

ANNUAL MEETING

ncme

April 11-14, 2024 – Philadelphia, PA



# WELCOME FROM THE PROGRAM CHAIRS

**Welcome to the 2024 NCME Annual Meeting! We are so excited to have you join us in Philadelphia!**

The theme of the 2024 NCME Annual Meeting is *Reconceptualizing Measurement Theory and Practice to Reduce Inequities*. As measurement researchers, we have a long history of studying educational inequities. Through this work we have identified many areas for improvement. At the same time, standardized tests themselves have been targeted in social, political, and legal arenas as being a major contributor to inequity. Responses in the measurement community have ranged from impassioned defenses on the grounds of fairness, transparency, and access; to critical re-examinations of the definitions of such terms; to active research toward developing more culturally responsive assessment techniques. These conversations are constructive, as they challenge us as measurement professionals to re-evaluate our roles in building equitable societies.

We invite you to attend the training and in-person presentations and engage in conversations with others on how we can truly reconceptualize measurement theory and practice to reduce inequities. Our hope for the 2024 Annual Meeting is to give attendees several ways to think about the theme and have further conversations on related measurement topics and research. There are a few changes to the Annual Meeting that we wanted to highlight.

NCME is planning to offer 19 different training sessions on Thursday April 11. Somewhat different than in years past, all the training sessions will be half-day sessions to allow attendees to attend two training sessions.

The first in-person presentation day, Friday April 12, will kick off with an opening plenary by Sam Kean, the author of *The Disappearing Spoon* and other books related to science history. Our hope is that his talk will challenge us to think about how historical events, inventions, and discoveries outside of the measurement field may have influenced how people think about measurement. The talk will be followed by a coffee break and networking time. The day will close with an invited session focused on the theme where we will hear several perspectives from outside the measurement community on the impact of testing and its role in creating an equitable society. This will be followed by the NCME Welcome Reception.

At different points during the conference, there will be opportunities to engage in hot topic roundtable discussions with colleagues as an alternative to paper sessions. Attendees will be randomly assigned to groups. Discussion prompts on contemporary issues in the field of measurement will be provided. In addition, paper sessions before lunch will be 60 minutes to allow for a longer and more natural lunch break. Exciting debates and important organized discussions will take place in the afternoon each day.

NCME will still have training sessions, innovation demonstrations, paper sessions, and coordinated sessions that people have come to love and expect on a wide range of measurement topics. Below are just a small number of selected sessions that attendees might be interested in each day.

## **Thursday, April 11 Training Sessions:**

- Data Cleaning for Data Sharing Using R (Morning)
- Examining MIRT Models from a Graphical Perspective using RShiny and Mathematica (Morning)
- Bayesian Latent Variable Models in Education Research (Afternoon)
- Professional Training for Graduate Students in Measurement (Afternoon)

## **Friday, April 12 In-Person Conference Sessions:**

- Opening Keynote and Coffee Social (Morning)
- How are State Assessment Programs and People Using Item-PLD Alignment These Days? (10:05 am to 11:05 am)
- NCME Career Award Presentation by Henry Braun (11:25 am to 12:25 pm)
- Testing Standard Revision Updates and Membership Input (11:25 am to 12:55 pm)
- Toward Balanced and Equitable Assessment Systems (1:15 pm to 2:45 pm)
- Fairness in the Next Generations of the Standards: A Discussion of Equity (3:05 pm to 4:35 pm)
- Invited Session Voices from the Measured: The Role of Testing in an Equitable Society (4:55 pm to 6:25 pm)
- NCME Welcome Reception (Evening)

# WELCOME FROM THE PROGRAM CHAIRS

## **Saturday, April 13 In-Person Conference Sessions:**

- NCME Walk/Run (6:00 am to 7:30 am)
- NCME Business Meeting and Presidential Address (9:00 am to 11:00 am)
- Years of Learning: Relating Changes in Student Attainment to Time (11:25 am to 12:25 pm)
- A Conversation about Rebuilding Pathways to Educational Measurement (1:15 pm to 2:45 pm)
- Foundational Challenges to Foundational Competencies: Debating the Task Force Report (1:15 pm to 2:45 pm)
- Strengthening Partnerships to Improve Classroom Assessment Research, Practice, and Policy (3:05 pm to 4:35 pm)
- One Year after GPT-4: How AI Changed Assessments (3:05 pm to 4:35 pm)
- Sparking a Debate on the Role of Artificial Intelligence in Educational Measurement (4:55 pm to 6:25 pm)
- Re-envisioning Fairness: Far-reaching Implications for Measurement and Assessment Research and Practice (4:55 pm to 6:25 pm)

## **Sunday, April 14 In-Person Conference Sessions:**

- Recent Innovations in Automated Scoring (7:45 am to 9:15 am)
- NCME Mission Fund Selected Projects Showcase (9:35 am to 11:05 am)
- Innovation Demonstration 3 (9:35 am to 11:05 am)
- Surveys of/for the Measurement Community (11:25 am to 12:25 pm)
- A Debate on the Merits of Embedded Standard Setting (1:15 pm to 2:45 pm)
- Personalization in the Service of Equity: What, Why, How, and for Whom? (1:15 pm to 2:45 pm)
- Moving from Accountability to Improvement: A Framework and Lessons from the Field (3:05pm to 4:35 pm)
- Validating Assessment in a Multicultural, Pluralistic Context (3:05p m to 4:35 pm)

## **Electronic Board Sessions Across Days:**

- Graduate Student eBoard Sessions 1 and 2 (Friday, April 12, 11:25 am to 12:25 pm; Saturday, April 13, 11:25 am to 12:25 pm)
- Innovation Demonstration Sessions 1 and 2 (Friday, April 12, 10:05 am to 11:05 am; Saturday April 13, 1:15 pm to 2:45 pm)
- Electronic Board Sessions 1, 2, 3, and 4 (Friday, April 12, 1:15 pm to 2:45 pm, Friday, April 12, 3:05 pm to 4:35 pm; Saturday, April 12, 3:05 pm to 4:35 pm; Sunday, April 14, 1:15 pm to 2:45 pm)

We are so thankful to all who have contributed to this engaging program, including submission authors and all volunteers. We are appreciative of the reviewers for providing helpful feedback as well as colleagues who volunteered as discussants and chairs. We want to thank Ji Seung Yang and Wesley Bonifay, the Training and Professional Development Committee Chairs, as well as Janine A. Jackson, Chair of the Graduate Students Issues Committee, for her work on the program. We are also very thankful to the previous NCME program chairs for their help as well as feedback and discussion with next year's program chairs. Finally, we want to thank the NCME President Michael Walker and NCME Executive Director Rich Patz for their help, support, and encouragement.

Jonathan Weeks and Adam Wyse  
2024 NCME Annual Conference Co-Chairs



# The SAT Suite Is Going **Digital**

## Meeting the Evolving Needs of Students and Educators

**The new digital format is more relevant for students than ever before.**



Shorter tests, with shorter reading passages



Built-in graphing calculator



Information about college and career opportunities

**The change is good news for educators too.**



Shorter tests mean shorter test days



No more packing, sorting, or shipping test materials



More flexibility for where and when the tests can be administered

Timeline for the transition to digital:

**Fall 2023:** PSAT™ 8/9 and PSAT/NMSQT®

**Spring 2024:** SAT® and PSAT™ 10

Learn more at  
[sat.org/digital-educators](https://sat.org/digital-educators)



© 2023 College Board.



# TABLE OF CONTENTS

General Meeting Information ..... 06  
 Floor Plans ..... 07  
 NCME Leadership ..... 09  
 Proposal Reviewers ..... 12  
 Schedule at a Glance ..... 16

## FULL SCHEDULE

Thursday, April 11 ..... 24  
 Friday, April 12 ..... 32  
 Saturday, April 13 ..... 70  
 Sunday, April 14 ..... 106

### edCount is pleased to sponsor NCME

Over 20 years of service to students and educators!



#### Our Belief Statement

Every individual brings unique experiences, skill sets, and perspectives that work to advance our purpose: continuously improving the quality, fairness, and accessibility of education for all students.

#### Our Services

- Assessment Design, Development, and Evaluation
- Instructional Systems and Capacity Building
- Policy Analysis and Technical Assistance



[www.edCount.com](http://www.edCount.com)  
 (202) 895-1502 | [info@edCount.com](mailto:info@edCount.com)



[www.acsventures.com](http://www.acsventures.com)  
[in acs-ventures-llc](https://www.linkedin.com/company/acs-ventures-llc)  
[✉ inquiries@acsventures.com](mailto:inquiries@acsventures.com)



#### Professional Certification • Professional Licensure • Education

ACS was established to address a need in the assessment community for design, evaluation, operational support, and quality assurance. Our staff members consult directly with state, regional, national, and international education and credentialing testing programs. ACS has designed and led numerous test development and validation activities, conducted alignment and standard setting studies for a wide variety of clients, and has provided actionable and appropriate guidance and feedback to program leaders on various testing related issues.

#### Design



Program Design & Redesign  
 Assessment Design & Redesign  
 Research Design

#### Operational Support



Standard Setting  
 Alignment  
 Psychometric Analysis  
 Technical Documentation

#### Quality Assurance



Psychometric Oversight  
 Program Review and Audit  
 Program Evaluation  
 Technical Advisory Committees

# GENERAL MEETING INFORMATION

## Welcome to the 2024 NCME Annual Meeting in Philadelphia, PA!

### NCME/AERA REGISTRATION & INFORMATION DESK

NCME/AERA check-in and registration is located at the Pennsylvania Convention Center, in conjunction with the AERA registration desk. Stop by the registration desk to pick up your conference materials including your name badge.

*NCME/AERA Registration will be open the following hours:*

- Wednesday, April 10 2:00 pm to 7:00 pm
- Thursday, April 11 7:00 am to 7:00 pm
- Friday, April 12 7:00 am to 7:00 pm
- Saturday, April 13 7:00 am to 7:00 pm
- Sunday, April 14 7:00 am to 4:30 pm

In addition to the registration desk, stop by the NCME Information Desk to ask questions about your membership, the program, the NCME Events app, and for any other questions! The Information Desk is located outside Room 126 on the Broad Street entrance of the Convention Center. If you are participating in the NCME 5K Fun Run, please stop by the desk to pick up your shirt.

*The NCME Information Desk will be open the following hours:*

- Thursday, April 11 7:30 am to 5:00 pm
- Friday, April 12 7:15 am to 5:00 pm
- Saturday, April 13 7:30 am to 5:00 pm
- Sunday, April 14 7:30 am to 12:00 pm

### NCME GIVES BACK

This year in conjunction with the NCME Annual Meeting, we are asking NCME members and attendees to consider supporting The Public Interest Law Center through a monetary donation. The Public Interest Law Center uses high-impact legal strategies to advance the civil, social, and economic rights of communities in the Philadelphia region facing discrimination, inequality, and poverty. Please visit [their website](#) or scan the QR code below to donate.



### DOWNLOAD THE OFFICIAL NCME EVENTS APP

- Download the eConference.io app from the App Store or Google Play.
- Open the app and enter NCME 2024 as the conference code.
- Once you are in the app, "Click to Log In" at the top of the screen and enter the log-in credentials from the Know Before You Go email to access.

