, beliefs, and morals, did not significantly predict GPA. However, scores on the first factor, representing change in sense of self and purpose, significantly interacted with gender to predict GPA, with a near significant main effect. In conclusion, results indicate growth in sense of identity during the first two years of college is multidimensional, and the two change factors found here differentially predict cumulative GPA.

25. Making Causal Inferences Using Education Observational Data with Multilevel Data Structures.
Jose M. Hernandez

Matching-based techniques are typically used for estimating causal effects from observational data when randomization to an intervention or control is not possible. One of the primary questions this study hopes to answer is: What are the implications of ignoring the sample properties of clustered or nested observational data found in education in the context of a matching estimator and estimation of Average Treatment Effects (ATE)? Through MC simulations that mimic real life research scenarios, this study illustrates and provide a set of guidelines that can be used by applied researchers when they choose to conduct a quasi-experiment using PSM methodologies in the presence of complex data. Results indicate that as the variation in the cluster level increases so does the bias in ATE estimation.

26. Parental Harshness and Warmth and Cognitive Outcomes in Hispanic American, African American and European American Families.
Elif Dede Yildirim & Jaipaul L. Roopnarine

Using the data from 5- and 9-year cohorts of the Fragile Families and Child Well Being Study, and latent class growth modeling, we examined how maternal psychological aggression, physical assault and warmth are related to children's cognitive abilities in a sample of 555 European American, 1256 African American, 637 Hispanic American mother-child dyads. Latent class growth modeling with 3 step approach revealed 4 parenting profiles across racial/ethnic groups; high psychological aggression-high physical assault-low warmth (latent class 1), high psychological aggression-high physical assault-high warmth (latent class 2), low psychological aggression-low physical assault-high warmth (latent class 3), low psychological aggression-low physical assault-low warmth (latent class 4). Results suggest that higher maternal psychological aggression and assault are associated with children's lower cognitive skills, with some differences in patterns by racial/ethnic groups.

27. Rasch Model Parameter Estimation via The Elastic Net.
Jon-Paul Paolino

In this study we further investigate the novel approach penalized joint maximum likelihood estimation (PJMLE) developed by Paolino (2013) for estimating the parameters of a Rasch model (Rasch, 1960). Here we use joint maximum likelihood estimation (JMLE) along with elastic net penalization using the glmnet package (Friedman et al., 2010) in R to obtain estimates for item difficulties and examinee abilities. Through simulation we compared the accuracy of PJMLE to conditional maximum likelihood estimation (CMLE), marginal maximum likelihood (MMLE), and marginal Bayes modal estimation (MBME). We show that PJMLE successfully estimates parameters of a Rasch model when the number of items is greater than the number of examinees, which is a shortcoming of traditional estimation techniques. In addition, we further show that PJMLE performs similarly to traditional techniques when

the number of examinees is greater than the number of assessment items without specifying a mixing distribution or a prior distribution.

28. Similar Fit Statistics and Similar Partitions: How Selecting a Suboptimal Solution Impacts the Accuracy of a Mixture Model.

Emilie Shireman

Model selection in mixture modeling is inherently difficult because choosing between a simpler and a more complicated model (models with increasing clusters or covariance parameters) cannot be tested by the typical likelihood ratio test statistic. Researchers typically rely on fit indices like the Bayes Information Criterion. Occasionally, researchers will select a model which is suboptimal on these statistics for several reasons. This poster discusses to what degree the quality of a solution is impacted by selection of a suboptimal solution, and whether large differences in the BIC are indicative of a jump in the quality of a solution. The results of two simulations and a real data demonstration will be presented examining (1) the classification agreement of alternative models and how that relates to differences in the BIC, and (2) the decrease in quality of a mixture model solution that can be expected when a researcher selects a suboptimal solution.

29. The Impact of Linear Dependence on Latent Variable Modeling.

Fraser Bocell

The current study aims to evaluate the impact of the underparameterization of correlated error terms on the estimation of parameters in a 2-factor confirmatory factor analysis (CFA), a popular latent variable model in educational measurement. The underparameterization of correlated error terms has the potential to bias estimates and standard errors. The present study aims to quantify and evaluate any resulting bias in model estimates. This study found the underparameterization of correlated error terms can produce bias in model results. Specifically, loadings within the factor containing a correlated error term were biased, differing according to loading magnitude. However, the bias did not appear to propagate across factors.

30. The New Methods for Detecting Aberrant Behavior In Educational Testing.

Kaiwen Man & Yunbo Ouyang

Response time has been widely considered recently in measuring and evaluating many aspects of a student's latent characteristics, including but not limited to intelligence, personality features, and skill level. By as such, response time can be utilized as a key to identifying aberrances behaviors on educational and psychological tests, which is essential for maintaining the effectiveness of test. By designing a Cheating index, test companies can find the student who is likely to cheat more efficiently and more quickly, which will save a lot of money and energy and keep the exam fair to all the examinees. In this way, the focus of this research is particularly on the identification of aberrant behavior in the test that indicate possible cheating, such as pre- knowledge of some items or attempts to copy answers from others with utilizing the response time.

31. Treatment of Missing Values in Classification and Regression Tree Analyses.

Shu Xu, Natalie Rubinchik, & Michael F. Lorber

Introduction: Classification and Regression Tree (CART) is a regression based analytical tool for classification. CART sequentially conducts binary splits of a sample so that the subsets are homogeneous as possible, with respect to the outcome. This method has been applied in many research areas, and recently, it has been applied in social sciences for prediction, classification, and

even variable selection. Like in many data analyses, missing data are common. Inappropriate treatment of missing values may lead to a loss of valuable information, decreased power in analysis, biased estimates, and invalid statistical inference. However, investigations of appropriate treatment of missing values in tree analysis have been limited. Thus, in this study, we propose an evaluation of the performance of five types of missing value treatments in CART. Methods: The missing value treatments considered in this study are: listwise deletion, mean substitution, multiple imputation under multivariate normal distribution, multiple imputation using chained equation, and surrogate splitting. We use complete data from the Air Force's (AF) Community Assessment survey, and then simulate missing values under missing completely at random (MCAR) and missing at random (MAR) mechanisms. We focus on cases with missing values on any predictors (e.g., types of aggression) rather than the outcome (i.e., cross-cutting patterns of family maltreatment and psychopathology). For each missing value treatment, we assess (1) the accuracy of the imputed values, and (2) the accuracy of prediction using CART. Results and conclusion: The results using the complete data serve as a baseline for comparison. The findings imply that the appropriate treatment of missing values improves the prediction accuracy of CART.

32. Type-I Error Rates and Power of Three Robust Chi-Square Difference Tests When Evaluating Measurement Invariance.

Jordan Brace, Victoria Savalei, & Jenny Chuang

A Monte Carlo simulation investigating power and Type-I error rates for several robust corrections to the Maximum Likelihood chi-square difference test in the context of evaluating measurement invariance was conducted. Studied statistics include the uncorrected difference test, D, Satorra's original computationally intensive correction, DS0, Satorra and Bentler's simplified correction, DSB1, and Satorra and Bentler's strictly positive correction, DSB10. Multiple-group data were generated from confirmatory factor analytic models invariant on some but not all parameters. A series of invariance models were fit to data using the Structural Equation Modelling package lavaan, and model fits were subsequently compared using difference statistics. DSB1 was found to perform well when testing between-group equality of factor loadings, while DSB10 performs well when testing simultaneous between-group equality of loadings and indicator intercepts. This pattern is consistent across a variety of population models. Differences in the relative performance of statistics were most pronounced in small samples.

