



2021 NCME ANNUAL MEETING

TRAINING SESSIONS

Full Day Training Sessions

Bayesian Networks in Educational Assessment | \$75

Part 1: Friday, June 4th | 9:00 AM - 1:00 PM EST

Part 2: Friday, June 4th | 1:30 PM - 5:30 PM EST

Duanli Yan (ETS); Russell G Almond, (Florida State University); Diego Zapata-Rivera (ETS); Roy Levy (Arizona State University)

The Bayesian paradigm provides a convenient mathematical system for reasoning about evidence. Bayesian networks provide a graphical language for describing complex systems, and reasoning about evidence in complex models. This allows assessment designers to build assessments that have fidelity to cognitive theories and yet are mathematically tractable and can be refined with observational data. The first part of the training course will concentrate on Bayesian net basics, while the second part will concentrate on model building and recent developments in the field.

Statistical Learning of Process Data: Methods, Software, and Applications | \$75

Part 1: Friday, June 4th | 9:00 AM - 1:00 PM EST

Part 2: Friday, June 4th | 1:30 PM - 5:30 PM EST

Jingchen Liu (Columbia University); Susu Zhang (University of Illinois at Urbana)

This full-day workshop introduces a selection of statistical learning methods for analyzing process data, that is, log data from computer-based assessments. Covered topics include (1) data-driven methods for extracting features from response processes; (2) sequence segmentation and subtask analysis with neural language modelling; (3) introduction to ProcData, an R package for process data analysis; and (4) applications of process features to practical testing and learning problems, including scoring, differential item functioning correction, computerized adaptive testing, and adaptive learning. Mode of instruction will be a blend of presentations, for topics (1) and (2), and concrete illustrations in R, for topics (3) and (4). Intended audience are researchers and practitioners interested in data-driven methods for analyzing process data from assessments and learning environments. To fully engage in the hands-on activities, familiarity with R and RStudio is expected. Running the ProcData package requires installation of R, Rcpp, and Python. Installation instructions and support will be provided. Participants are expected to bring their own laptop with Windows or Mac operating system. By the end of the workshop, participants are expected to get a composite picture of process data analysis and know how to conduct various analyses using the ProcData package.

Split Full Day Training Sessions

Creating Custom Interactive Applications with R and Shiny | \$75

Part 1: June 4th | 1:30 PM - 5:30 PM EST

Part 2: June 7th | 1:30 PM - 5:30 PM EST

Christopher Runyon (National Board of Medical Examiners); Joshua Goodman (National Commission on Certification of Physician Assistants); Marcus Walker (National Commission on Certification of Physician Assistants)

This session explores the use of R and the Shiny package for creating unique statistical apps. In many testing, commercial, and academic contexts, there is a need for specialized statistical apps for custom tasks and analyses. Many of the commercially available programs are offered in a one-size-fits-all format, and thus often lack the flexibility needed across multiple contexts. Shiny is a free, open-source resource that can be used to build applications that can be developed and maintained by persons with only a modest level of R programming skill. These apps can be hosted on a webpage or deployed as standalone executable files, and end users of such apps do not need to know any R programming to successfully use them.

Using psychometric tasks as motivating examples, we guide session participants through building a simple app in Shiny. After teaching the foundations of a Shiny program, we expand to showcase some of the advanced capabilities of Shiny use, including generating reports and building standalone executable programs. Participants should have at least a moderate level of R programming ability. More advanced R programmers will still benefit from Shiny information that goes well beyond “hello world” examples often found on Shiny resource pages.

Cognitive Diagnosis Modeling: A General Framework Approach and Its Implementation in R | \$75

Part 1: June 7th | 1:30 PM - 5:30 PM EST

Part 2: June 8th | 1:30 PM - 5:30 PM EST

Jimmy de la Torre (University of Hong Kong); Wenchao Ma (University of Alabama)

The primary aim of the workshop is to provide participants with the necessary practical experience to use cognitive diagnosis models (CDMs) in applied settings. Moreover, it aims to highlight the theoretical underpinnings needed to ground the proper use of CDMs in practice.