### SOCIAL MEDIA

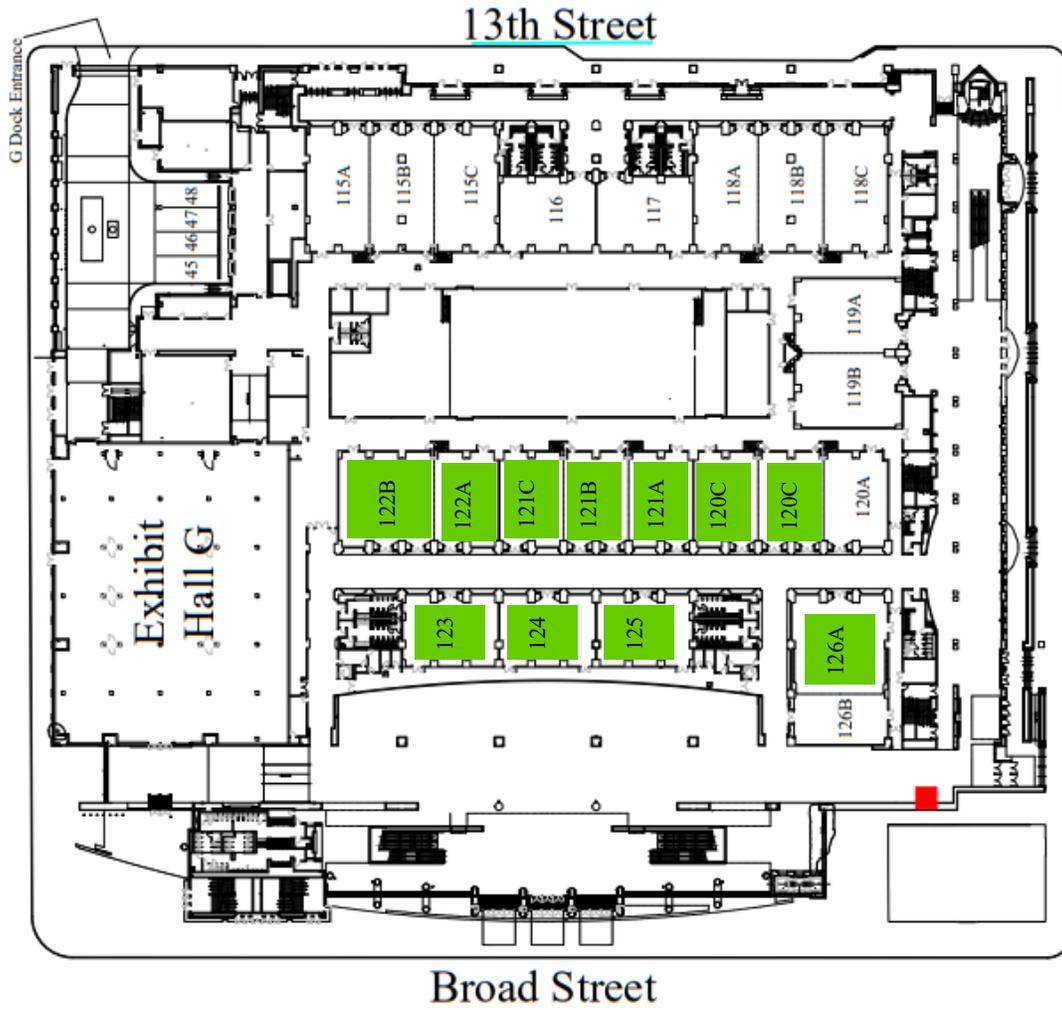
Share your experience at the NCME Annual Meeting by using #NCME2024 in your posts!

### FUTURE ANNUAL MEETING

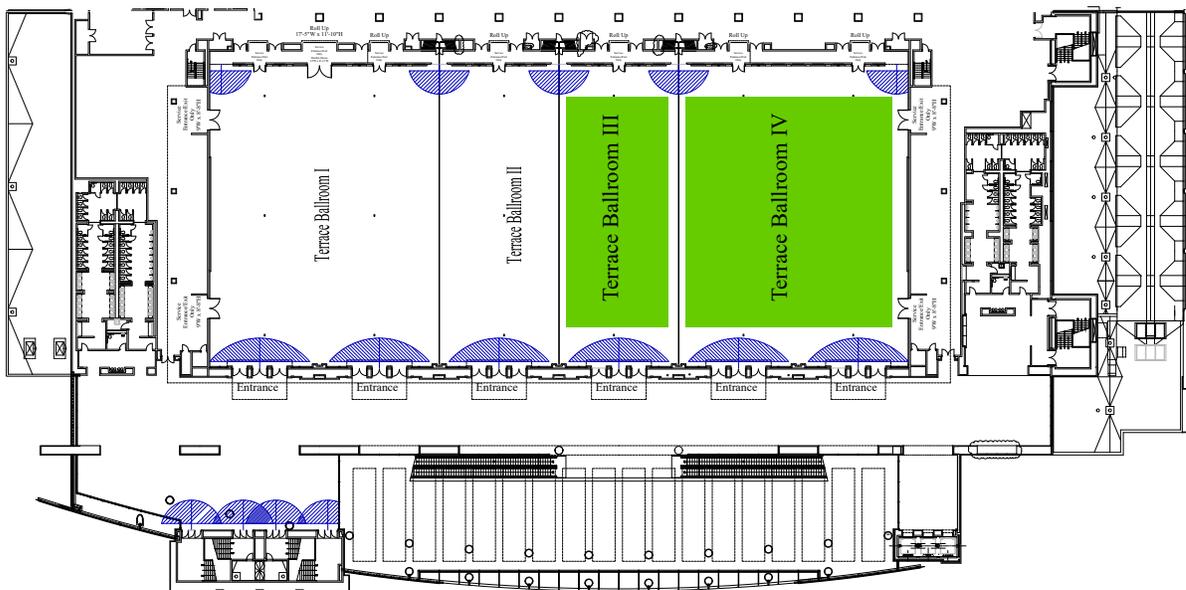
Denver, CO  
April 24-27, 2025

# FLOOR PLANS

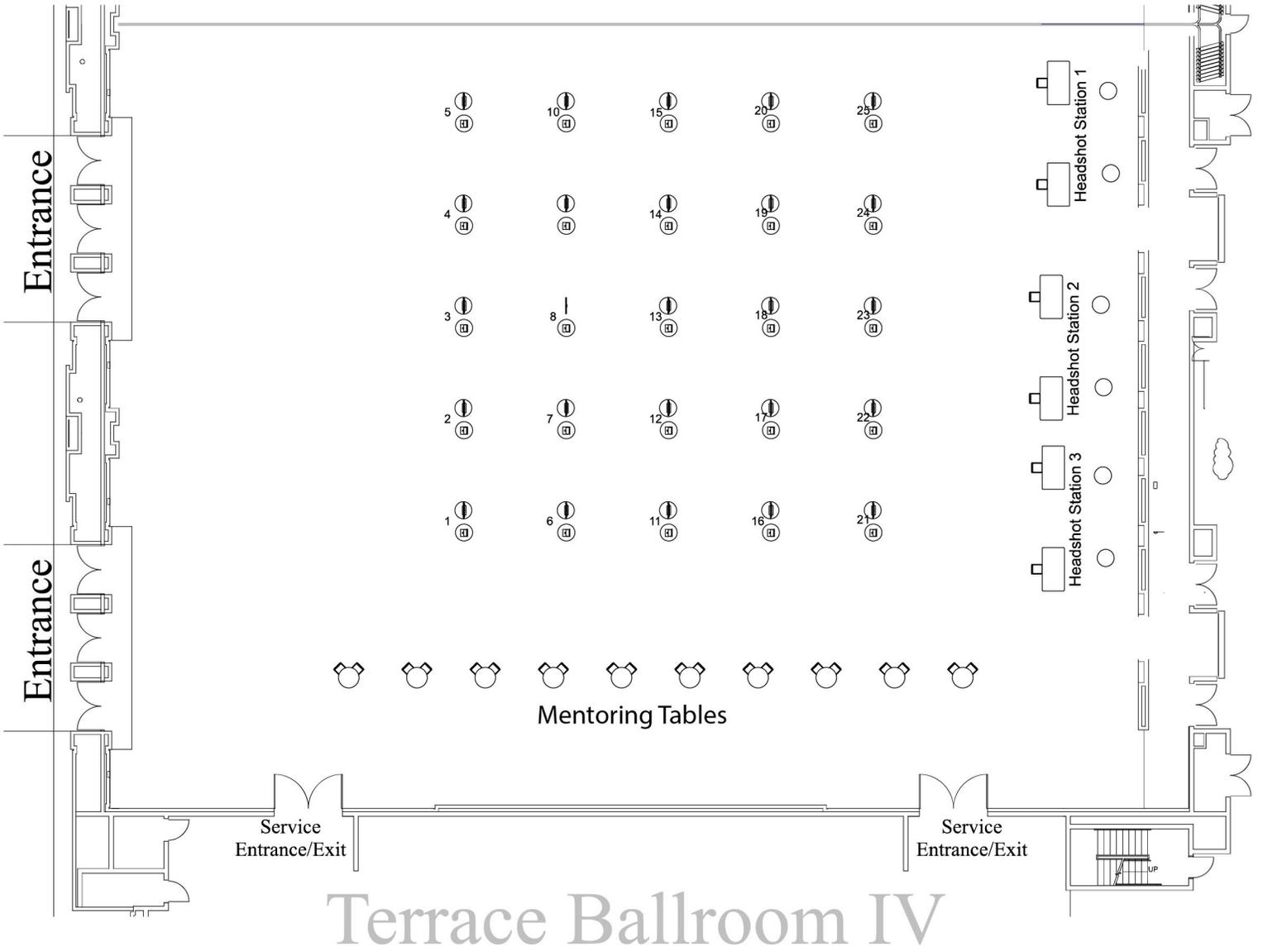
## LEVEL 1



## LEVEL 4



# FLOOR PLANS



## Terrace Ballroom IV

# NCME LEADERSHIP

## NCME OFFICERS

President	Michael Walker, HumRRO
President-Elect	Andrew Ho, Harvard University
Past President	Deborah Harris, University of Iowa

## NCME DIRECTORS

Antionette Stroter Liberty University	Kyndra Middleton Howard University
Susan Davis-Becker ACS Ventures	Li Cai UCLA CRESST
Zachary Warner New York State Education Department	Rochelle Michel Smarter Balanced

## NCME EXECUTIVE DIRECTOR

Richard J. Patz

# EDITORS

Journal of Educational Measurement	Chun Wang University of Washington
Chinese/English Journal of Educational Measurement and Evaluation	Yi Zheng Arizona State University
	Hua-Hua Chang Purdue University
	Tao Xin Beijing Normal University
Educational Measurement: Issues and Practice	Dr. Zhongmin Cui CFA Institute
ITEMS Editor	Brian C. Leventhal James Madison University
NCME Book Series	Kadriye Ercikan ETS
News@NCME	Cheng Hua University of Montevallo
NCME Website	Nathan Dadey Center for Assessment

# 2024 ANNUAL MEETING CHAIRS

Annual Meeting Program Chairs

Jonathan Weeks  
ETS

Training and Professional Development  
Committee Chairs

Adam Wyse  
Renaissance  
Wes Bonifay  
University of Missouri

Graduate Student Sessions Chair

Ji Seung Yang  
University of Maryland

Fitness Run/Walk Director

Janine Jackson  
Morgan State University  
Katherine Furgol Castellano  
ETS

The American Board of Internal Medicine is delighted to welcome our colleagues in educational measurement and psychometrics to our home in historic Philadelphia.





ETS is launching something  
*groundbreaking.*

**April 9, 2024**

**[ets.org/research](https://ets.org/research)**

Copyright © 2024 by ETS. All rights reserved. ETS and the ETS logo are registered trademarks of ETS in the United States and other countries.