33. Variable Selection for Propenstity Score Matching on Prognostic Strata.

Jiaqi Zhang & Chrispother M. Swoboda

The propensity score is defined as the conditional probability of assignment to a particular treatment given a vector of observed covariates. As an alternative, the prognostic score has been developed to summarize covariates' association with potential outcomes, also reducing the dimension of covariates. It has been suggested that the combination of PS and PG might improve upon just PS matching. Combining PS with the prognostic score explores a new perspective on matching, especially when there is little overlap of propensity score when the PS model is misspecified or there is little overlap for propensity scores. Even when PSM provides an acceptable matching result, combining the prognostic score could help reduce the overall bias more. Model misspecification has potential to derail the benefits of either matching approach. This study informs researchers on the magnitude of benefit of combining PS and PG across different types of variables and model specifications.

Wednesday May 20th – Session 3
8:00 am – 9:00 am

Session 3.1- Symposium (Laurel Hall 301)

Methods for Analyzing Secondary Outcomes in Public Health Case-Control Studies.
Ofer Harel (Chair), Elizabeth D. Schifano, & Haim Bar

Case-control studies are common in health and public-health research. In these studies, cases are chosen based on the primary outcome but there are usually many other related variables which are collected. While the analysis of the association between the primary outcome and exposure variables is generally the main focus of the study, the association between secondary outcomes and exposure variables may also be of interest. Since the experiment was designed for the analysis of the primary outcome, the analysis of secondary outcomes may suffer from selection bias. In this symposium we will introduce the problem and the potential biased inference that can result from ignoring the sampling design. We will discuss and compare a design-based and model-based approach to account for the bias, and demonstrate the methods using a public health data set.

Session 3.2 (Laurel Hall 305)

Introducing N-Level Structural Equations Modeling: Framework, Software and Applications.
Paras D. Mehta

A general n-Level Structural Equation Modeling (NL-SEM) for complex dependent data is introduced. NL-SEM allows latent variable models with arbitrary number of levels. Each level may include a complete SEM model with observed and latent variables. Regression among observed and latent variables is allowed across any two levels that share a parent-child relationship. In effect, a full NL-SEM model is a network of SEM models. The framework blurs the distinction between a structural equation modeling and mixed-effects model or multilevel modeling. Data with complex dependency structure such as partial nesting, cross-classification, longitudinal data with switching classification, repeated-measures at higher levels, and reciprocal ratings data obtained in round-robin designs (Social Relations Model) can all be specified as NL-SEM models and estimated using xxM (http://xxm.times.uh.edu). A number of empirical examples illustrating key features of the software are presented.

Session 3.3 (Laurel Hall 306)

Comparison of Advanced Methods for Data Imputation in the Context of Item Response Theory: A Monte Carlo Simulation.
Julianne M. Edwards & W. Holmes Finch

Missing data is a common problem faced by psychometricians and measurement professionals. To address this issue, there are multiple techniques that have been proposed to handle missing data in regard to Item Response Theory. These methods include types of data imputation - corrected item mean substitution imputation, response function imputation, multiple imputation, and the EM algorithm, and non-imputation methods - treating the item as not presented, coding missing responses as incorrect, or as fractionally correct. Of these methods, multiple imputation have demonstrated the less biased estimates; however, biased estimates are still present. Due to this, this study's goal is to determine how other imputation methods (random forest imputation (RF), multivariate imputation by chained equations (MICE), a combination of MICE and RF (MICE-RF), and MICE with recursive partitioning (MICE-CART)) compare to multiple imputation.

Wednesday May 20th – Session 3
8:00 am – 9:00 am

Multidimensional Item Calibration and Plausible Value Imputation in Large-Scale Educational Assessments using the Metropolis-Hastings Robbins-Monro Algorithm.

Lauren Harrell & Li Cai

The Metropolis-Hastings Robbins-Monro algorithm is adapted to perform multidimensional IRT calibration simultaneously with latent regression. Using the estimated item parameters, regression parameters, and response data, plausible values can be drawn from the posterior distribution. This method is used to fit complex IRT models, including bifactor and two-tier models, to data from the National Assessment for Educational Progress (NAEP) Science assessment to better reflect the data generating process assumed by the framework. Simulations are generated from two-tier models with background covariates to show the parameter recovery and properties of the models when covariates are included and excluded from item calibration as well as under less complex model formulations.

Session 3.4 (Laurel Hall 106)

Methological Illustration for Regression Mixture Models: Current Issues of Estimation Problems.

Minjung Kim, Andrea Lamont, & M. Lee Van Horn

Regression mixture models are an increasingly popular statistical approach for estimating heterogeneity in effects. These models can be used as an exploratory method to find individual variability in the effects of a predictor on an outcome beyond what can typically be seen when using interactions in regression models. While potentially powerful, these regression mixtures require strong assumptions and are particularly sensitive to model misspecification. This presentation will provide an example to motivate the use of the model and then review the findings from simulation studies examining: (1) non-normality within class error variances, (2) consequences of constraining the class-specific residual variances, (3) consequences of ignoring the relationship between independent variables and latent classes, and (4) model sensitivity to small sample sizes. The presentation will focus on the implication of these findings for those looking to use regression mixtures with applied data and will illustrate the estimation of these models in Mplus.

Session 3.5 (Laurel Hall 107)

Reconciling Factor-Based and Composite-Based Approaches to Structural Equation Modeling.

Edward E. Rigdon

Advocates for factor-based and composite-based approaches to structural equation modeling have been presenting the same arguments (and exchanging the same epithets) for years. As a step toward liberating this conversation, this presentation offers new perspectives on four claims which are part of the pro-factor argument. Do composite methods produce biased parameter estimates? Not if the model is correctly specified. Does the factor model "account for measurement error?" No, observed variable residual variance is only repackaged into factor indeterminacy. Do only factor methods model "latent variables?" No, both approaches create proxies to represent unobserved conceptual variables which should be validated against prior knowledge. Does the factor method's overall fit test make the method "more rigorous?" It does not address the most important question—the validity of the common factors as proxies for the conceptual variables—though it could be used in that way.

Wednesday May 21st - Concurrent Session 4
9:15 am – 10:45 am (90 minutes)

Session 4.1- Symposium (Laurel Hall 301)

The "What", "Why", and "How" of Partial Approximate Measurement Invariance.
Chair: Katherine Masyn

Some degree of measurement invariance (MI) is necessary when one wants to compare latent variable scores across groups. Those who dutifully investigate MI in their data also know how overwhelming it can be to navigate what van de Schoot et al. (2013) aptly describe as the treacherous and narrow passage between full invariance and no invariance. The new approach of Bayesian partial approximate measurement invariance offers a promising, robust vessel for traversing that passage but applied researchers have been left without a rudder or paddle. In this symposium, the first paper presents a principled model-building approach to partial approximate measurement invariance that intersects the established best-practices for investigating partial MI in the maximum likelihood framework with the approximate MI specification in the Bayesian estimation framework. The second paper provides a step-by-step illustration of this approach with data from a large-scale international study of civic knowledge, skills, and attitudes.

Paper 1: Navigating the Full-No Measurement Invariance Passage with Partial Approximate Measurement Invariance
Katherine Masyn & Julia Hidgon

It is well-known that measurement invariance (MI) is necessary when one wants to compare latent variable scores across groups. And the dangers of imposing measurement invariance in a model when items are, in fact, non-invariant, are well-documented. The recently-introduced approach to measurement invariance using Bayesian estimation techniques, termed "approximate measurement invariance" (Muthén & Asparouhov, 2012, 2013), offers a promising enhancement to traditional maximum likelihood-based methods of diagnosing and testing for non-invariance. In this approach, informative prior distributions are used for measurement parameter differences across the groups in place of equality constraints, allowing for small, substantively-negligible differences, in an attempt to strike a balance between model-data fit on the one hand and measurement invariance on the other. In this paper we present a principled model-building approach to partial approximate measurement invariance that intersects the established best-practices for investigating measurement invariance in the maximum likelihood framework (see Millsap, 2011) with the approximate measurement invariance specification in the Bayesian estimation framework. An advantage of our approach is that it allows the researcher to explore areas of partial invariance among subsets of groups, with those subsets varying across item-level parameters. We illustrate the use of both numeric and graphical summaries of non-invariance signals in the data to guide the application of partial invariance constraints, at both the metric and scalar levels, in reaching a well-fitting final partial approximate measurement invariance model. We also demonstrate the use of a novel approach to visualizing the degree to partial approximate invariance across items and groups using social network plots to enable researchers to assess whether there is sufficient invariance to proceed with group comparisons of the latent variable means, variances, and covariances.