In this workshop, participants will be introduced to a proportional reasoning (PR) assessment that was developed from scratch using a CDM paradigm. Participants will get a number of opportunities to work with PR assessment-based data. Moreover, they will learn how to use GDINA, an R package developed by the instructors for a series of CDM analyses (e.g., model calibration, evaluation of model appropriateness at item and test levels, Q-matrix validation, differential item functioning evaluation). To ensure that participants understand the proper use of CDMs, the theoretical bases for these analyses will be discussed.

The intended audience of the workshop includes anyone interested in CDMs who has some familiarity with item response theory (IRT) and R programming language. No previous knowledge of CDM is

required. By the end of the session, participants are expected to have a basic understanding of the theoretical underpinnings of CDM, as well as the capability to conduct various CDM analyses using the GDINA package.

Full Day Training Sessions (4-hour asynchronous, 4-hour session)

Using Stan for Bayesian Psychometric Modeling | \$75

June 7th | 1:30 PM - 5:30 PM EST

Yong Luo (ETS); Manqian Liao (Duolingo)

This session will provide audience with systematic training on Bayesian estimation of common psychometric models using Stan. The estimation of model parameters for common psychometric models will be illustrated and demonstrated using Stan, with a particular emphasis on IRT models. Further the advantages and disadvantages of Stan comparing to traditional Bayesian software programs such as OpenBUGS and JAGS will be discussed.

This session consists of lecture, demonstration, and hands-on activities of running Stan. It is intended for intermediate and advanced graduate students, researchers, and practitioners who are interested in learning the basics and advanced topics related to parameter estimation of common psychometric models using Stan. It is expected the audience will have some basic knowledge of the Bayesian theory, but not required. Attendees will bring their own laptop and download the software program free online. It is expected that attendees will master the basics of writing Stan codes in running standard and extended psychometric models; further they can develop Stan codes for new psychometric models for their own research and psychometric modeling.

Blatent Psychometrics: Bayesian Psychometric Modeling in R | \$75

June 8th | 9:00 AM - 1:30 PM EST

Jonathan Templin (University of Iowa); Alfonso Martinez (University of Iowa); Catherine Mintz (University of Iowa); Alexis Oakley (University of Iowa); Jihong Zhang (University of Iowa); Mingying Zheng (University of Iowa)

Bayesian modeling techniques have become popular estimation methods used in the analysis of psychometric data. In this workshop, we present a unified approach to the estimation and use of Bayesian psychometric models with the new R package, blatent. Building upon the basis of generalized linear mixed models, this workshop will feature hands-on demonstrations of Bayesian estimation of popular psychometric model families including Item Response Theory (unidimensional and multidimensional), Confirmatory Factor Analysis, Diagnostic Classification (cognitive diagnosis) Models, Structural Equation Models, and Bayesian Networks.

The goal of this workshop is to enable participants to implement Bayesian psychometric models in their own research or operational work. Participants should have, at least, both a basic understanding of R and one psychometric measurement model (e.g., item response, factor analytic, or diagnostic). All slides and R syntax will be available via a workshop website. Participants should bring a laptop with R installed if they wish to follow along with examples.

Morning Half-Day Sessions

A Visual Introduction to Computerized Adaptive Testing | \$45

June 8th | 9:00 AM - 1:00 PM EST

Yuehmei Chien (College Board); David Shin (Pearson)

The training will provide the essential background information on operational computerized adaptive testing (CAT) with an emphasis on CAT components (including ability estimation, item exposure control and content balancing methods--weighted penalty model and shadow tests) and CAT simulation. Besides the traditional presentation through slides, this training consists of hands-on demonstrations of several key concepts, with visual and interactive tools and a CAT simulator.