## PROPOSAL REVIEWERS

Christiana Aikenosi Akande  
Tony Albano  
Isaac Amoako  
Benjamin Andrews  
Judit Antal  
Elizabeth Ayers-Wright  
Alison Bailey  
David Bamat  
Erin Banjanovic  
Michael Beck  
Magdalen Beiting-Parrish  
Daniel Bolt  
Chad W. Buckendahl  
Michael Brannen Bunch  
Donna J. Butterbaugh  
Serap Büyükkıdık  
Heather Buzick  
Tiago A. Caliço  
Luciana Cancado  
Maritza Casas  
Ruben Castaneda  
Kamal Chawla  
Jianshen Chen  
Britte Haugan Cheng  
Yiling Cheng  
Yi-Chen Chiang  
Hye-Jeong Choi  
Dante Cisterna-Alburquerque  
Amanda Clauser  
Mengyao Cui  
Laurie Davis  
Susan Davis-Becker  
Teresa Dawber  
Lawrence T. DeCarlo  
Onur Demirkaya  
Ben Domingue  
Howard Everson  
Nnamdi C. Ezike  
Fen Fan  
Rich Feinberg  
Anthony D. Fina  
Holmes Finch  
Yanyan Fu  
Katherine Furgol Castellano  
Brian Gane  
Yuan Ge  
Melissa L. Gholson  
Kylie Gorney  
Irina Grabovsky  
Hongwen Guo  
Yage Leah Guo  
Ahmet Guven  
Kathryn Hille  
Igor Himelfarb  
Hyeri Hong

Cheng Hua  
Anne Corinne Huggins-  
Manley  
Janine Jackson  
Kuan Yu Jin  
Marc Anthony Johnson  
Andrew Jones  
Edmund Jones  
Eli Andrew Jones  
Daniel Jurich  
Paraskevi (Voula) Kanistra  
Klint Kanopka  
Yusuf Kara  
Hacer Karamese  
Leslie Keng  
Justin L. Kern  
Eunhee Keum  
Stella Kim  
Jennifer Kobrin  
Charalambos Kollias  
Rodrigo Schames  
Kreitchmann  
Huan Kuang  
Alexander Kwako  
Hollis Lai  
Wendy Lam  
Ida Lawrence  
Anh Thu Le  
Brian C. Leventhal  
Jennifer L. Lewis  
Isaac Li  
Yuan-Ling Liaw  
Hwanggyu Lim  
Youn Seon Lim  
Ami Lin  
Séverin Lions  
Chunyan Liu  
Jinghua Liu  
Jing Lu  
Hotaka Maeda  
Xia Mao  
Danette Waller McKinley  
Janet Mee  
Ivy Mejia  
Jing Miao  
Michalis Michaelides  
Andrew J. Middlestead  
Luke Weisman Miratrix  
Mubarak O. Mojinyinola  
Kristin M. Morrison  
Tim Moses  
Aaron Myers  
Xinyu Ni  
Katherine Nolan  
Steven Nydick

Francis O'Donnell  
Yesim Ozer Ozkan  
Seohee Park  
Thanos Patelis  
Richard Patz  
Han-Hui Por  
Sonya Powers  
Stanley N. Rabinowitz  
Kelly Rewley  
Michael C. Rodriguez  
Louis Roussos  
Merve Sarac  
Paulius Satkus  
Edynn Sato  
Matthew Schultz  
Robert Schwartz  
Carl Setzer  
S.Kanageswari Suppiah  
Shanmugam  
Dorota Staniewska  
Elizabeth Stone  
Yu-Lan Su  
Joshua Sussman  
Dubravka Svetina Valdivia  
Kyoko Taniguchi  
Wei Tao  
Lisette Tolentino  
Emily Karen Toutkoushian  
Anne Traynor  
Secil Ugurlu  
Stephanie Underhill  
Siyu Wan  
Min Wang  
Zhuoran Wang  
Jonathan Weeks  
Alexander Weissman  
Jordan M. Wheeler  
Natasha Jayne Williams  
Phoebe C. Winter  
Tong Wu  
Yi-Chen Wu  
Yi-Fang Wu  
Adam E. Wyse  
Jiawei Xiong  
Ji Seung Yang  
Lihua Yao  
Hanwook Yoo  
Paul Zavitkovsky  
Fabian Zehner  
Ji Zeng  
Liru Zhang  
Mo Zhang  
Rongchun Zhu

## TRAINING SESSION REVIEWERS

Wes Bonifay  
Youngjin Han  
Jeneve Nicola Swaby  
Ji Seung Yang  
Xiaying Zheng

## GRADUATE STUDENT ABSTRACT REVIEWERS

Mohammed Abulela  
Nana Amma Asamoah  
Kayla Burt  
Sandra Liliana Camargo  
Salamanca  
Carlos Chavez  
Guanyu Chen  
Hong Chen  
Jeongwon Choi  
Jiayi Deng  
Deborah Fabiyi  
Maria-Paz Fernandez  
Tavia Flowers  
G Curt Fulwider  
Tuba Gezer  
Ihnwhi Heo  
Qi Huang  
Latif Kadir  
Radhika Kapoor  
Olasunkanmi Kehinde  
Seongeun Kim

Yun-Kyung Kim  
Alexandra Levin Kislyonkova  
Kevin Krost  
Ferdinand Banji Kumolalo  
Haeju Lee  
Hyeryung Lee  
Meng-Hsun Lee  
Ziying Li  
Min Liang  
Xiangyi Liao  
Liu Liu  
Jinwen Luo  
Wanjing Anya Ma  
Constanza Mardones  
Catherine Elizabeth Mintz  
Nicolas Mireles  
Malitsitso Moteane  
Megan M. Mulvihill  
Roza Nalbandyan  
Kelsey Nason  
Daniel O. Oyeniran

Onur Ramazan  
He Ren  
Yelisey A. Shapovalov  
Julian Maximilian Siebert  
Dan Song  
Olushola Olufemi Soyoye  
Alvin Daiz Tenorio  
Bowen Wang  
Yu Wang  
Tarid Wongvorachan  
Yi-jung Wu  
Xingyao Xiao  
Mingfeng Xue  
Sinan Yavuz  
Ye Yuan  
Biao Zeng  
Mingying Zheng  
Xinchang Zhou  
Montserrat B. Valdivia  
Medinaceli

### Fostering Greater Student Learning For Over 20 Years

*Supporting states and local  
districts across the country  
to design and implement  
high-quality assessment  
and accountability systems*

X@NCIEA1

[www.nciea.org](http://www.nciea.org)



31 MOUNT VERNON STREET • DOVER, NH 03820 • 603.516.7900

# Empower educators. Lift students.

Navy, a comprehensive classroom assessment suite, provides real-time, actionable and, most crucially, reliable results at the granular standards-level to correctly inform personalized instruction.



Navy gives you trustworthy learning evidence that's personal to each student to guide targeted instruction.

Keep track of each student's progress on the standards they're expected to learn throughout the school year and *make the vision of personalized learning a reality.*

## Learn more from the creator of Navy

*The Impact of Large-scale Diagnostic Classroom Assessment on Student Achievement*

Convention Center | Room 121C  
11:25 a.m. – 12:25 p.m.

On Sunday, April 14, hear from Navy creator, Dr. Laine Bradshaw, about why classroom standards-level assessment systems hold promise for improving learning trajectories by providing detailed, just-in-time feedback for informing personalized learning in the classroom.

Dr. Bradshaw will examine the impact of Navy, a large-scale, diagnostic measurement-based classroom assessment system for learning and student achievement across subjects, grades, and over time.

800-627-7271 | [PearsonAssessments.com](https://www.pearsonassessments.com)

Copyright © 2024 Pearson Education, Inc. or its affiliates. All rights reserved. Pearson and Navy are trademarks, in the US and/or other countries, of Pearson plc. C9107 ML 2/24



## ACT's Vision For the Future

**By 2032, ACT will empower 20.2 million more learners to exit high school ready for postsecondary and work opportunities.**

ACT's objective, research-based solutions are designed to help students know that they are ready for life. Our solutions empower educators and encourage students at all stages of their journeys by recognizing, rewarding, and supporting learning wherever and however it happens. We are focused on creating solutions for a world of evolving opportunities and providing more support for learners to chart pathways toward greater mobility and economic prosperity.

We are working to create a world in which all learners have the opportunity to fulfill their potential.

Learn more at [act.org](https://act.org).



## The future of testing. Here today.



### Expand your applicant pool

Tap into a diverse pool of candidates from 210+ countries and territories of origin, who have taken the Duolingo English Test because of its radical accessibility.



### Built on the latest assessment science

The Duolingo English Test is a computer adaptive test powered by rigorous research and AI. Results are highly correlated with other assessments, such as the TOEFL and the IELTS.



### Innovative test security

Industry-leading security protocols, individual test proctoring, and computer adaptive technology help prevent fraud and cheating and ensure results you can trust.



### Convenient results dashboard

Access candidates' certificates, video interviews, and writing samples through a free and secure dashboard. Quickly and easily view applicant data with filtering tools.

[englishtest.duolingo.com/edu](https://englishtest.duolingo.com/edu)

# The AAMC Strengthens the World's Most Advanced Medical Care



Representing institutions that deliver the world's most advanced medical care, the AAMC provides programs and services that support the entire spectrum of education, research, and health care. We believe in medical education that prepares physicians and scientists to meet the nation's evolving health needs, support a diverse and culturally competent health care workforce, lead medical advancements that prevent disease and alleviate suffering, and more.

Learn more at [aamc.org](https://aamc.org).

24-061 (03/24)

Association of  
American Medical Colleges



## Drive improvement with curriculum-aligned assessment

Gain insight. Read our whitepaper at [cognia.org/accountability](https://cognia.org/accountability)

## DRC Is Honored To Be A Silver Sponsor Of The NCME Annual Meeting



[datarecognitioncorp.com](https://datarecognitioncorp.com)



## REASSESS THE FUTURE: SUPPORTING THE NEXT GENERATION OF PHYSICIANS

In collaboration with the medical education community, NBME is exploring how new, innovative and inclusive assessment tools can help future physicians develop not only the knowledge, but the skills and behaviors needed to better serve their patients.

Learn about the initiatives and research currently underway to make this possible at [ReassesstheFuture.org](https://ReassesstheFuture.org).