Paper 2: Partial Approximate Measurement Invariance in Action: Measuring Intergroup Attitudes in Europe
Julia Hidgon & Katherine Masyn

As part of an ongoing research effort to understand intergroup conflict and improve intergroup relations, we examine intergroup attitudes among adolescents in Europe. In this paper, we demonstrate the use of partial approximate measurement invariance in the Bayesian estimation framework (Muthén & Asparouhov, 2013; van de Schoot et al., 2013) to examine evidence of

measurement invariance, identify areas of non-invariance, and, when appropriate, compare countries on means, variances and covariances of the underlying attitude factors. We use data from the 2009 International Association for the Evaluation of Educational Achievement (IEA) International Civic and Citizenship Education Study (ICCS) (IEA, 2009; Schultz et. al, 2009). Beginning with a well-fitting, seven-group configural invariance model, we follow the process outlined in Paper 1 of this symposium to investigate the degree of measurement non-invariance in the three factors across the countries. We illustrate, step-by-step, how we arrived at our final partial approximate measurement invariance model and discuss the substantive implications of the areas of non-invariance as well as the across-country differences in the underlying constructs.

Session 4.2 (Laurel Hall 106)

Row Fit Derivative Clustering for Heterogeneity Analysis.
Timothy R. Brick

In fitting Structural Equation Models, it is often assumed that the best fit solution is the best fit to the entire data set. Yet in the case where the data consist of distinct clusters, the optimum represents a weighted average of the optimal parameter values for each group. If one were to change a parameter in the direction of one cluster, the likelihoods for the elements of that cluster would improve, while the likelihoods of the other clusters would degrade. This difference in the response of each group to change is characterized by the row-level derivatives of the likelihood fit function for each parameter--the row fit derivatives. By applying standard clustering methods to the computed row fit derivatives, it is possible to detect and explore heterogeneity in an efficient, data-driven manner. In this paper, I present the method of row fit derivative clustering and illustrate its use with examples.

The Method of State Space Mixures.
Michael D. Hunter

Increasingly, psychologists encounter data where several individuals were measured on multiple variables over numerous occasions. Many current methods combine these data, assuming everyone is a randomly equivalent. An extreme alternative is to assume no one randomly equivalent. This presentation proposes a method as a compromise. The goal is to find people that are undergoing similar change processes over time. Data were simulated under various conditions to explore what factors influenced the ability to correctly estimate the change process and find people with the same process. It was found that sample size had the greatest influence on parameter estimation and the dimension of the change process had the greatest impact on correctly grouping people together, likely due to the distinctiveness of their patterns of change. The method, state space mixture modeling, offers one of the best-performing methods for simultaneously drawing conclusions about individual change processes while also analyzing multiple people.

Session 4.3 (Laurel Hall 107)

Analyzing Long-duration and High-frequency Data Using the Time-varying Effect Model.
Haiyi Xie, Robert E. Drake, Sunny Jung Kim, & Gregory J. McHugo

Longitudinal data with a long time series and/or high frequency of measurements has become common in health services research. These data typically exhibit complex and irregular patterns of change, and the relationship between variables may also change over time. Existing longitudinal statistical methods are not flexible enough to capture this dynamic relationship. A new method, the

time-varying effect model (TVEM), uses non-parametric smoothing techniques and permits modeling this complexity. This paper introduces TVEM and illustrates its application by analyzing data from a 16-year longitudinal study of 223 participants with serious mental illness and substance use disorders. We examined the impact of a time-varying predictor, work status on general life satisfaction over 16 years; a static variable, psychiatric diagnosis at baseline, was included as a covariate. The results indicated that the effect of work status on general life satisfaction was positive over most of the time period. But the effect was not constant over time. This analytic approach may help investigators to gain a better understanding of the relationships among variables of interest.

Mediational Processes in Latent Growth Curve Modeling: Investigation of the Longitudinal Effect of Technology-Based Substance Use Treatment on Drug Abstinence.

Sunny Jung Kim, Lisa A. Marsch, & Haiyi Xie

The present study demonstrates the longitudinal effect of technology-based substance use treatment on opioid abstinence and a mediational role of cocaine abstinence in the treatment process. First, intercept and slope modeling was performed to examine whether the intervention successfully enhanced opioid abstinence over a 12-month study period. Second, based on the longitudinal effect of intervention on opioid abstinence, parallel latent growth curve modeling was conducted to understand the mediating role of early cocaine abstinence in opioid abstinence. The estimated mediation effect of early cocaine abstinence on the opioid outcome revealed a significant indirect effect ($z\neg\alpha\beta$ =.062, $\sigma\alpha\beta$ =.03, p = .03). Another mediation path testing the intervention effect on opioid abstinence via the growth trajectory of cocaine abstinence was not significant ($z\neg\alpha\beta$ = .152, $\sigma\alpha\beta$ =.10, p = .12). Overall, the total effects explained by these two parallel mediation paths were significant (Z = 0.214 SE = 0.11, p = 0.049).

Using Autoregressive Fractional Integrated Moving Average (ARFIMA) Models to Analyze Daily High School Attendance over the Long Term.

Matthijs Koopmans

Given that many teenagers drop out of high school before graduating, it is surprising that there is so little empirical work to describe high school attendance patterns over time. The literature typically reports traditional frequencies and measures of central tendency. This presentation describes a more dynamical approach, using Autoregressive Fractional Integrated Moving Average (ARFIMA) modeling to analyze daily high school attendance trajectories over a seven year period in each of five urban high schools. Statistical models are fitted to the attendance trajectories of each school to illustrate the approach and demonstrate its utility. The analysis reveals that short term dependencies, including seasonal cycles are significant in all five schools, whereas long-term dependency (self-organized criticality) is found in only one, suggesting the strain of attending that school and perhaps the vulnerability of the trajectory to exogenous influences (e.g., parental support). These time fluctuations are important, but concealed in traditional summary statistics.

Session 4.4 (Laurel Hall 108)

Detailed Effect Analysis Using Structural Equation Modeling.

Axel Mayer, Lisa Dietzfelbinger, Yves Rosseel, & Rolf Steyer

We present a framework for estimating average and conditional effects of a discrete treatment variable on a continuous outcome variable, conditioning on categorical and continuous covariates. Using the new approach, researchers can consider conditional treatment effects given values of all covariates in the analysis and various aggregates of these conditional treatment effects such as average effects or conditional effects given values of a subset of covariates. Building on an extended

multigroup structural equation model with stochastic group sizes, the approach combines the following strengths: (1) it allows for latent covariates and outcome variables, (2) it permits (higher order) interactions between the treatment variable and categorical and (latent) continuous covariates, and (3) covariates can be treated as stochastic rather than fixed. We show how the new approach can be used to analyze different kinds of average and conditional effects using a comprehensive example.

Structural Equation Models for Comparing Dependent Means and Proportions.
Jason T. Newsom

This presentation will discuss structural equation modeling (SEM) approaches to comparing means and proportions for repeated measures designs. I first discuss how SEM approaches encompass conventional repeated-measures ANOVA for comparing means with continuous variables and McNemar's chi-square for comparing two dependent proportions when single measured variable is used. After an overview of several equivalent methods for comparing means, an extensive discussion of one specification approach provides a novel demonstration how the approach can be used for multiple indicators, multiple occasions, planned contrasts, more complex coding schemes, and trend analyses. This is followed by an overview of extensions for mixed between and within-factorial interaction tests using MIMIC or multigroup analyses and incorporation of time invariant and time-varying covariates. Finally, I discuss the advantages over conventional analysis for comparing means or proportions.