Practitioners, researchers, and students are invited to participate. A background in IRT is recommended. Participants should bring their own laptops and item pools in CSV file format, as they will access the tools that were designed to help the participants understand important CAT concepts and visualize the results. Installation instruction of the tools will be provided via email prior to the conference. Upon completion of the workshop, participants are expected to have 1) a broader picture about CAT; 2) a deeper understanding of the fundamental CAT techniques; 3) appreciation of the visual techniques used to analyze and present results in an intuitive and pleasing way.

Addressing the Data Challenges from Next-generation Assessments: Data Science Upskilling for Psychometrician | \$45

June 8th | 9:00 AM - 1:00 PM EST

Oren Livne (ETS); Jiangang Hao (ETS)

Digitally Based Assessments (DBAs) offer promising opportunities into insights of test takers' response process information. Yet the significantly increased volume, velocity, and variety of data pose new challenges to psychometricians for handling, analyzing, and interpreting the data to materialize their value. Data science is an emerging interdisciplinary field aimed at obtaining such insights from structured and unstructured data. Data science techniques and practices could and should be adopted into the toolkit of next generation psychometrics to help address the data challenges accompanying DBAs. This workshop is intended on providing a basic toolkit and modeling strategies in the context of DBAs to help psychometricians and data analysts become better equipped to work with the increasingly big and complex data from next-generation assessments.

An Overview of Operational Psychometric Work in Real World | \$45

June 8th | 9:00 AM - 1:00 PM EST

Hyeon-Joo Oh (ETS); JongPil Kim (Riverside Insights); (Presenter) Jinghua Liu (Enrollment Management Associates); Sarah Quesen (Pearson); Hanwook Yoo (ETS)

An overview of the psychometric work routinely done at various testing organizations will be presented in this training session. The training session will focus on the following topics: (1) outline of operational psychometric activities across different testing companies, (2) hands-on activities to review item and test analyses output, (3) hands-on activities to review equating output, and (4) overview of

computerized adaptive testing (CAT) and multi-stage testing (MST) and hands-on activities. If time allows, there will be a brief discussion session regarding factors that affect operational psychometric activities in the CAT and MST environment. We are hoping that through this training session, participants will get a glimpse of the entire operational cycle, as well as gain some understanding of the challenges and practical constraints that psychometricians face at testing organizations. It is targeted toward advanced graduate students who are majoring in psychometrics and seeking a job in a testing organization and new measurement professionals who are interested in an overview of the entire operational testing cycle. Representatives from major testing organizations (e.g., ETS, Pearson, and etc.) will present various topics related to processes in an operational cycle.

Optimal Test Design Approach to Fixed and Adaptive Test Construction using R | \$45

June 8th | 9:00 AM - 1:00 PM EST

Seung W. Choi (University of Texas Austin); Sangdon Lim (University of Texas Austin)

In recent years, fixed test forms and computerized adaptive testing (CAT) forms coexist in many testing programs and are often used interchangeably on the premise that both formats meet the same test specifications. In conventional CAT, however, items are selected through computer algorithms to meet mostly statistical criteria along with other content-related and practical requirements, whereas fixed forms are often created by test development staff using iterative review processes and more holistic criteria. The optimal test design framework can provide an integrated solution for creating test forms in various configurations and formats, conforming to the same specifications and requirements. This workshop will present some foundational principles of the optimal test design approach and their applications in fixed and adaptive test construction. Practical examples will be provided along with an R package for creating and evaluating various fixed and adaptive test formats.