[REASSESTHEFUTURE.ORG](https://REASSESTHEFUTURE.ORG)

© NBME 2024

# SCHEDULE AT-A-GLANCE

(SESSIONS MARKED WITH AN \* HAVE AN ADDITIONAL COST)

## THURSDAY, APRIL 11

Begin Time	End Time	Room	Session Title
8:45 AM	12:45 PM	120B	An Introduction to Creating Video Games for Measurement: From Design to Analysis*
8:45 AM	12:45 PM	120C	An Introduction to the Theory of Standard Setting*
8:45 AM	12:45 PM	121A	Data Cleaning for Data Sharing Using R*
8:45 AM	12:45 PM	121B	Cognitive Diagnosis Modeling: A General Framework Approach and Its Implementation in R*
8:45 AM	12:45 PM	121C	Designing Learning Progressions to Advance Equity in Assessment and Learning*
8:45 AM	12:45 PM	122A	Demystify Amazon Web Services (AWS): Cloud Computing and Psychometric Applications*
8:45 AM	12:45 PM	122B	Applying Data Mining Methods to Detect Test Fraud*
8:45 AM	12:45 PM	123	Examining MIRT Models from a Graphical Perspective Using RShiny and Mathematica*
8:45 AM	12:45 PM	125	Organizational Leadership for Measurement Experts*
1:00 PM	5:00 PM	120B	Diagnostic Classification Modeling with R and Stan*
1:00 PM	5:00 PM	120C	Educational Testing and Psychometrics with R Package Dexter*
1:00 PM	5:00 PM	121A	Bayesian Latent Variable Modeling In Education Research*
1:00 PM	5:00 PM	121B	Longitudinal Diagnostic Classification Models: Theory and Applications*
1:00 PM	5:00 PM	121C	Comprehensive Statistical Model Evaluation: Traditional, Bayesian, and Information-theoretic Methods and User-friendly Software*
1:00 PM	5:00 PM	122A	Optimal Test Design Approach to Fixed and Adaptive Test Construction using R*
1:00 PM	5:00 PM	122B	Computerized Multistage Testing: Theory, Practical Issues, and Solutions (Book by Routledge)*
1:00 PM	5:00 PM	123	Building Monte Carlo Simulations in R for Measurement, Causal Inference, and Beyond!*
1:00 PM	5:00 PM	124	Professional Training for Graduate Students in Measurement*
1:00 PM	5:00 PM	125	Vertical Scaling: Hands-on Practice and Evaluation of IRT Linear and Non-linear Methods*

## FRIDAY, APRIL 12

Begin Time	End Time	Room	Session Title
7:45 AM	9:00 AM	Ballroom AB Floor 3	Opening Keynote
9:00 AM	9:50 AM	Ballroom AB Floor 3	Coffee Social
10:05 AM	11:05 AM	126A	Historically Marginalized Group Mixer (HMG/CODIT)
10:05 AM	11:05 AM	120B	Nonparametric IRT
10:05 AM	11:05 AM	120C	Learning Progressions
10:05 AM	11:05 AM	121A	Oral Reading Fluency Analyses
10:05 AM	11:05 AM	121B	Credentialing
10:05 AM	11:05 AM	121C	Background Questions and Surveys
10:05 AM	11:05 AM	122A	Network Models
10:05 AM	11:05 AM	122B	How Are State Assessment Programs and People Using Item-PLD Alignment These Days?
10:05 AM	11:05 AM	123	Dimensionality
10:05 AM	11:05 AM	124	Explorations of GPT
10:05 AM	11:05 AM	125	Scaling, Linking, and Equating
10:05 AM	11:05 AM	Terrace 3	Using Generative AI in Education: Results from Large-Scale EdTech Platform R&D
10:05 AM	11:05 AM	Terrace 4	Innovation Demonstration I
11:25 AM	12:25 PM	120B	Rater Models
11:25 AM	12:25 PM	120C	Reliability Estimation
11:25 AM	12:25 PM	121A	Surviving and Understanding the US Department of Education's State Assessment Peer Review
11:25 AM	12:25 PM	121B	Longitudinal Credentialing
11:25 AM	12:25 PM	121C	Computational Thinking and Problem Solving
11:25 AM	12:25 PM	122A	IRT Calibration
11:25 AM	12:25 PM	122B	School Accountability Models in the USA and Abroad: Past, Present, Future
11:25 AM	12:25 PM	123	Response Times
11:25 AM	12:25 PM	124	Applications of AI
11:25 AM	12:25 PM	125	Situational Judgment Tests
11:25 AM	12:25 PM	Terrace 3	Career Award
11:25 AM	12:25 PM	Terrace 4	GSIC Paper Session I
11:25 AM	12:25 PM	201C	Testing Standards Revision Updates and Membership Input (Joint AERA/NCME Session)
1:15 PM	2:45 PM	120B	Towards A New Framework For Measurement In Higher Education Admissions Decision-Making
1:15 PM	2:45 PM	122B	Innovations in the Measurement of 21st Century Skills

## FRIDAY, APRIL 12

Begin Time	End Time	Room	Session Title
1:15 PM	2:45 PM	121A	Challenges and Innovations in Creating Interactive Reports of Student Progress and Growth
1:15 PM	2:45 PM	121B	Navigating Challenges in Operational Psychometric Environments
1:15 PM	2:45 PM	121C	Assessment on the AI: Challenges and Opportunities of Leveraging Large Language Model
1:15 PM	2:45 PM	122A	Cognitive Reasoning and Complexity
1:15 PM	2:45 PM	123	Revolutionizing Alignment and Tagging Using AI: Results from the Front Lines
1:15 PM	2:45 PM	124	Tackling Unfairness in AI Scoring from Multiple Angles
1:15 PM	2:45 PM	125	Automated Scoring I
1:15 PM	2:45 PM	Terrace 3	Toward Balanced and Equitable Assessment Systems: Considerations for Implementation and Use (Joint AERA/NCME Session)
1:15 PM	2:45 PM	Terrace 4	Electronic Board I
3:05 PM	4:35 PM	120B	Comprehensive Assessment and Score Development: The NIH Infant and Toddler Toolbox
3:05 PM	4:35 PM	120C	Randomly Parallel Tests: Theory, Use, and Benefits
3:05 PM	4:35 PM	121A	Furthering research underpinning OECD Programme for International Student Assessment - PISA
3:05 PM	4:35 PM	121B	Linguistically Equitable Reading Screening
3:05 PM	4:35 PM	121C	Using Assessment Engineering and Item Complexity Modeling to Enhance Score Interpretation
3:05 PM	4:35 PM	122A	Bayesian Techniques
3:05 PM	4:35 PM	122B	Unpacking Test-Taker (Dis)Engagement: Innovative Approaches to Enhance Validity in Educational Assessment
3:05 PM	4:35 PM	123	Subscores: A Practical Guide for Their Production and Consumption
3:05 PM	4:35 PM	125	Cognitive Diagnostic Models I
3:05 PM	4:35 PM	Terrace 3	What Are Important Quality Criteria For Assessments In A Competency-Based Educational Program?
3:05 PM	4:35 PM	203AB	Fairness in the Next Generation of the Standards: A Discussion on Equity (Joint AERA/NCME Session)
3:05 PM	4:35 PM	Terrace 4	Electronic Board II
4:45 PM	6:25 PM	Terrace 3	Voices From the Measured: The Role of Testing in an Equitable Society
6:25 PM	8:00 PM	Terrace 4	NCME Welcome Reception

## SATURDAY, APRIL 13

Begin Time	End Time	Room	Session Title
6:00 AM	7:30 AM	Meet in ALOFT HOTEL Lobby	NCME Fitness Run/Walk*
7:30 AM	8:30 AM	Terrace 4	Hot Topics in Measurement - Roundtable Discussion I
9:00 AM	10:00 AM	Terrace 3	Awards Ceremony and Business Meeting
10:00 AM	11:00 AM	Terrace 3	Presidential Address
11:25 AM	12:25 PM	120B	Evaluating Fit
11:25 AM	12:25 PM	120C	Reliability and Scoring
11:25 AM	12:25 PM	121A	Longitudinal ML Models
11:25 AM	12:25 PM	121B	Considerations in Assessing Special Populations
11:25 AM	12:25 PM	121C	ORF Models
11:25 AM	12:25 PM	122A	Effort Responding
11:25 AM	12:25 PM	122B	Methodological Challenges and Solutions for Administering Large-Scale Assessments in India
11:25 AM	12:25 PM	123	Missing Data
11:25 AM	12:25 PM	Terrace 3	Years of Learning: Relating Changes in Student Attainment to Time
11:25 AM	12:25 PM	Terrace 4	GSIC Paper Session II
12:15 PM	1:15 PM	126A	Past President's Luncheon (invite only)
1:15 PM	2:45 PM	120B	Leveraging Experiences of Latin American and Muslim Women for Equitable Measurement Practices
1:15 PM	2:45 PM	120C	Reconceptualizing Diagnostic Classification Models: Applications and New Development
1:15 PM	2:45 PM	121A	Investing in Research and Development to Foster Innovation in PISA
1:15 PM	2:45 PM	121B	Putting Research into Practice
1:15 PM	2:45 PM	121C	Tackling Challenges in Automated Scoring: Data Augmentation, Multilingualism, and Feedback Mechanisms
1:15 PM	2:45 PM	122A	Computer Adaptive Testing Applications
1:15 PM	2:45 PM	122B	A Conversation about Rebuilding Pathways to Educational Measurement
1:15 PM	2:45 PM	123	Scaling, Linking, and Equating in Automatic Item Generation
1:15 PM	2:45 PM	124	Promoting Competency-Based Education
1:15 PM	2:45 PM	125	Classroom Assessment and Culturally Responsive Teaching
1:15 PM	2:45 PM	Terrace 3	Foundational Challenges to Foundational Competencies: Debating the Task Force Report
1:15 PM	2:45 PM	Terrace 4	Innovation Demonstration II
3:05 PM	4:35 PM	120B	What a Cluster! Examining Local Item Dependence in Multidimensional Science Assessment.
3:05 PM	4:35 PM	120C	Equity by Design: New Models for Implementing Universal Design

## SATURDAY, APRIL 13

Begin Time	End Time	Room	Session Title
3:05 PM	4:35 PM	121A	Based on True Story: Developing CAT-based High-stakes Exam with Latest Innovations
3:05 PM	4:35 PM	121B	Approaches to Evaluating Item Quality
3:05 PM	4:35 PM	121C	Automatic Item Generation
3:05 PM	4:35 PM	122A	Validity and Test Use
3:05 PM	4:35 PM	122B	Strengthening Partnerships to Improve Classroom Assessment Research, Practice, and Policy (Joint AERA/NCME Session)
3:05 PM	4:35 PM	123	Pipeline Trends in Certification and Licensure: Psychometric Implications
3:05 PM	4:35 PM	124	Rooted in Community: Equitable Design and Practice in Assessment
3:05 PM	4:35 PM	125	Multistage Testing
3:05 PM	4:35 PM	Terrace 3	One Year After GPT-4: How AI Changed Assessments?
3:05 PM	4:35 PM	Terrace 4	Electronic Board III
4:55 PM	6:25 PM	120B	Validity of Socioculturally Responsive and Culturally Sustaining Assessments: Issues and Practice
4:55 PM	6:25 PM	120C	Findings from a Pilot Study of Classroom-based NGSS Aligned Prototype Assessments
4:55 PM	6:25 PM	121A	Unveiling Narratives: Advancing Equity in Psychometric Testing Through Black Scholars' Perspectives
4:55 PM	6:25 PM	121B	Is That a Mode Difference? Demystifying Mode Effect Detection and Adjustment Methods
4:55 PM	6:25 PM	121C	Item Response Theory
4:55 PM	6:25 PM	122A	Alternative Ways of Estimating Performance in Large Scale Assessments
4:55 PM	6:25 PM	122B	Re-envisioning Fairness: Far-reaching Implications for Measurement and Assessment Research and Practice (Joint AERA/NCME Session)
4:55 PM	6:25 PM	123	Next-Gen Assessments: Probing Higher Cognitive Dimensions in the Common Core Era
4:55 PM	6:25 PM	124	Writing Assessments and Rater Effects
4:55 PM	6:25 PM	125	Test Security and Test-Taking Behavior
4:55 PM	6:25 PM	Terrace 3	Sparking a Debate on the Role of Artificial Intelligence in Educational Measurement