Session 4.5: Latent Transition Analysis (Laurel Hall 305)

A Guide to the Application of Multilevel Structural Equation Modeling: Bayesian and Frequentist Implementations.
James P. Clifton & Sarah Depaoli

Multilevel structural equation modeling (MSEM) is gaining popularity in the social sciences as a framework for estimating latent variable models in the presence of hierarchical data. A number of research papers have been published on technical developments in MSEM; however, there are currently no tutorials on how to properly implement these complex models. Furthermore, applied researchers may be unfamiliar with how to implement a Bayesian estimation approach to MSEM, despite the fact that it has distinct advantages in this modeling context (e.g., Depaoli & Clifton, in press; Hox et al., 2012). To that end, this paper serves as a tutorial on the implementation and application of MSEM using Bayesian and frequentist methods. We illustrate the implementation of MSEM with an application using data from the Program for International Student Assessment (PISA). Findings highlight the benefits of a Bayesian estimation approach to MSEM, and the impact of priors with small samples.

Misspecification of the Random Effect Structure: Implications for the Linear Mixed Model.
Brandon LeBeau

The linear mixed model is a commonly used model for longitudinal or nested data due to its ability to account for the dependency of nested data. Researchers typically rely on the random effects to adequately account for the dependency, however serial correlation can also be used. If the random effect structure is misspecified, can the addition of serial correlation overcome this misspecification and allow for unbiased estimation and accurate inferences? This study explored this with a simulation. Simulation results show that the fixed effects are unbiased, however inflation of the empirical type I error rate occurs when a random effect is missing from the model. Implications for applied researchers are discussed.

Robust Bayesian Methods in Growth Curve Modeling.
Xin Tong & Zhiyong Zhang

Despite wide applications of growth curve models, few studies have dealt with a practical issue -- nonnormality of data. Previous studies have typically used Student's t distributions to remedy the nonnormal problems. In this study, robust distributional growth curve models are proposed from a semiparametric Bayesian perspective, in which intraindividual measurement errors follow unknown random distributions with Dirichlet process priors. Based on Monte Carlo simulations, we evaluate the performance of the robust semiparametric models and compare the semiparametric Bayesian methods to the robust methods using Student's t distributions. We conclude that the semiparametric Bayesian methods are more robust to nonnormal data. An example about the development of mathematical abilities is also provided to illustrate the application of robust Bayesian growth curve models.

Session 4.6 (Laurel Hall 109)

Comparing the Performance of the Mean- and Variance-Adjusted ML Chi-Square Test Statistic with and without Satterthwaite df Correction.
Jonathan M. Lehrfeld & Heining Cham

In structural equation modeling, the chi-square test statistic under normal-theory ML with nonnormal data can result in incorrect results. The purpose of this paper is to compare the performance of two mean- and variance-adjusted maximum likelihood (ML) chi-square test statistics. The first utilizes a Satterthwaite (1941) degrees-of-freedom (df) correction and often results in non-integer values of df. The second was proposed by Asparouhov and Muthén (2010) and retains the integer-valued df of the normal-theory ML chi-square test statistic. We designed a Monte Carlo study based on Curran, West, and Finch (1996) to investigate the performance of the two test statistics. Conditions that vary are whether the model was correctly specified, the underlying data distribution (normal vs. nonnormal vs. severely nonnormal), and five different sample sizes ranging from 100 to 5,000. Results focus on power and Type-I error rates of the test statistics' empirical distributions compared to the reference chi-square distribution.

Robust Joint Modeling: Questioning the Distributional Assumptions.
Lisa McCrink, Adele Marshall, Karen Cairns, & Damian Fogarty

In the joint modeling of longitudinal and survival data it is commonly assumed that the longitudinal random effects and random error terms follow normality assumptions. This research investigates the impact of such conventions when longitudinal outliers are present. By replacement of normality with t-distributional assumptions, robust joint models can accommodate and down weigh the negative effects of outliers on the parameters obtained. The robust joint model presented is flexible as it allows the degree at which outliers are adjusted for to be dictated by the data where the standard joint model, which assumes normality, is given as a special case. Such detrimental effects of outliers are clearly demonstrated within this research both through a simulation study and an illustrative example focusing on Northern Irish renal patient data, with evidence indicating that the bias in the parameter estimates grows with increasing proportions and, in particular, increasing extremity of outliers.

Two F Approximations to the Distribution of Test Statistics in SEM.
Hao Wu & Johnny Lin

When the assumption of multivariate normality is violated in structural equation models or when the sample size is small, the null distribution of the test statistic from maximum Wishart likelihood (MWL)

method may not be chi-square distributed. In addition, if a fitting method other than MWL is used, such as the diagonally weighted least square, the test statistic is not chi-square distributed even with normal data and large sample size. Methods exist to approximate this null distribution by matching moments of this distribution, but are mostly based on scaled or linearly transformed chi-square distributions. We propose two new approximations: a scaled F-distribution and a linearly transformed F-distribution. Simulation studies are used to examine their performance.

Session 4.7 (Laurel Hall 306)

Errors-in-variables System Identification using Structural Equation Modeling.
David Kreiberg

Over the years, Structural Equation Modeling (SEM) has become a well-established statistical technique with numerous uses in the social sciences. Applications of SEM are most often seen in psychological, educational and sociological research. However, due to its versatility, SEM may have usage that go beyond the ones just mentioned. In this study, we demonstrate how SEM can be applied for the purpose of identifying Error-in-variables (EIV) Single-Input Single-Output (SISO) systems. EIV SISO systems belong to a class of time series models that are typically found in engineering applications. Two schemes for how such systems can be formulated as SEMs are presented. The proposed formulations allow for quick implementation using standard SEM software. We additionally show how an approximate efficient weighting matrix can be obtained by applying a Vector Autoregressive (VAR) process to the mean-centered covariances. Monte Carlo examples illustrate the SEM implementing of EIV SISO systems.

New Variable Selection Criteria in Model Selection.
Ji hoon Ryoo, Snigdhansu Chatterjee, & Dingjing Shi

Model selection plays an important role in an exploratory analysis. Nevertheless, the model selection has been considered a very difficult issue because strategies for model building are different across research areas and also dependent on data structures. In spite of complexity and difficulty in the model selection, the procedure of model selection is crucial and necessary because accuracy of model selection directly influences that of parameter estimation. In this project, we narrow down various and wide issues in the model selection to the variable selection criteria and propose new variable selection criteria using the biases and variances in the linear regression model that can easily be applied to structured data such as multilevel and longitudinal data.

Wednesday May 21st - Concurrent Session 5
11:00 am – 12:00 pm (60 minutes)

Single Session 5.1 (Laurel Hall 305)

Confirmatory Composite Analysis – Making Structural Equation Modeling Fit for Design Research.
Jörg Henseler

Causal analysis of meta-analytic data presents unique challenges when compared to the modeling of primary source data. An iterative process that involves the concurrent consideration of primary and secondary data as well as static and longitudinal studies is considered. The process begins with theory-driven models of primary source data and extends to the synthesis of secondary data. Strategies for addressing variance due to third variables and artifacts across studies are suggested. Heterogeneity due to moderator variables requires techniques for differentiating additive and multiplicative effects. The iterative meta-causal analysis was applied to data on the effects of job loss and financial strain on social support and social undermining. The two communication variables partially mediate the effect of these employment-related stressors on self-esteem, depressed affect, and anxiety. The concept of mental health is evaluated as a second-order factor. Trauma is examined as a moderator variable. Several artifacts are identified.