Python, Machine Learning, and Applications - A Gentle Introduction | \$45

June 8th | 9:00 AM - 1:00 PM EST

Zhongmin Cui (CFA Institute)

Machine learning is getting popular in recent years. Its applications span a vast range: from agriculture to astronomy, from business to biology, from communication to chemistry, from data mining to dentistry, from education to economy; the list goes on. The interest in machine learning continues growing as indicated by related presentations and publications. The goal of this lecture-style training is to provide a gentle introduction on this topic. Although other languages are available for machine learning, Python will be introduced as a starter in this training. The main course has two dishes, supervised machine learning and unsupervised machine learning. Dessert samples of using machine learning in educational measurement research conclude the training. Participants do not need to have any experience in machine learning or Python. Upon completion, participants are expected to have a general idea of machine learning and know how to use Python on a simple machine learning project. Participants do not need to bring their laptops or install software; the training will be as gentle as possible so that it is tasty to a broad audience. Having said this, following an example with a laptop near the end of the training would make the dessert taste sweeter.

ReCo: A Shiny App for Automatically Coding Short Text Responses in Assessments | \$45

June 8th | 9:00 AM - 1:00 PM EST

Fabian Zehner (DIPF | Leibniz Institute for Research and Information in Education, Centre for International Student Assessment (ZIB); Nico Andersen (DIPF | Leibniz Institute for Research and Information in Education)

In this training session, participants will learn to use the ReCo shiny app (Automatic Text Response Coder) for automatically coding text responses in assessments. For example, this can be used for scoring constructed responses as correct or incorrect. The session will start with an introduction to the employed methodology (i.a., Latent Semantic Analysis, classification and its evaluation) but will have its focus on hands-on activities. Participants will use a graphical interface in R for automatically coding text responses from an English response data set. Participating assessment developers, practitioners, as well as researchers will be empowered to automatically code constructed responses in their own assessments.

Afternoon Half Day Sessions

Bridging Research and Practice by Examining the Consequences of Assessment Design and Use | \$45

June 8th | 1:30 PM - 5:30 PM EST

David Slomp (University of Lethbridge); Maria Elena Oliveri (Buros Testing Center, University of Nebraska-Lincoln)

Participants will learn how to systematically examine the consequences of assessment design and use for both classroom and large-scale assessment programs.

Participants will be introduced to two approaches—Integrated Design and Appraisal Framework (Slomp, 2016) and Theory of Action (Bennett, 2010)—for integrating attention to consequences into the design and appraisal of assessment programs. The IDAF approach provides a taxonomy for considering questions of fairness, validity and reliability in an integrated fashion that highlights the intended and unintended consequences of decisions made at each stage of an assessment's design and use. The Theory of Action (ToA) model applies logic models to the articulation and testing of claims regarding both how program information is used, and the impact using this information has on individuals or organizations.

An overview of the literature on the consequences of assessment design and use will be provided. Participants will then be guided through two case studies illustrating the application of the IDAF and ToA models. Participants will then work collaboratively on building a plan of action, extrapolated from these frameworks, that they will apply to a third case study.

Computerized Multistage Testing: Theory and Applications | \$45

June 8th | 1:30 PM - 5:30 PM EST

Duanli Yan (ETS); Alina A. von Davier (DuoLingo); Kyung (Chris) Han (Graduate Management Admission Council)

This course provides a general overview of a computerized multistage test (MST) design and its important concepts and processes. The MST design is described, why it is needed, and how it differs from other test designs, such as linear test and computer adaptive test (CAT) designs, how it works, the methodologies involved, and its simulations.

Modeling Writing Process Using Keystroke Logs | \$45

June 8th | 1:30 PM - 5:30 PM EST

Mo Zhang (ETS); Hongwen Guo (ETS); Xiang Liu, (ETS)

In this half-day workshop, participants will have an opportunity to learn about and analyze a newer type of the educational data that is being progressively used in writing research; namely, the keystroke logs collected during the writing process. Information contained in the keystroke logs goes much beyond a holistic evaluation on the written product. From the keystroke logs, one may identify, for example, whether a writer had trouble with retrieving words, edited what was written before the submission, or spent sufficient time and effort on the task.