## SUNDAY, APRIL 14

Begin Time	End Time	Room	Session Title
7:45 AM	9:15 AM	120C	Recent Innovations in Automated Scoring
7:45 AM	9:15 AM	121A	Assessment and Instruction Coherence for Dynamic Measurement of Student Proficiency
7:45 AM	9:15 AM	121B	Social Emotional Learning
7:45 AM	9:15 AM	121C	Automated Scoring II
7:45 AM	9:15 AM	122A	Differential Item Functioning I
7:45 AM	9:15 AM	122B	Large Scale Assessment
7:45 AM	9:15 AM	123	Developing Curriculum-Agnostic, Instructionally Relevant Assessments
7:45 AM	9:15 AM	124	Student Growth
7:45 AM	9:15 AM	125	Applications of AI II
7:45 AM	9:15 AM	Terrace 3	Exploring Validity of NAEP: Multiple Approaches and Perspectives
9:35 AM	11:05 AM	120B	Innovation Demonstration III
9:35 AM	11:05 AM	120C	Test Taking Effort and Careless Responding
9:35 AM	11:05 AM	121A	Innovations and Practices of AI and Language Models in Large-Scale Assessments
9:35 AM	11:05 AM	121B	NCME Mission Fund
9:35 AM	11:05 AM	121C	Equity-Centered Innovations in Early Learning Assessment
9:35 AM	11:05 AM	122A	Co-Design/Participatory Design Approaches to Assessment Development: Learnings from Four Implementation Studies.
9:35 AM	11:05 AM	122B	Educational Measurement Fifth Edition: A New Approach (Joint AERA/NCME Session)
9:35 AM	11:05 AM	123	Deciphering US Students' Post-COVID Performance Results from National Assessment of Educational Progress
9:35 AM	11:05 AM	124	Machine Learning Computational Psychometrics
9:35 AM	11:05 AM	125	Alignment and Standard Setting: Interrogating Methodologies in Support of Instructional Utility
9:35 AM	11:05 AM	Terrace 3	Is Your Test Instructionally Useful? How Do You Know?
9:35 AM	11:05 AM	Terrace 4	Hot Topics in Measurement - Roundtable Discussion II
11:25 AM	12:25 PM	120C	Classification
11:25 AM	12:25 PM	121A	Model Considerations
11:25 AM	12:25 PM	121B	Application Essays and Characters in Higher Education Admissions
11:25 AM	12:25 PM	121C	Cognitive Diagnostic Models Applications
11:25 AM	12:25 PM	122A	Models for International Assessment
11:25 AM	12:25 PM	122B	Text & Measurement: Critical Approaches to Measuring Text & Measuring Our Field
11:25 AM	12:25 PM	123	Instructing and Assessing Competencies for Future Success: Research From International Baccalaureate

## SUNDAY, APRIL 14

Begin Time	End Time	Room	Session Title
11:25 AM	12:25 PM	126A	Creating a Cadre of Culturally Responsive Measurement Professionals - GSIC and Digital Promise Joint Session
11:25 AM	12:25 PM	Terrace 3	Surveys of/for the Measurement Community
1:15 PM	2:45 PM	120B	The Digital SAT: Concordance Studies and Post-Admin Analyses
1:15 PM	2:45 PM	121A	Revisiting Common Assumptions and Practices in Vertical Scaling
1:15 PM	2:45 PM	121B	Incorporating Multiple Data Source to Explore Group Strategy Variations
1:15 PM	2:45 PM	121C	Cognitive Diagnostic Models II
1:15 PM	2:45 PM	122A	Process Data
1:15 PM	2:45 PM	122B	A Debate on the Merits of Embedded Standard Setting
1:15 PM	2:45 PM	123	Equity and Fairness: Building Relationships that Bridge Gaps
1:15 PM	2:45 PM	124	Culturally Responsive Assessment in Classrooms and Large-Scale Contexts: Theory, Research, and Practice
1:15 PM	2:45 PM	125	Medical and Credentialing Test Applications
1:15 PM	2:45 PM	Terrace 3	Personalization in the Service of Equity: What, Why, How, and for Whom?
1:15 PM	2:45 PM	Terrace 4	Electronic Board IV
3:05 PM	4:35 PM	120C	Culturally Responsive and Sustaining Approaches to Constructed Response Scoring in Assessment
3:05 PM	4:35 PM	121A	A Principled Design Approach to Developing an Alternate English Language Proficiency Assessment
3:05 PM	4:35 PM	121B	Large Language Models in Assessment: Guidelines, Characteristics and Biases
3:05 PM	4:35 PM	121C	Addressing Fairness Concerns in Digital-First, AI-Powered Testing Environments
3:05 PM	4:35 PM	122A	Process Data and Data Quality
3:05 PM	4:35 PM	122B	Moving from Accountability to Improvement: A Framework and Lessons from the Field
3:05 PM	4:35 PM	124	Performance Levels and Standard Setting
3:05 PM	4:35 PM	125	Differential Item Functioning II
3:05 PM	4:35 PM	Terrace 3	Validating Assessment in a Multicultural, Pluralistic Context



Click or scan to learn more about *i-Ready Assessment*.



## Support Every Student with One Coherent Assessment Program

*i-Ready Assessment* delivers clear data, connects to precise instruction, and is backed by committed service. *i-Ready* is highly rated by the National Center on Intensive Intervention and trusted by educators to deliver screening, progress monitoring, and benchmark assessments to 12 million students nationwide.

**Cambium Assessment is a leading provider of online assessments that shape the educational outcome for millions of students**

Cambium Assessment offers summative, interim, and formative assessments. Visit [cambiumassessment.com](http://cambiumassessment.com) to find out more.



Item Response Theory Modeling with

**flexMIRT<sup>®</sup>**  
**IRTPRO<sup>™</sup>**



Driven by our mission,  
rooted in research

Learn more at [nwea.org/research](http://nwea.org/research)



# FULL SCHEDULE

THURSDAY, APRIL 11

## 001. An Introduction to Creating Video Games for Measurement: From Design to Analysis

**Training Session**

**8:45 am to 12:45 pm**

**Convention Center: Level 1 - 120B**

Participants will learn about considerations integral to the creation of videogames for measuring player learning, including the affordances of different game mechanics and design choices on gameplay data and how to derive meaningful indicators from gameplay data. We will use a variety of games to demonstrate how particular game mechanics impact the collection of gameplay data, the analyses that consequently can be performed with those data, and what they can reveal about player learning. This introductory session is for people interested in learning more about designing or using games for measurement purposes. It will not cover advanced statistical modeling or data mining. The training session will have three parts. Part 1: Identifying Game Mechanics for Measurement will offer an overview of the relationship between game design and gameplay data. Part II: Extracting Meaningful Events and Indicators from Gameplay Data will offer hands-on experience with the critical analytical process involved identifying important events and deriving indicators. Part III: Examples of Indicators and Analyses of Gameplay Data will focus on basic data analysis approaches that can be used to make sense of gameplay data. Participants should bring a laptop, tablet, or smartphone to access games for the hands-on activities.

Session Organizer:

*Elizabeth Redman, UCLA CRESST*

Presenters:

*Gregory Chung, UCLA CRESST*

*Tianying Feng, UCLA*

*Jeremy Roberts, PBS KIDS*

## 002. An Introduction to the Theory of Standard Setting

**Training Session**

**8:45 am to 12:45 pm**

**Convention Center: Level 1 - 120C**

This training session is designed to give attendees a thorough understanding of the goals for the use of standard setting procedures and the basic components of the standard setting procedures that are commonly in use. The intent is to show how the structure and processes used for setting standards of proficiency on educational tests can be used to support a validity argument for the interpretations that are commonly made about examinee performance relative to standards located on reporting score scales. The training session is not designed to train participants on a specific standard setting procedure. Examples will be provided of common standard setting procedures with a focus on how to evaluate whether those standard setting procedures are functioning as intended. The examples will include a discussion of the typical steps in a standard setting procedure including panelist selection, training, selection of test items, standard setting tasks, feedback, and final approval of standards. Some statistical methodology will be described and demonstrated for determining the functional characteristics of a standard setting procedure. The training session is appropriate for anyone who is interested in learning about the conceptual underpinnings of standard setting procedures.

Session Organizer:

*Mark Reckase, Psychometric Solutions*

## 003. Data Cleaning for Data Sharing Using R

**Training Session**

**8:45 am to 12:45 pm**

**Convention Center: Level 1 - 121A**

Before sharing research study data, it should be vetted to ensure that it is interpretable, analyzable, and reliable. This half-day virtual workshop will provide a foundational understanding of how to organize data for the purpose of data sharing. This workshop is for any education researcher who could benefit from guidance on how to take a messy raw dataset and organize it into a shareable data product. This workshop assumes you have some experience working with rectangular data, as well as a basic

# FULL SCHEDULE

THURSDAY, APRIL 11

working knowledge of the R programming language and experience working in RStudio. This course will focus on functions in the Tidyverse so familiarity with that package will be helpful, but is not required. All participants will need a computer and have both R and RStudio installed. More information will be provided prior to the training.

Session Organizer:

*Crystal Lewis, Crystal Lewis Consulting, LLC*

## **004. Cognitive Diagnosis Modeling: A General Framework Approach and Its Implementation in R Training Session**

**8:45 am to 12:45 pm**

**Convention Center: Level 1 - 121B**

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.

Session Organizer:

*Wenchao Ma, University of Alabama*

Presenter:

*Jimmy de la Torre, University of Hong Kong*

## **005. Designing Learning Progressions to Advance Equity in Assessment and Learning Training Session**

**8:45 am to 12:45 pm**

**Convention Center: Level 1 - 121C**

A learning progression (LP) has the potential to provide actionable information to teachers and students by providing personalized, domain-specific information about what students already know and understand. In this training session, we focus on the design, development, and validation of an LP for the concept of function, which is central to mathematics learning in algebra and beyond. It is important that an LP is validated with the students it is intended to serve. In our work, we have focused on developing and validating an LP and associated activities for students in primarily high-needs schools. During the session, participants will engage in an experiential activity and discuss how it links to the mathematical ideas in the LP, learn how LPs can be developed and applied in a way that advances equity, see examples from steps in an LP validation cycle, and consider how small group work can be interpreted through the lens of an LP. Participants will leave the session with ideas about how to approach LP design and validation, and which psychometric methods can be used for empirical verification.

Session Organizer:

*Edith Aurora Graf, ETS*

Presenters:

*Cheryl Lizano, Southern Illinois University Edwardsville*

*Peter van Rijn, ETS GLOBAL*

### **006. Demystify Amazon Web Services (AWS): Cloud Computing and Psychometric Applications**

#### **Training Session**

**8:45 am to 12:45 pm**

**Convention Center: Level 1 - 122A**

Cloud computing has become increasingly popular over the past few decades, allowing people to store a massive volume of data, access the newest version of software and use the virtual machine with unbeatable computing power. As practitioners who handle assessment data and do various computing tasks daily, it can be helpful to explore how cloud computing technology can be leveraged to improve efficiency and enable effective communication of test results. Given the limitations of existing training materials, a workshop is proposed which targets at audiences who do not come from IT background. In this workshop, we will cover several AWS core services which can be used to accomplish analytical tasks, store outcome in the database, and display results on a dashboard. Participants do not need to have AWS experience. Upon completion, they will be able to streamline typical psychometric tasks, conduct ML and NLP research on cloud, and communicate findings effectively with other stakeholders inside or outside of the organization via interactive dashboard on the cloud. This is a heavy hands-on training and participants are strongly encouraged to bring their laptops to follow along in order to optimize the learning outcomes from this eight-hour training.

Session Organizer:

*Huijuan Meng, Amazon Web Services AWS*

Presenters:

*Vinita Talreja, Amazon Web Services AWS*

*Ye Ma, Amazon Web Services AWS*

*CJ Taylor, Amazon Web Services AWS*

*Jodine Oquendo, Amazon Web Services AWS*

### **007. Applying Data Mining Methods to Detect Test Fraud**

#### **Training Session**

**8:45 am to 12:45 pm**

**Convention Center: Level 1 - 122B**

This session will provide audience with systematic training on applying various data mining models using software programs: R and/or Python to detect fraud in different test formats, such as computer based, computer adaptive or multistage settings. It covers the basics of these two software programs, theories of selected unsupervised and supervised learning methods, including K-Means, Gaussian Finite Mixture, Self-Organization Mapping, K-Nearest Neighbor, Random Forest, Supported Vector Machine, Neural Network with R/Python demonstrations. Further, the advantages and disadvantages of using each software program will be discussed. This session consists of lectures, demonstrations, and hands-on activities of running various commonly used data mining methods. It is intended for intermediate and advanced graduate students, researchers, and practitioners who are interested in learning the basics and advanced topics related to data mining methods. It is expected the audience will have some basic knowledge of R and Python programming, but not required. Attendees will bring their own laptop and download the software programs free online. It is expected that attendees will master the basics of specify various data mining models and applying these models to detect aberrantly behaved test-takers; further, they can apply the skills to their own research and datasets.