Single Session 5.2 (Laurel Hall 301)

Context Questionnaire Rotation and Imputation with Implications for Estimation of Plausible Values in Large-Scale Assessments.
David Kaplan & Dan Su

This talk concerns two issues of missing data that are relevant to the design and analysis of large-scale surveys. The first issue concerns the desire to expand background questionnaire content through rotation designs. Rotated questionnaires lead to a specific type of missing data problem that can be addressed via various forms of multiple imputation. However, not all methods of imputation lead to the same results and it is important to provide measures of the validity of the imputation. I will discuss imputation of rotated background questionnaires focusing on PISA 2012 with special reference to plausible value estimation. The second issue concerns fusing two different surveys. Data fusion also presents a missing data problem that is similar to that found in rotation designs. I present recent work on data fusion with an application involving the fusion of OCED PISA and TALIS surveys.

Group Session 5.3 (Laurel Hall 106)

Estimating Latent Variable Interactions with Incomplete Exogenous Items.
Heining Cham & Evgeniya Reshetnyak

Since the seminal contribution by Kenny and Judd (1984), different approaches for estimating interaction effects between latent variables have been developed. Although different Monte Carlo simulations studies have been conducted to investigate their performances under different conditions (e.g., interaction models, sample sizes, and reliability, numbers and distributions of observed exogenous items), no study has been conducted to investigate the performances of these two approaches when the observed exogenous items have missing values. The proposal conducts a simulation study to investigate the performances of latent moderated structural equations approach (LMS; Klein & Moosbrugger, 2000) and three variants of the product indicator approach (PI): constrained PI (CPI; Jöreskog & Yang, 1996), generalized appended PI (GAPI; Wall & Amemiya, 2001), and unconstrained PI (UPI; Marsh et al., 2004). The results support that the LMS approach which produce unbiased interaction effect across different missing data conditions and sample sizes.

Wednesday May 21st - Concurrent Session 5
11:00 am – 12:00 pm (60 minutes)

Performances of Mixture Latent Moderated Structural Equations Approach.
Evgeniya Reshetnyak & Heining Cham

Monte Carlo simulation studies in latent variable interactions have urged for the need to develop new approaches that are robust to both nonnormal and missing at random exogenous items (e.g., Cham & Reshetnyak , 2015; Cham, West, Ma, & Aiken, 2012). It is hypothesized that the mixture latent moderated structural equations approach (mixture LMS; Kelava, Nagengast, & Brandt, 2014) can produce correct model results in this condition. The proposed study also attempts to improve this approach by reducing its computational time. The proposed study will conduct a Monte Carlo simulation study to investigate the performances of the LMS approach in various nonnormal exogenous items conditions.

Group Session 5.4 (Laurel Hall 306)

Simultaneous Two-Way Fuzzy Clustering of Multiple Correspondence Analysis.
Sunmee Kim & Heungsun Hwang

Multiple correspondence analysis (MCA) is useful for exploring the associations among dummy-coded categorical variables. We propose to combine MCA with fuzzy k-means in a unified framework to identify common clusters of both respondents and variable categories. The proposed method provides a low-dimensional joint map of variable categories and cluster centroids as well as fuzzy cluster memberships of respondents and variable categories. An application is presented to demonstrate the empirical usefulness of the method. It is shown that the joint map offered by the method is easy to interpret. Moreover, the clustering information on respondents and variable categories greatly contributes to describing clusters in a cleaner manner.

Using a Scale-Adjusted Latent Class Model to Establish Measurement Equivalence in Cross-Cultural Surveys: An Application with the Myers-Briggs Personality Type Indicator (MBTI).
Jay Magidson

The Myers-Briggs personality indicator (MBTI) is one of the most commonly used and well known personality assessments in the world, with the items now being translated/adapted into 17 different languages across more than 20 countries. In this paper we describe how the challenge of establishing measurement equivalence across cultures and different languages was achieved using a scale adjusted latent class (SALC) model. In addition to describing the SALC model, we explain the assumptions underlying worldwide measurement equivalence, and describe the results.

Session 5.5 (Laurel Hall 107)

Are Indirect Effects Better Captured by Multiple Group Analyses? Benefits of Multiple-Group Structural Modeling in Testing Causal Mediation.
Emil Coman, Judith Fifield, & Monique Davis-Smith

We present a 2-group structural model for testing indirect effects that repositions in an intuitive manner the modern causal mediation approaches advocated in the last decade. We unpack the common Tx->M->Y mediation setup as a 2 group M->Y model (for some two Tx=x and Tx=x'). We list potential benefits of the 2-group approach, like more intuitive grasp of the potential concomitant moderating role of the mediator, a natural return to comparing means (intercepts, technically) as opposed to regression coefficients, more meaningful parameters and ability to compare them across treatment groups, along with more flexibility in holding parameters equal across conditions (or not). We compare the 2-group mediation tests to the classic and then contrast to causal mediation options

available currently, as implemented in Mplus, SAS, Stata, and R, using a Diabetes Prevention Program (DPP) translated for implementation in African-American communities. The models match in their finding of no interaction, and in general in the size and significance of the indirect effect, with somewhat larger estimates from the 2-group models than regular 1-group ones. Direct effect conceptualization remains to be illuminated, beyond the mere difference between total and indirect effects. Comparative Effectiveness research (CER) especially can benefit from this approach, as alternative treatments produce distinct 'causal worlds' that need to be compared on multiple levels.

Comparing Recent Advances in Causal Mediation: How Medical Researchers and Practitioners Can Better Understand Causal Mediation and use it for Personalized Medicine.

Emil Coman & Judith Fifield

We review all currently available options for causal mediation testing, with an emphasis on providing a more intuitive view of the Potential Outcomes(PO)/Counter-Factual (CF) perspective of causal modeling. We provide an introduction to PO/CF alternative definitions of indirect and direct effects (Imai, Muthen, & Pearl, Robins, and VanderWeele], including the recent 'unification' of mediation and moderation proposal [VanderWeele]. Judea Pearl's Causal Structural Modeling (CSM) definitions that operate at the 'pre-linear' (nonparameteric) modeling level are also explicated, with examples of how to create counterfactual scenarios from estimated linear causal model parameters. We illustrate the approaches with an applied data from the SisterTalk project, a community-based intervention conducted in African American churches (SisterTalk) aimed at weight loss among women at risk for diabetes, through lifestyle behavioral changes. We provide syntax/code and interpretation of Mplus, SAS, Stata, and R outputs. We compare the software implementation approaches in terms of flexibility and modeling options, and try to explain the differences in estimates; sensitivity analysis options will be also compared. We provide concrete examples of CF questions about how individual participants would have fared under the other condition, instead of the one they were actually assigned to.

Lunch, 12:00 pm –1:00 pm
Student Union Food Court (voucher provided)

Wednesday May 21st - Concurrent Session 6
1:00 pm – 2:30 pm (90 minutes)

Session 6.1 - Symposium (Laurel Hall 106)

Applications of Advanced Latent Variable Modeling in the Study of Reading and Motivation.
Chair: Paulina Kulesz

The symposium focuses on applications of advanced latent variable modeling in the study of reading and motivation. The presentations discuss criterion contamination in a SEM model of reading comprehension, descriptive item response models evaluating functioning of test items measuring motivation to read, and application of explanatory item response models to study developmental changes in reading and visual processing as a function of person and item characteristics.