As much as the opportunities and potential applications given by this type of timing and process data, it also poses many challenges to researchers and practitioners, which includes construct-relevant evidence identification from the logs, evidence extraction/feature engineering, and statistical treatment and modeling of such complex data.

Students and professionals in the areas of writing research and educational measurement are invited. The format of this workshop will be a mix of lecture-style presentation, hands-on data analyses, and group discussion. Some background on statistical analyses will be preferred. Sample R codes for applying Markov or semi-Markov process and other graphical models will be provided. Participants should bring personal laptops with the statistical software R installed.

Principles and Methods in Psychometric Evaluation of Educational Assessments | \$45

June 8th | 1:30 PM - 5:30 PM EST

Louis Roussos (Cognia); Han Yi Kim (ACT); Liuhan (Sophie) Cai (Cognia)

Dr. Louis Roussos has 15 years of experience in evaluating test forms from a variety of assessment programs. He is in the early stages of writing a textbook and will share the “tricks of the trade” he has learned, including guiding principles, methods that flow from these principles, and a variety of real-life examples. Special consideration will be given to importance of communication and complex decision making. The session will entail a mixture of lecture, dialogic learning through interactive discussion and sharing of experiences by participants, methods demonstration, and practical exercises in which participants implement the principles and methods. The exercises will result in constructive feedback to both the presenters and participants.

Using SAS for Monte Carlo Simulation Studies in Item Response Theory | \$45

June 8th | 1:30 PM - 5:30 PM EST

Brian Leventhal (James Madison University); Allison Ames Boykin (University of Arkansas)

Data simulation and Monte Carlo simulation studies are important skills for researchers and practitioners of educational measurement, but there are few resources on the topic. This four-hour workshop presents the basic components of Monte Carlo simulation studies (MCSS). Multiple examples will be illustrated using SAS including simulating total score distribution and item responses using the two-parameter logistic IRT, bi-factor IRT, and graded response model. Material will be applied in nature with considerable discussion of SAS simulation principles and output.

The intended audience includes researchers interested in MCSS applications to measurement models as well as graduate students studying measurement. Comfort with SAS base programming and procedures will be helpful. Participants are not required to have access to SAS during the session. The presentation format will include a mix of illustrations, discussion, and hands-on examples.

As a result of participating in the workshop, attendees will: 1) Articulate the major considerations of a Monte Carlo simulation study, 2) Identify important SAS procedures and techniques for data simulation, 3) Adapt basic simulation techniques to IRT-specific examples, and 4) Extend examples to more complex models and scenarios.

Using School-Level Data from the Stanford Education Data Archive | \$45

June 7th | 1:30 PM - 5:30 PM EST

Sean Reardon (Stanford University); Andrew Ho, (Harvard Graduate School of Education); Benjamin R. Shear (University of Colorado Boulder); Erin Michelle Fahle (St. John's University)

The Stanford Education Data Archive (SEDA) is a growing, publicly-available database of academic achievement and educational contexts. The nationally-comparable achievement data is based on roughly 500 million standardized test scores for students in nearly every U.S. public school in third through eighth grade from the 2008-09 through 2017-18 school years. Initially, SEDA included only estimates of school district and county-level achievement. Subsequently, the data were expanded to include estimates of average school-level achievement and were made accessible to a broader audience through a new, interactive website, The Educational Opportunity Project Data Explorer.

This workshop is intended to introduce researchers of all levels, practitioners, and policymakers to the school-level SEDA achievement and context data. We will provide an overview of both the contents of the SEDA database and the statistical and psychometric methods used to construct the database. The workshop will include presentations by the instructors and hands-on activities designed to help users engage directly with the school-level data. All attendees should have a computer accessible during the training in order to engage in the activities. Attendees who are interested in using the data for research purposes should have statistical software (e.g., R or Stata) installed on their computers.

If you have any questions regarding the training sessions,
please reach out to Sarah Quesen at sarah.quesen@pearson.com