Session Organizers:

*Kaiwen Man, University of Alabama*

*Sarah Linnea Toton, Caveon Test Security*

Presenter:

*Yiqin Pan, University of Florida*

### **008. Examining MIRT Models from a Graphical Perspective using RShiny and Mathematica**

**Training Session**

**8:45 am to 12:45 pm**

**Convention Center: Level 1 - 123**

The purpose of this workshop is to walk researchers through 15 different item and test analyses based upon two-dimensional IRT compensatory model item parameters estimated from assessment data. We will illustrate each analysis graphically using the software RShiny and Mathematica. Two software packages were developed for this Workshop: MIRTGraphRS and MIRTGraphM. The packages consist of a suite of RShiny and Mathematica programs that require the user to input a file containing their estimated two-dimensional compensatory item parameters and then the user can select to run one of 15 different graphics programs. All graphics can be downloaded to include in reports or articles. The MIRTGraphRS and MIRTGraphM packages are divided into five categories of graphics for Total Test, Test Information, Two-Dimensional Ability Estimation, and Differential Item Functioning. The goal of this workshop is to provide graphical tools that will yield greater insight into what tests are measuring and inform future test development. The session will have an interactive format where the audience can use their laptops and follow the presenters as they describe the various graphics. Attendees will receive a manual, detailing each of the graphics, and a copy of all the R Shiny and Mathematica programs that are discussed.

Session Organizer:

*Terry Ackerman, University of Iowa*

Presenters:

*Mingqin Zhang, National Council of State Boards of Nursing (NCSBN)*

*Qing Xie, University of Iowa*

### **010. Organizational Leadership for Measurement Experts**

**Training Session**

**8:45 am to 12:45 pm**

**Convention Center: Level 1 - 125**

Participants in this training course will learn to identify the skills needed to function effectively as an organizational leader, learn from the real-world experience of measurement professionals who have been effective in leadership roles, and have the opportunity to reflect upon and explore their own interests and skills. Open to all but specifically intended to address the interests of early and mid-career professionals who are engaged in or contemplating organizational leadership roles, individuals from all backgrounds, including those from groups traditionally under-represented in leadership roles, are highly encouraged to participate. The facilitators have both rigorous scientific training and extensive organizational leadership experience. Leveraging NCME's Foundational Competencies in Educational Measurement, dimensions of effective leadership are presented and discussed broadly, informed by real-world experience from within testing organizations and higher education contexts. The training session will be interactive, affording opportunities for self-reflection, small-group activities, and large-group discussions.

Session Organizer:

*Richard Patz, University of California, Berkeley*

Presenters:

*Ye Tong, NBME*

*Michael C. Rodriguez, University of Minnesota*

*Jennifer Dunn, Pearson Assessments*

*Richard Patz, University of California, Berkeley*

### **011. Diagnostic Classification Modeling with R and Stan**

**Training Session**

**1:00 to 5:00 pm**

**Convention Center: Level 1 - 120B**

Diagnostic classification models (DCMs; also known as cognitive diagnostic models [CDMs]) have gained interest in recent years due to their ability to provide fine-grained actionable feedback while keeping test lengths short. In this workshop, we will cover how to easily estimate and evaluate DCMs with R and Stan. The workshop will include hands-on examples of defining a DCM, estimating the model, and evaluating the fit (e.g., test- and item-level fit, classification accuracy and consistency, etc.). The goal of the workshop is to enable participants to implement DCMs in their own work, and thus is intended for anyone who uses, or would like to use, DCMs for applied or research uses (e.g., psychometricians, faculty, applied researchers, graduate students). Although not necessary, prior experience with R will be helpful. All workshop materials, including slides, examples, and solutions will be available on a workshop website. Participants should have access to a laptop they can bring to the workshop in order to follow along with the examples. Instructions for installing any necessary software will be provided.

Session Organizer:

*W. Jake Thompson, University of Kansas*

### **012. Educational Testing and Psychometrics with R Package dexter**

**Training Session**

**1:00 to 5:00 pm**

**Convention Center: Level 1 - 120C**

Among the 250 core and regular packages listed in the psychometric view of CRAN, dexter stands out by offering a comprehensive system for the management and analysis of test data. This includes a solid SQL data base created automatically, numerous and advanced diagnostic techniques for items and tests, and efficient support for ability estimation (both standard and plausible value based). dexter can calibrate connected multi-booklet designs of great complexity, and analyze large data sets very fast. It includes many innovative techniques for DIF detection, equating, profile analysis, and more. An easy to use graphical interface is provided by the companion package dextergui, while another package, dexterMST, supports a version of multi-stage testing particularly attractive in the context of high stakes exams. dexter is appropriate for production purposes and actually used in production on a daily basis. It was published in the hope of making IRT more appropriate, attractive, and accessible for the many countries and testing establishment around the world that still rely exclusively on classical test theory and traditional equating methods.

Session Organizer:

*Ivailo Partchev, CITO*

Presenter:

*Ivailo Partchev, CITO*

### **013. Bayesian Latent Variable Modeling In Education Research**

**Training Session**

**1:00 to 5:00 pm**

**Convention Center: Level 1 - 121A**

The goal of the session is to provide background on Bayesian latent variable modeling and to illustrate the models using education data. We will consider factor analysis models, item response models, structural equation models, and two-level variations, with distinctions between the models becoming especially blurry under a Bayesian viewpoint. Specific topics include theoretical background, advantages and disadvantages of adopting a Bayesian modeling approach, prior and posterior checking, model estimation, and model extension. The topics will be illustrated via case studies in R, especially focusing on the blavaan package paired with Stan. Session attendees are assumed to have some knowledge of item response and/or structural equation models, though not necessarily in depth.

Session Organizer:

*Edgar Merkle, University of Missouri*

### **014. Longitudinal Diagnostic Classification Models: Theory and Applications**

**Training Session**

**1:00 to 5:00 pm**

**Convention Center: Level 1 - 121B**

Diagnostic classification models (DCMs) are psychometrics tools that focus on providing actionable feedback in the form of examinee attribute classifications. Longitudinal DCMs have been developed and applied as psychometric tools for analyzing diagnostic assessments administered over multiple occasions. Different from traditional psychometric frameworks for growth, which typically provide continuous and norm-referenced growth estimates, longitudinal DCMs provide categorical growth estimates with criterion-referenced interpretations. This workshop focuses on longitudinal DCMs and their application as the psychometric foundation for categorical growth models. After completing this workshop, participants will understand the structure of longitudinal DCMs, be able to estimate longitudinal DCMs using a newly developed R package, and interpret software output. This session is appropriate for graduate students, researchers, and practitioners at the emerging or experienced level. Participants are expected to have a basic knowledge of DCMs and psychometrics to enroll. This session presents both conceptual and technical content and also provides hands-on experience for participants to apply what they learn. Content will mostly be delivered through lecture, and content will be reinforced using hands-on activities. Instructor will encourage audience participation through questions and allow time for discussions among participants and the instructor.

Session Organizer:

*Matthew James Madison, University of Georgia*

### **015. Comprehensive Statistical Model Evaluation: Traditional, Bayesian, and Information theoretic Methods and User-friendly Software**

**Training Session**

**1:00 to 5:00 pm**

**Convention Center: Level 1 - 121C**

Statistical models play a central role in scientific analysis, inference, and decision-making, so it is imperative that researchers diligently and thoroughly evaluate their models before disseminating them. This training session offers an immersive exploration of three perspectives on statistical model evaluation. Attendees will gain a theoretical and methodological understanding of (1) traditional goodness-of-fit testing and bootstrapping procedures, (2) Bayesian prior and posterior predictive model checking, and (3) information-theoretic techniques that adhere to the principle of minimum description length. This discussion will culminate in a simple framework that integrates all three perspectives. Session leaders will then demonstrate a user-friendly Shiny software application that allows users to upload data, specify a statistical model, select any or all of the methods within the framework, and generate a customized model evaluation report. Attendees should bring their own laptops, with R and RStudio installed prior to the session. The intended audience comprises researchers who use statistical models of any form. The methods covered in this session have been applied to item response theory, factor analysis, and structural equation models, and could be extended to other modeling frameworks. There are no prerequisites for this session, other than general familiarity with the practice of statistical modeling.

Session Organizer:

*Wes Bonifay, University of Missouri*

Presenters:

*Wes Bonifay, University of Missouri*

*Sonja D Winter, University of Missouri, Columbia*

*Hanamori Skoblow, University of Missouri*

### **016. Optimal Test Design Approach to Fixed and Adaptive Test Construction using R**

**Training Session**

**1:00 to 5:00 pm**

**Convention Center: Level 1 - 122A**

Many testing programs use fixed test forms and computerized adaptive testing (CAT) forms interchangeably, but conventional CAT selects items through computer algorithms while fixed forms are created using iterative review processes. The optimal test design framework provides an integrated solution for creating test forms in various configurations and offers a simplified workflow for meeting complex blueprint requirements. This workshop covers the principles of the optimal test design framework and their applications in fixed and adaptive test construction, along with practical examples and an R package for creating and evaluating various fixed and adaptive test formats.

Session Organizer:

*Seung W. Choi, University of Texas at Austin*

Presenter:

*Sangdon Lim, Cambium Assessment*

### **017. Computerized Multistage Testing: Theory, Practical Issues, and Solutions (Book by Routledge)**

**Training Session**

**1:00 to 5:00 pm**

**Convention Center: Level 1 - 122B**

(Revised from NCME Training session 2016-2022) This course provides a general overview of a computerized multistage test (MST) design and its important concepts and processes in the age of AI. The MST design is described alongside the considerations for test development, psychometrics, and engineering. We will also discuss how it differs from other test designs, such as a linear test and a computer adaptive test (CAT), and how to design and conduct simulations for an optimal test.

Session Organizer:

*Duanli Yan, ETS*

Presenters:

*Duanli Yan, ETS*

*Alina A. von Davier, Duolingo*

*Kyung (Chris) T. Han, GMAC*

### **018. Building Monte Carlo Simulations in R for Measurement, Causal Inference, and Beyond!**

**Training Session**

**1:00 to 5:00 pm**

**Convention Center: Level 1 - 123**

In this course we will learn how to write Monte Carlo simulations in R. Monte Carlo simulations are an essential tool of inquiry, useful both for small-scale investigations and for formal methodological research. Simulation can assess, for example, how much re-testing biases a regression discontinuity design due to undermining the measurement properties of a test score. Our focus is on the best practices of simulation design. Overall, we will show how a specific simulation framework allows for rapid exploration of the impact of different design choices, measurement qualities, and data concerns, and show how simulation can answer questions that are hard to answer using direct computation. For example, available algebraic formulas are often based on asymptotic approximations, which might not “kick in” if sample sizes are moderate; simulation can provide an answer. In this session we will work through two simulation case studies, showcasing a modular programming approach for good design. Code, along with an on-line open-source textbook, will be provided and demonstrated. Students should bring laptops, with R and RStudio installed, to follow along. By the end, students will be able to adapt provided code to their own purposes moving forward. Some familiarity with R is assumed.