Paper 1: Evaluating the Impact of Common Method Variance in a SEM Model of Reading Comprehension
Yusra Ahmed, David J. Francis, Jack M. Fletcher, Mary York, & Marcia A. Barnes

Purpose: The goal of this study was to evaluate the influence of common method variance in a model of reading comprehension including five direct and five indirect relations among the following latent variables: background knowledge, vocabulary, word reading, strategies, inference, and reading comprehension. The final model included a general method factor (*g*) that removed variability due to common method. The results indicated that controlling for common method variance in the statistical analyses altered the importance of the various components of the model as predictors of comprehension. Although vocabulary and knowledge were important predictors of comprehension, both directly and indirectly through their effects on inferencing, the magnitude of their direct effect on comprehension was diminished when common method variance was controlled. In contrast, the direct effect of inferencing increased when common method variance was controlled. In both models, inference predicted comprehension over and above vocabulary and knowledge. However, when the measurement model controlled for criterion contamination of the predictors, inference had the largest direct effect on comprehension. Implications for researchers and practitioners include the use of latent variable modeling to mitigate the effects of method bias in the context of reading.

Paper 2: Measuring Motivation to Read Through Self-Report: New Insights Through the Application of Polytomous Item Response Models
Paulina A. Kulesz, David J. Francis, Mary York, and Chris A. Wolters

The goal of the project is to improve understanding of how test items work to measure motivation through the application of two-parameter polytomous item response models, namely graded response, generalized partial credit and rating scale models. The models are contrasted to determine which of them are most appropriate when one is measuring multidimensional motivation construct with response alternatives structured as a rating scale. Three factor analytic multidimensional item response models were used to examine which of the models are most appropriate when one is measuring motivation with a rating scale as a response alternatives format. Contrasted item response models included: (a) graded response model, (b) generalized partial credit model, and (c) rating scale model. The item response models were fitted within three data forms: (a) correlated factors (i.e., a correlated trait model), (b) bifactor with uncorrelated individual-specific factors, and (c) bifactor with correlated individual-specific factors. The three data forms were used to examine how the items and factors relate to each other. The AIC and BIC fit indices indicated that: (a) the graded response model was the best fitting model regardless of data form, and (b) the bifactor data structure with correlated individual-specific factors was best at capturing relations between the items and factors. At the same time, the most stringent in terms of parameters' constrains model (i.e., rating scale model) was the worst fitting model. Together, the results suggest: (a) item thresholds

and factor loadings are not constant across items, and (b) presence of a general factor and individual-specific factors which are correlated with each other.

Paper 3: An Application of Explanatory Item Response Models to Study Developmental Changes in Reading and Visual Processing Skills in Grades K-2

Shiva Khalaf, Paulina A. Kulesz, Kristi L. Santi, & David J. Francis

The current study explores the impact of the development of early reading skills on the visual processing skills of children. Explanatory item response models were used to uniquely study the interplay of task demands, as measured by item features, and student characteristics, as measured by time varying covariates of reading and reading related skills, to understand how the development of reading affects the development of visual processing as measured on standard tests of visual discrimination and visual-motor processing. Nine-hundred and thirty-two first, and second grade students were assessed four times per year on a variety of: (a) reading and reading precursor measures, (b) standardized achievement and intellectual assessments, and (c) measures of visual motor skills and visual discrimination over three years. Explanatory item response models using cross-classified structure with separate random intercepts for people and items were utilized to determine whether variation in item difficulties for test items from the visual processing measures could be attributed to developmental growth in reading ability or due to maturation unrelated to reading as reflected simply by students' age. The cross-classified random effects structure was used to deal with dependencies among the responses to items as these dependencies result from administering all items to all students with all students responding to all items. The results revealed different developmental patterns for five types of visual processing test items, but failed to show consistent effects of learning to read on changes in item difficulty. Findings suggest that visual processing skills are related to the person abilities, and uniquely related to phonological awareness and spelling. However, they do not support the idea that learning to read changes how children process visual information, as no evidence for differential effects of person characteristics over time was found. Instead, the results indicate that visual processing and reading ability improve together over time with no evidence to suggest cross-domain influences. The results imply that any concern among students, parents, and teachers for lack of negative effects of learning to read on the processing of visual information in standard educational assessment seems to be unwarranted.

Session 6.2 - Symposium (Laurel Hall 305)

Validating Methods for Predicting Individual Treatment Effects.

Chair: Andrea Lamont

The effectiveness of an intervention is typically assessed using the Average Treatment Effect (ATE) or some fairly simple version of subgroup analyses (e.g., interactions with sex, ethnicity) resulting from a randomized trial. Interventions that have a favorable ATE are recommended for implementation. The problem with the ATE is that it ignores known individual heterogeneity in treatment response; and, therefore is only useful for tailoring treatments to the extent that the average treatment effects, or simple subgroup-specific effects, reflect the response of the individual. In applied practice, treatment planning would be greatly enhanced if interventionists had access to access to predictions for how well an individual is likely to respond before starting any treatment regimen. Newly developed methods show promise in moving us beyond reliance on group averages by explicitly modeling variability in treatment effects (e.g., Basu, 2014; Freidlin, McShane, Polley, & Korn, 2012; Huang, Gilbert, & Janes, 2012; Imai & Ratkovic, 2013; Zhang, Wang, Nie, & Soon, 2013). At the current time, however, these methods have remained at the level of subgroup analysis (e.g., to search for differential effects in a subpopulation). They have not been extended down to the level of the individual, for the specific purpose of predicting how an individual (who was not involved in the original clinical trial) will respond to treatment. This symposium proposes a novel approach for

predicting an individual's treatment response. Based in a Potential Outcomes framework, the Predicted Individual Treatment Effect (PITE) is a general approach of obtaining predictions for how well an individual will respond to treatment. The PITE is defined for each individual as the difference between the individual's potential outcomes in each treatment arm. The PITE approach builds upon existing methods for estimating heterogeneity in treatment effects, and allow the analyst to derive predicted value for each individual under both treatment and control. These predicted values provide model-based estimates that circumvent the fundamental problem of causal inference (i.e., that realized values for only one treatment arm are typically available), thereby allowing for the estimation of an individual-level treatment effect. In this symposium, we present the initial validation of two algorithms for coming up with predictions. The first paper overviews the PITE approach and its theoretical underpinnings, and tests the performance of an imputation-based estimator to obtain PITEs. The second paper tests the performance of random decision trees to estimate individual-level predictions, and compares the relative performance of the random decision trees to the imputation-based method. The third paper extends the work of the first paper by testing the effects of sample size and number of nuisance variables on the quality of predictions. This symposium relies on a set of Monte Carlo simulation studies to test the performance of two algorithms for developing predictive models. Each paper in this symposium uses a shared dataset and results are compared. Given the novelty of the approach, this symposium focuses on a set of optimal conditions to demonstrate the feasibility of the approach.

Paper 1: An Imputation-Based Approach for Predicting Individual Treatment Effects.
Andrea Lamont, Mike Lyons, & Lee Van Horn

At a basic level, the fundamental problem of causal inference can be conceptualized as a missing data problem – i.e., outcome data is observed from one condition and missing in the counterfactual condition – and handled using modern missing data techniques. In particular, multiple imputation can be used to impute predicted values under treatment and control for each individual, and these predictions can be used to derive individual-level treatment effects. Purpose. In this paper, we test the performance of an imputation-based method for computing predictions in the PITE approach. We specifically examine the accuracy and bias associated with the imputation-based estimator. Methods. We implemented a Baysian approach to imputation using the chained equations algorithm. The PITE was defined as the mean difference between treatment and control across 100 imputed datasets. Results. Results show that the imputation-based estimator does a satisfactory job at recapturing true values (correlation between true treatment effect and the predicted treatment effect is .97) and shows minimal bias across simulations. Discussion. We discuss how this method performed well under ideal conditions and future directions for better understanding and validating the approach.