Session Organizer:

*Luke Weisman Miratrix, Harvard Graduate School of Education*

### **019. Professional Training for Graduate Students in Measurement**

#### **Training Session**

**1:00 to 5:00 pm**

**Convention Center: Level 1 - 124**

This training session will address practical topics for graduate students in measurement to find a job and start a career. First, what to do now while they are still in school to best prepare for a job, which includes the types of training employers look for and how to obtain it (classes, workshops, online training, etc.), how to find a topic and complete a dissertation, how to maximize experiences with networking, internships, social media, and volunteering. Second, how to locate and interview for a job, which includes finding open positions and the application process, including tailoring cover letters, references, and resumes. Third, what to expect in the interview process, including online and in-person interviews, job talks, questions to ask, and negotiating an offer. Last, starting a career, adjusting to the work environment, evaluations, and people (clients, students, co-workers, bosses, mentors, etc.), establishing a career path, searching for grants/funds/awards opportunities, work-life balance, dealing with a bad fit, and staying current. The session addresses working-from-home, publishing, professional associations and service, layoffs, and the wide variety of jobs available. The session is interactive, geared to addressing the attendees' particular interests during the session, and providing resource material on all topics as a takeaway.

Session Organizer:

*Deborah J. Harris, The University of Iowa*

Presenters:

*Deborah J. Harris, The University of Iowa*

*Nathan Wall, EMETRIC*

*Yi-Fang Wu, Cambium Assessment, Inc.*

### **020. Vertical Scaling: Hands-on Practice and Evaluation of IRT Linear and Non-linear Methods**

#### **Training Session**

**1:00 to 5:00 pm**

**Convention Center: Level 1 - 125**

This training demonstrates the empirical application of vertical scaling approaches developed for newly launched large-scale assessments. Vertical scales are a current interest in K-12 testing programs for supporting the measuring and tracking of growth across grade levels. This training session focuses on the following topics: (1) conceptual and technical background of vertical scaling, including the definition of student growth, data collection design, and scaling methods, (2) step-by-step guides to conduct item response theory (IRT) vertical scaling with linear and non-linear methods of converting student ability estimates to scaled scores, (3) a hands-on exercise related to evaluating scaling outcomes and comparing scaled scores, and (4) practical issues and future research topics on vertical scaling. Through the training session, we hope participants will achieve empirical vertical scaling work experience and an understanding of the challenges and limitations related to vertical scaling that psychometricians face in new test development. The training is targeted toward advanced graduate students or new measurement professionals who want to practice vertical scaling. We recommend that participants have some background knowledge in IRT-based equating and scaling, but it is not required. Attendees should bring their own laptops to conduct the hands-on activities.

Session Organizer:

*Hyeon-Joo Oh, Riverside Insights*

Presenters:

*Tim Moses, College Board*

*Hanwook Yoo, Ascend Learning*

# FULL SCHEDULE

FRIDAY, APRIL 12

## 021. Keynote Lecture and Conference Opening

**Invited Session**

**7:45 to 9:00 am**

**Convention Center: Level 3 - Ballroom AB**

Presenter:

*Sam Kean*

Sam Kean is the New York Times bestselling author of six books, including *The Ickpick Surgeon*, *The Bastard Brigade*, *The Dueling Neurosurgeons*, and *The Disappearing Spoon*.

His stories have appeared in *The Best American Science and Nature Writing*, *The New Yorker*, *The Atlantic*, *Slate*, and *Psychology Today*, among other places, and his work has been featured on NPR's "Radiolab", "Science Friday", and "All Things Considered," among other shows. *The Bastard Brigade* was a "Science Friday" book of the year, while *Caesar's Last Breath* was the Guardian science book of the year. *The Disappearing Spoon* was a runner-up for the Royal Society book of the year. Both *The Violinist's Thumb* and *The Dueling Neurosurgeons* were nominated for PEN's literary science writing award.



**Coffee Social immediately following**

## 022. Nonparametric IRT

**Paper Session**

**10:05 to 11:05 am**

**Convention Center: Level 1 - 120B**

Chair:

*Haeju Lee, University of North Carolina Greenboro*

Participants:

**Evaluating Data Quality Using a Combination of Nonparametric and Parametric IRT Models**

*Stefanie A. Wind, University of Alabama; Benjamin Kweku Lugu, University of Alabama*

We used real and simulated data to examine the performance of sequential MSA procedure and parametric IRT models to evaluate measurement data quality. The findings show that MSA is an effective tool for screening items and participants with careless responses before applying parametric models.

**Item-Explanatory Mokken Scaling: Using Nonparametric Item Response Theory to Explore Item Characteristics**

*Stefanie A. Wind, University of Alabama*

We considered how nonparametric IRT techniques can be used to explore item characteristics. We used two real-data illustrations and evaluated the sensitivity of the approach using a simulation. Results suggested that MSA techniques can be extended to examine item properties for evidence of monotonicity and invariant ordering.

**Comparing Genetic Algorithms and NIRT for Autism Screening Gender Bias**

*Yiling Cheng, Kaohsiung Medical University; Mark Reckase, Psychometric Solutions*

It has been known that gender differences exist in ASD. Previous approaches often remove items that demonstrate unequal performances across groups. Because certain disorders might exhibit diverse phenotypes, removing items could potentially lead to the exclusion of subgroups. An alternative method for creating screening tools tailored to gender was proposed.

Discussant:

*Daniel Bolt, University of Wisconsin, Madison*

### 023. Learning Progressions

#### Paper Session

10:05 to 11:05 am

Convention Center: Level 1 - 120C

Chair:

*Na Liu, Georgia Institute of Technology*

Participants:

#### **Cultural Sensitivity of Assessments based on Learning Progressions within Early Childhood**

##### **Assessment Systems**

*Joshua Sussman, University of California, Berkeley*

This paper examines the learning progressions (LPs) within early learning and kindergarten entry assessment (EL-KEA) systems. We examine the degree to which models of learning and development operationalized by LPs apply equally well to culturally and linguistically diverse (CLD) families and children identified as White.

#### **Fine-Tuning Gpt-3 For Automatic Scoring In Learning Progression**

*Mingfeng Xue, University of California - Berkeley; Mark Wilson, Berkeley School of Education,*

*University of California, Berkeley; Linda Morell, University of California, BER*

Generative Pre-trained Transformer (GPT) is powerful in natural language processing. This research investigates its potential in automatic scoring for items from a learning progression. We leverage the text classification ability of GPT-3 by fine-tuning the models. Results show the fine-tuned GPT-3 models perform satisfactorily in automatic scoring.

#### **Using Machine Learning to Detect Student Learning Levels along a Learning Progression**

*Duy N. Pham, University of Massachusetts, Amherst; Viet Lai, University of Oregon*

We introduced a novel method of using Natural Language Processing and Machine Learning to detect student learning levels using student short responses to a question measuring a learning progression of functions. Accuracy rate of our model was at 86%. Our precision, recall, and F1 scores ranged from .77 to .92.

Discussant:

*Mark Wilson, Berkeley School of Education, University of California, Berkeley*

### 024. Oral Reading Fluency Analyses

#### Paper Session

10:05 to 11:05 am

Convention Center: Level 1 - 121A

Chair:

*Yu-Lan Su, Center for Applied Linguistics*

Participants:

#### **Modeling Differential Speediness of Readers in Oral Reading Fluency Assessments**

*Yusuf Kara, Southern Methodist University; Xin Qiao, Shandong Academy of Sciences*

This study proposes an extended speed and accuracy model to relax the assumption of constant reading speed in model-based oral reading fluency assessment. Variable reading speed is defined in the form of a quadratic growth model. Theoretical and practical implications are demonstrated through the analysis of an empirical dataset.

#### **Sentence Order Fluency and Formative Assessment of Reading Comprehension**

*Adam Lekwa, Rutgers, The State University of New Jersey*

Formative assessment of reading comprehension has been a psychometric and practical challenge. This presentation introduces Sentence Order Fluency (SOF) as a novel measurement strategy, including evidence of its relationship with comprehension, as well as text characteristics that might promote creation of higher-quality SOF probes.

### **Validity Evidence of Spanish and English Oral Reading Fluency**

*Anh Thu Le, University of Notre Dame; Doris Luft Baker, The University of Texas at Austin; Deni Basaraba, Amplify Education*

We explore whether the way we measure oral reading fluency in bilingual students is reliable across time points. We used 18 passages in Spanish and English to examine this validity. We ran CFA and found strong evidence for the validity of these passages.

Discussant:

*John Sabatini, Institute for Intelligent Systems, University of Memphis*

## **025. Credentialing**

### **Paper Session**

**10:05 to 11:05 am**

**Convention Center: Level 1 - 121B**

Chair:

*Amanda Clauser, National Board of Medical Examiners*

Participants:

### **Assessing Complex Reasoning Through the Design and Development of a Credentialing Program**

*Susan Davis-Becker, ACS Ventures, LLC; Kelley Stethen, ACS Ventures, LLC; Shelby Williamson, National Strength and Conditioning Association*

Credentialing programs are continually challenged to build examinations that evaluate whether candidates can perform job tasks related to professional judgments and evaluations. This presentation details how one program worked towards this goal by implementing a multi-dimensional framework of professional expectations to develop items designed to measure these higher-order thinking tasks.

### **Integrating Content Area Feedback for Subject Exams in Support of Licensure Preparation**

*Onur Ramazan, Washington State University; Irina Grabovsky, NBME; Francis O'Donnell, National Board of Medical Examiners; Carol Morrison, NBME*

This research introduces new composite scores for content area feedback on the NBME® Clinical Science Subject Exams. Analyses provide strong validity evidence based on internal structure, reliability, and relation to other variables. The results indicate that the composite scores may be helpful to medical students preparing for Step 2.

### **Word Embeddings For Machine Learning Scoring On Medical Examination Short Answer Items**

*Yooyoung Park, NBOME; Xia Mao, NBOME*

The study investigates how word embeddings can be used in machine learning-based scoring algorithms to score text responses from short answer items on medical examinations. The performances of scoring algorithms using word embeddings in several machine learning approaches are compared. Finally, factors affecting the efficiency of these algorithms are discussed.

Discussant:

*Andrew Jones, American Board of Surgery*

## **026. Background Questions and Surveys**

### **Paper Session**

**10:05 to 11:05 am**

**Convention Center: Level 1 - 121C**

Chair:

*Yuan Ge, The College Board*

Participants:

### **Leveraging Background Knowledge Measures to Predict Performance in Scenario-Based Assessments**

*Jesse R. Sparks, ETS; Teresa Ober, ETS; Blair Lehman, ETS; Diego Zapata-Rivera, ETS*

Personalization of tasks based on students' funds of knowledge, can be a useful method to foster their engagement and learning. This study examined multiple approaches to measure students' background knowledge related to a scenario-based assessment topic and the degree to which different measures predicted performance on the assessment.

### **Optimizing Background Questionnaire Score Comparability and Model Fit in International Survey Assessments**

*Selene Sunmin Lee ETS; Frederic Robin, ETS*

To achieve both accurate model data fit and high comparability of scores across many countries and language groups, an item fit RMSD statistic threshold value of 0.20 may be optimal to trigger the estimation of group-specific item parameters for a wide range of background questionnaire scales.

### **Investigating “Books at Home” as an Indicator of Student Socioeconomic Status**

*Samuel Dale Ihlenfeldt, University of Minnesota; Kyle Nickodem, University of Minnesota - Twin Cities; Michael C. Rodriguez, University of Minnesota*

We investigated the use of a self-reported count of books at home as an indicator of socioeconomic status on a large-scale student survey. Low correlations between books at home, county median household income, and poverty rates, suggest the need for a better indicator. Possible alternatives are discussed.