Paper 2: A Random Forest Approach to Predicting Individual Treatment Effects.
Mike Lyons, Andrea Lamont, & Lee Van Horn

Classification (or machine learning/data mining methods) methods have shown success in estimating individual heterogeneity in data and in obtaining estimates for "out-of-sample" individuals. Specifically, random forests™ are a classification method that uses decision trees to classify individuals and to obtain individual-level predictions. Purpose. The purpose of this paper is to apply random decision trees to estimate predicted individual treatment effects (PITES) and to compare performance to the imputation-based approach presented in paper #1. Methods. We obtained predictions using the RandomForest package in R software. Forests were grown from 5,000 cases were used as a training set and the remaining 5,000 cases were used to estimate PITEs. Results. Results show that random forests are one promising approach to estimate PITEs; however the random forest approach may show more bias than the imputation-based approach under certain conditions. Discussion. A unique advantage to random forests is their ability to handle higher-order interactions. We will discuss how random forests may outperform the imputation-based approach

under less-than-ideal conditions (e.g., a scenario where treatment effects are defined by higher-order interactions) and suggest this as an area of future research.

Paper 3: Sample Size Requirements for Imputation-Based PITE.
Kathleen Jocoy, Andrea Lamont, & Lee Van Horn

The first two papers in this symposium test the performance of the PITE approach under optimal conditions; in essence, they demonstrate the viability of the method and lay foundational work for moving forward in testing the limits of this approach. This paper is the first paper to test performance under more realistic conditions. Purpose. The purpose of this paper is to explore how nuisance variables and sample size impact the precision of estimated predicted individual treatment effects (PITES) for both in-sample and out-of-sample individuals. Methods. This paper utilized generated data with varying numbers of nuisance variables (0, 68, and 134) and sample sizes (250, 1000, and 5000). PITE accuracy was measured as the absolute value of the difference between the PITE and the true treatment effect. Differences between in-bag (generated individuals who were part of the training sample) and out-of-bag (generated individuals who were not part of the training sample) samples were explored. Results. Results suggest that increased nuisance variables decrease the precision of PITE estimates and increased sample size improves the precision of PITE estimates. Discussion. The findings of this study are discussed in terms of next steps for methodological testing.

Session 6.3 (Laurel Hall 301)

Because it Might not Make a Big DIF: Improved Differential Test Functioning Measures.
David B. Flora, R. Philip Chalmers, & Alyssa Counsell

Differential test functioning (DTF) occurs when one or more items in a test demonstrate differential item functioning (DIF) and this DIF accumulates to bias test-level scoring. If DIF effects are weak or the DIF in one set of items is opposite to the DIF in other items, then DTF may be negligible. This presentation introduces new IRT-based DTF measures that account for sampling variability in item parameter estimates. When DIF items systematically favor one group over another, these DTF statistics are large, reflecting the accumulation of DIF across items. Yet, when there is bidirectional DIF such that some items favor one group while other items favor the other, the DTF statistics correctly indicate that the overall DTF effect is negligible. These novel DTF measures can be effectively applied to determine the extent to which individual DIF effects accrue to cause biased scoring for a test as a whole.

Methods for the Comparison of DIF Across Assessments.
Holmes Finch, Maria Finch, & Brian French

Testing practitioners must determine the best assessment to use with their particular population. This study represents some of the first research comparing differential item functioning across tests to aid in this purpose. Four DIF effect size indices are evaluated for three cognitive measures administered to the same participants.

Permutation Randomization Methods for Testing Measurement Equivalence and Detecting Differential Item Functioning.
Terrence D. Jorgensen, Benjamin A. Kite, Po-Yi Chen, & Stephen D. Short

We propose a permutation method for testing the omnibus H0 of measurement invariance using alternative fit indices (AFIs), as well as for testing differential item functioning (DIF) using a single reference distribution of the maximum DIF observable under H0 (conceptually similar to the q distribution used in Tukey's HSD post hoc test). We present our proposed method and evaluate it

using a simulation study. Results show adequate control of Type I errors for the omnibus test, with RMSEA having slightly higher than nominal levels, and CFI, SRMR, Mc, AIC, and BIC having slightly lower than nominal levels. DIF tests, however, showed Type I error rates close to 0% and power < 50% regardless of α level. We discuss how to improve power of DIF tests with several items and multiple groups, which we validate in a larger simulation study comparing permutation to commonly used cutoffs for ΔCFI and modification indices.

Session 6.4 (Laurel Hall 107)

Checking Robustness of Longitudinal Results Across Two Types of Gain Scores.
Robert E. Larzelere, Mwarumba Mwavita, Taren Swindel, Ronald Cox, Jr., & Isaac Washburn

In response to recent calls for replications of results across alternative analyses, this paper illustrates how to check whether results from analyses of residualized gain scores replicate for analyses of simple gain scores and vice versa. Since Lord's (1967) paradox, social scientists have known that results can vary across these two types of gain scores, but social scientists usually analyze and/or report analyses of only one type of gain score. This paper shows how to use estimated correlations to estimate the causally relevant coefficients for both types of gain scores, illustrated from research on corrective actions for problems detected in children (disciplinary punishments, homework assistance, psychotherapy, and Ritalin). Checking for robust results across both types of gain scores would reduce the tendency to make over-generalized causal conclusions from tiny effects in longitudinal analyses, a correction that is needed for some substantive topics.

Three Steps Toward Improving Causally Relevant Conclusions from Longitudinal Studies.
Robert E. Larzelere, Ronald B. Cox, Jr., & Taren Swindle

This presentation describes three ways to improve causally relevant conclusions from longitudinal studies, based on a March 2015 article on critical replications (Larzelere, Cox, & Swindle, 2015). First, researchers need to recognize that almost all statistical adjustments for confounding variables are fallible and imperfect. Second, longitudinal analyses could yield stronger causal evidence if the referent time period for the purported cause clearly came after the referent time for Wave-1 scores on the outcome variable. Finally, conclusions need to recognize that individual causal effects vary around the average treatment effect (ATE), which can include some individual causal effects in the opposite direction from the ATE when the ATE is small, even if significant.

Session 6.5 (Laurel Hall 306)

A Unified Approach to Functional Principal Component Analysis and Functional Multiple-Set Canonical Correlation.
Ji Yeh Choi & Heungsun Hwang

Functional principal component analysis (FPCA) and functional multiple-set canonical correlation analysis (FMCCA) are data reduction techniques for functional data that are collected in the form of curves varying over a continuum (e.g., time or space). In FPCA, components are extracted from a single functional dataset to explain the variance of the dataset as much as possible, whereas in FMCCA, components are obtained from each of multiple functional datasets such that the associations among the components are maximized across the datasets. We propose a unified approach to FPCA and FMCCA, which includes the two techniques as special cases. Moreover, this approach permits a compromise between the techniques in such a way that components are extracted to maximize their associations across different functional datasets, while accounting for the variance of the datasets well. We propose a single optimization criterion for the proposed approach

and develop an alternating regularized least squares algorithm to minimize the criterion. The usefulness of the approach is illustrated by analyzing multiple-subject functional magnetic resonance imaging data.

"To Parcel or Not to Parcel" 2.0: Parceling Indicators in Latent Class Models.
Katherine Masyn & Todd Little

Direct applications of mixture models are typically motivated by the desire to capture individual heterogeneity in a multidimensional phenomenon space. In these applications, it often happens that there are varying numbers of indicators with varying levels of reliability representing the various dimensions of the phenomenon space. Researchers, following the well-established "more-is-better" principle of measurement, include all substantively-relevant items as latent class indicators in the hopes of achieving an optimally-inclusive representation of said phenomenon. And, in doing so, create conditions likely to violate the standard local independence assumption of latent class analysis. In this presentation, we explore the potential of dimension-representative item parceling to mitigate the threats to model validity when local dependence of the items is present. Using simulations, we evaluate this potential in comparison to other techniques that have been applied to remedy local dependence of indicators, namely, log-linear residual associations and factor mixture models.

Feasible Sample Size Determination in Confirmatory Factor Analysis
Jennifer Koran

Many factors have been shown to affect minimum sample size requirements in latent variable models, with most research investigating confirmatory factor analysis models using maximum likelihood estimation. Adding further complication, many criteria defining adequate minimum sample size have been applied in the sample size determination literature. However, no studies to date have applied more than just a few of these criteria, and existing recommendations often lack precision due to the use of factorial designs with fixed sample sizes. This paper provides minimum sample size recommendations that simultaneously meet six criteria for three factor CFA models using maximum likelihood estimation. The number of indicators per factor and the magnitude of the homogeneous loadings were varied. The initial sample sizes investigated were selected based on information from past studies and then adjusted until the recommended minimum sample size to meet six criteria simultaneously was found to within an interval of 25 observations.

Session 6.6 (Laurel Hall 108)

Confidence Sets and Exchangeable Weights in Multiple Linear Regression.
Jolynn Pek & R. Philip Chalmers

When statistical models are employed to provide a parsimonious description of empirical relationships, the extent to which strong conclusions can be drawn rests on quantifying the uncertainty present in parameter estimates. In multiple linear regression (MLR), regression weights carry two distinct kinds of uncertainty represented by confidence sets (CSs) and exchangeable weights (EWs). Confidence sets quantify uncertainty in estimation whereas the set of EWs quantify uncertainty in the substantive interpretation of regression weights. As CSs and EWs share certain commonalities, we clarify the relationship between these two distinct kinds of uncertainty about regression weights. First, the estimation of CSs and the set of EWs for regression weights are reviewed. Then the analytical relationship between CSs and sets of EWs is established. We illustrate the usefulness of CSs and EWs for drawing strong scientific conclusions with an empirical example, and discuss the importance of considering both CSs and EWs as part of the scientific process.

Wednesday May 21st - Concurrent Session 6
1:00 pm – 2:30 pm (90 minutes)

OLS and HCSE Estimation in Linear Models: An Investigation of Non-normality, Heteroscedasticity, and Measurement Error.

Lynn Foster-Johnson & Jeffrey D. Kromrey

The actor-partner interdependence model (APIM) has been widely used for the analysis of pairs of individuals who interact with each other. The goal of this presentation is to detail in a non-technical way how the APIM for binary or count outcomes can be implemented and how actor and partner effects can be estimated using generalized estimating equations (GEE) methodology. Both SPSS- and SAS-syntax needed to estimate the model and the interpretation of the output are illustrated using data from a study exploring the effect of satisfaction with the relationship before the break-up on unwanted pursuit behavior after the break-up in formerly married partners. The use of this GEE method will allow researchers to test a wide array of research hypotheses.

Wednesday May 21st – Closing Keynote Address
2:45 pm – 4:15 pm
Laurel Hall Room 102

Dr. Thomas Cook, Northwestern University

Systematic Empirical Evidence on Quasi-Experiments That Often Reproduce Causal Estimates from an Experiment Sharing the Same Treatment Group

This presentation summarizes an ongoing line of work in which estimates from an experiment are compared to various quasi-experimental design and analytic practices where the treatment group is shared with the experiment but the way of forming the non-equivalent group obviously is not. Work of this kind on regression discontinuity (RD) and comparative RD is summarized in terms of bias reduction and precision both at and away from the RD cutoff. Also summarized is work on interrupted time series (ITS) designs and comparative ITS designs. But most attention is paid to simpler non-equivalent control group designs to illustrate practices in this area that reproduce experimental estimates. Included here is work on various ways of selecting intact but non-equivalent comparison groups and work on various ways of selecting covariates to control for any selection that remains after non-equivalent comparison groups have been chosen.

Thomas Cook is interested in social science research methodology, program evaluation, school reform, and contextual factors that influence adolescent development, particularly for urban minorities. Cook has written or edited 10 books and published numerous articles and book chapters. He received the Myrdal Prize for Science from the Evaluation Research Society in 1982, the Donald Campbell Prize for Innovative Methodology from the Policy Sciences Organization in 1988, the Distinguished Scientist Award of Division 5 of the American Psychological Association in 1997, and the Sells Award for Lifetime Achievement, Society of Multivariate Experimental Psychology in 2008, and the Rossi Award from the Association for Public Policy Analysis and Management in 2012. Cook was chair of the board of the Russell Sage Foundation from 2006 to 2008. He was elected to the American Academy of Arts and Sciences in 2000 and was inducted as the Margaret Mead Fellow of the American Academy of Political and Social Science in 2003. He was part of the congressionally appointed committee evaluating Title I (No Child Left Behind) from 2006 to 2008.

For more information about Dr. Cook, please visit his website:
www.ipr.northwestern.edu/faculty-experts/fellows/cook-t.html

Modern Modeling Methods Conference
May 24-25 2016
University of Connecticut

Call for Papers - Proposals due February 1st, 2016

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 23rd to 26th, 2016.

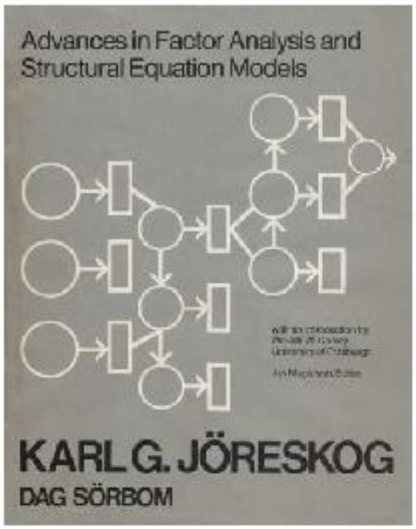
Keynote speakers for M^3 2016 include **Andrew Gelman** and **Bengt Muthén**. Bengt will also offer a full day pre-conference workshop on Mplus on May 23rd.

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.

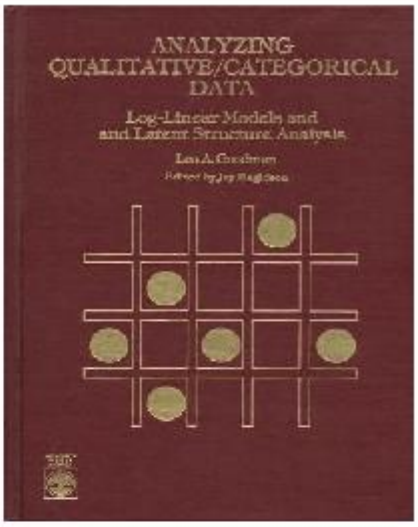
Proposals for the 2016 conference are due February 1st, 2016. Notifications of presentation status will be emailed by February 18th, 2016. If you have any questions about the conference, please email D. Betsy McCoach at betsy.mccoach@uconn.edu.



Statistical Innovations

Thinking outside the brackets!™

Latent GOLD®



The ground-breaking books from Leo A. Goodman (1978) and Karl G. Jöreskog (1979) inspired the editor of these books, Jay Magidson, to found Statistical Innovations in 1981. Statistical Innovations (SI) offers innovative software, training, and consulting.

SI's Latent GOLD® program is today's #1 package for latent class modeling. It has both a point and click menu (GUI), and a powerful and flexible syntax language to estimate latent class, latent trait, and extended models based on variables of different scale types including continuous, ordinal, nominal, and count. The various types of models include:

Latent GOLD®

co-developed with Jeroen K. Vermunt

- Latent Class (LC) Cluster models
- Discrete Factor (DFactor) models
- Latent Class Regression models
- Latent (hidden) Markov models
- Multilevel models
- Step3 Latent Class models

Flexible annual and perpetual licensing options, and special student rates are available.

To learn about the latest new features in Latent GOLD ® attend our talk/session:

- **Innovative Developments and Applications in Latent Class Analysis (May 19 1.30 pm)**
- **Using a Scale-Adjusted Latent Class Model to Establish Measurement Equivalence in Cross-Cultural Surveys: An Application with the Myers-Briggs Personality Type Indicator (May 20 11 am)**

www.statisticalinnovations.com