Discussant:

*Stanley N Rabinowitz, EdMetric LLC*

## **027. Network Models**

### **Paper Session**

**10:05 to 11:05 am**

**Convention Center: Level 1 - 122A**

Chair:

*Jiawei Xiong, Pearson*

Participants:

### **Additive Bayesian Network: Exploring The Latent Networks From Textual Responses and Scores**

*Constanza Mardones, University of Georgia; Shiyu Wang, University of Georgia; Allan Cohen, University of Georgia; Xiaoming Zhai, University of Georgia*

This study used the additive Bayesian network (ABN) model to build latent networks between latent attributes learned from a diagnostic classification model and latent topics learned from a latent Dirichlet allocation model. Preliminary results showed that the ABN is useful for detecting the complex relationships between these latent variables.

### **IRT and FA are Neural Networks: a Friendly Introduction**

*Denis Federiakin, Johannes Gutenberg University Mainz; Lidia Dobria, Wilbur Wright College; Kevin Shenavai, Johannes Gutenberg University Mainz; Susanne Schmidt, Johannes Gutenberg-University Mainz; Olga Zlatkin-Troitschanskaia, Johannes Gutenberg University*

We describe how to configure a Neural Network to enable it to estimate the parameters of either Item Response Theory or Factor Analysis models. For this purpose, we utilize the architecture of Variational Autoencoders, which is based on Variational Inference. Real data examples are used for benchmarking.

### **Using Network Psychometrics to Identify Wording Effects in a Socioemotional Skills Survey**

*Ashley Clelland, University of Alberta; Okan Bulut, University of Alberta*

We explored the effectiveness of using network psychometrics to identify wording effects in five scales from the Organization for Economic Cooperation and Development’s Social and Emotional Skills Survey dataset. We validated network psychometric models against confirmatory factor analysis models and found that results were strongly consistent for two scales.

### 028. How Are State Assessment Programs and People Using Item-PLD Alignment These Days?

#### Coordinated Paper Session

10:05 to 11:05 am

Convention Center: Level 1 - 122B

Abstract (for the programs, 200 words) Measurement researchers, psychometricians, and item development leaders now recognize the importance of matching item response demands with knowledge and skill expectations in performance level descriptors (PLDs). While perhaps most important to test score interpretations and uses, item PLD alignment also is important for understanding and using PLDs for instructional planning, alignment of test items to test blueprints and content standards, test forms assembly, and standard setting. It's also relevant to defining item cognitive complexity. Item-PLD alignment also can reduce inequities for test takers whose exposure to state content standards is limited. In this session, three experts in using item-PLD alignment for operational testing programs will describe their recent work in practical applications of item-PLD alignment. The discussant will comment on the papers and extend the work described in the session to research on and concepts of item cognitive complexity.

Session Organizer:

*Steve Ferrara, HumRRO*

Participants:

#### The Expanding Uses of Item-to-Range-ALD Alignment

*Christina Schneider, Cambium Assessment, Inc.*

#### Matching Items to PLDs and Annotating to Guide Instruction

*Arthur Thacker, HumRRO; Andrea Sinclair, HumRRO*

#### Why and How to Use ALDs for Alignment

*Ellen Forte, edCount, LLC; Melissa Fincher, edCount, LLC*

Discussant:

*Steve Ferrara, HumRRO*

### 029. Dimensionality

#### Paper Session

10:05 to 11:05 am

Convention Center: Level 1 - 123

Chair:

*Ji Zeng, Michigan Department of Education*

Participants:

#### Using Bifactor Models to Assess the Dimensionality of Information and Computer Literacy

*Yuan-Ling Liaw, IEA Hamburg; Kerstin Drossel, Universität Paderborn; Birgit Eickelmann, Universität Paderborn; Mojca Rozman, IEA Hamburg*

Large-scale assessments evaluate education systems in domains such as reading and mathematics, but some argue they mainly reflect overall cognitive ability. Using German ICILS data, we examine the extent to which students' scores in computer and Information literacy, computational thinking, reading, and cognitive tests reflect their overall and subject-specific abilities.

#### Variational Approximation for Efficient Estimation of Latent Variable Models

*Jihang Chen, Boston College; Zhushan Mandy Li, Boston College*

This study introduces variational approximation as an efficient alternative to Hamiltonian Monte Carlo (HMC) for latent variable model (e.g., CFA and IRT models) parameter estimation. Through a simulation study, we demonstrate that Variational Approximation offers precise results in high-dimensional, small-sample settings while significantly reducing computational time.

Discussant:

*Joseph A. Martineau, ETS*

### 030. Explorations of GPT

#### Paper Session

10:05 to 11:05 am

Convention Center: Level 1 - 124

Chair:

*Paulius Satkus, GMAC*

Participants:

#### Assessment of University Students' Essays through OpenAI GPT

*Ayfer Sayin, Gazi University; Deniz Melanlioglu, Istanbul University*

This study examines the applicability of the OpenAI GPT for scoring essays. GPT scores exhibit positive and moderate correlation with expert scores. It is observed that GPT scores content well but has difficulties in scoring text organization. The research underlines the applicability of automatic scoring in languages other than English.

#### Bridging Bilingual Gaps: An In-depth Exploration of ChatGPT's Transadaptation Capabilities

*Xingyao Xiao, Berkeley School of Education, University of California, Berkeley; Pooya Razavi, Edmentum, Inc; Sonya Powers, Edmentum, Inc.*

This research examines ChatGPT's prowess in transadapting English to Spanish educational content, specifically assessment items in ELA and mathematics. Contrasting ChatGPT outputs with human benchmarks, we gauge linguistic accuracy, contextual relevance, and efficiency. Initial results hint at potential AI improvements and emphasize the delicate equilibrium between rapidity and quality.

#### Evaluating the Performance of GPT-4 on the GMAT Exam Questions

*Yanyan Fu, GMAC; Paulius Satkus, GMAC; Kyung (Chris) T. Han, GMAC*

A recent study found that the language model GPT-4 could achieve above the 90th percentile in major assessments (e.g., SAT, LSAT, and GRE). This study examined GPT-4's performance on the GMAT exam and found that it can achieve the 93rd and 66th percentiles on the Verbal and Quant sections, respectively.

Discussant:

*Jiangang Hao, ETS*

### 031. Scaling, Linking, and Equating

#### Paper Session

10:05 to 11:05 am

Convention Center: Level 1 - 125

Chair:

*Leslie Keng, National Board of Medical Examiners*

Participants:

#### A Comparison of Robust Linking Methods with Multiple Outlying Common Items

*Seongeun Kim, University of North Carolina at Greensboro; Kyung Yong Kim, University of North Carolina Greensboro*

Common items (CIs) play a crucial role in scale linking. When existing outlying CIs, retaining them is recommended to maintain content balance and psychometric characteristics rather than removing them. Robust linking methods, reducing the influence of outlying CIs have been suggested, and those are compared through simulation under various conditions.

#### Selecting Parameters in Cubic Spline PostSmoothing Using Cross-Validation

*Stella Kim, University of North Carolina at Charlotte; Won-Chan Lee, University of Iowa; Hwanggyu Lim, GMAC; Yeonwho Kim, Seoul National University*

The purpose of this study is to propose a new approach to selecting a smoothing parameter in cubic-spline functions in equating using a cross-validation technique. Results from the proposed method are evaluated in comparison to various levels of parameters typically used in cubic-spline postsmoothing and log-linear presmoothing through simulation studies.

### **Vertical Scale Implications When Accelerated Students are Not Double Tested**

*Elizabeth Ayers-Wright, Cambium Assessment, Inc.; Hongwook Suh, Cambium Assessment, Inc.; Mario Gonzalez, Cambium Assessment*

When states have provisions that allow all students to prepare for and take advanced mathematics coursework, eighth grade students are not required to double test. We explore the effect of vertical scale recovery when accelerated eighth grade students test complete an end-of-course assessment instead of the Grade 8 assessment.

Discussant:

*Tim Moses, College Board*

### **032. Using Generative AI in Education: Results from Large-Scale EdTech Platform R&D**

#### **Coordinated Paper Session**

**10:05 to 11:05 am**

**Convention Center: Level 4 - Terrace Ballroom III**

Generative Artificial Intelligence (GenAI) made a major technical breakthrough in 2022 with the public release of ChatGPT, DALL-E, LLAMA-2 and other applications that rely on large language models (LLM) to create new text/images in response to human prompts. This technology is early in the hype cycle, but is having substantial impacts in many industries, and education is one of the most promising areas identified for its application. Measurement experts and assessment service providers are familiar with LLMs and use them frequently in automated scoring and other contexts. However, GenAI brings significant challenges to foundational principles in measurement. LLMs are inherently difficult to explain – even by the people creating them. Their training data is neither representative nor static, and in some cases is not even disclosed (e.g. GPT-4). In this context, we suggest that while the concepts of measurement are still important, their practice requires a change in methods. This session seeks to pierce the hype and hyperbole surrounding this new technology through papers that describe Item Development with Duolingo, Tutoring Dialogue with Khanmigo, Automated Distractor Generation & Item Feedback with EEDI, Math Feedback with NAEP Data, and interpretability research using eRater/Textevaluator. This session is organized through NCME's Generative AI SIGIME.

Session Organizer:

*John Whitmer, Federation of American Scientists*

Moderator:

*Chris Ormerod, Cambium Assessment, Inc.*

Participants:

**Automated Item Development of Complex Reading and Listening Tasks for Language Assessments**

*Andrew Runge, Duolingo*

**Automated Distractor and Feedback Generation for Math Multiple-choice Questions**

*Andrew Lan, University of Massachusetts at Amherst*

**Providing dialogue-based stepwise feedback in mathematics**

*Kristen DiCerbo, Khan Academy*

**Creating Partially Interpretable Embeddings From Large Language Models Using Best Interpretable Orthogonal Transformations**

*Paul Deane, ETS*

Discussant:

*Magdalen Beiting-Parrish, Federation of American Scientists*

### 033. Innovation Demonstration I

#### Paper Session

10:05 to 11:05 am

Convention Center: Level 4 - Terrace Ballroom IV

Participants:

#### **Aberrance: An R Package For Detecting Aberrance In Test Data**

*Kylie Gorney, Michigan State University; Jiayi Deng, University of Minnesota*

Aberrant behavior refers to any type of unusual behavior that would not be expected under normal circumstances. In this demonstration, we introduce an R package for detecting several types of aberrant behavior, including preknowledge, answer copying, rapid guessing, and more.

#### **Collaborative Problem Solving Training Program: Arming the 21st Century**

*Carol McGregor Forsyth; Jessica Andrews Todd, ETS; Emily Kerzabi, ETS; Guangming Ling, ETS*

This paper presents a Collaborative Problem Solving (CPS) training program developed within a facilitation system known as EPCAL, with the aim of providing individuals with much-needed training and assessment on how to be effective collaborative problem solvers. This system will be presented as a hands-on activity.

#### **ShinyBHB: A Shiny App for Bayesian Historical Borrowing in Large-Scale Assessments**

*Weicong Lyu, University of Washington, Seattle; Shaojie Ye, Pinterest; Jianshen Chen, College Board; Sinan Yavuz, University of Wisconsin-Madison; David Kaplan, University of Wisconsin, Madison*

ShinyBHB is a comprehensive application package built on the R package Shiny to implement Bayesian historical borrowing for single-level, multilevel, and longitudinal data. It covers Bayesian models with no borrowing, pooling, aggregated data priors, power priors, commensurate priors, and dynamic priors, with flexible model building and various output formats available.

#### **A Human-Centered AI-Augmented Approach To Contextualizing Test Performance On Large-Scale Assessments**

*Hongwen Guo, ETS; Matthew S Johnson, ETS; Luis Saldivia, ETS; Michelle Worthington, ETS; Kadriye Ercikan, ETS*

Process data, collected from large-scale assessments, offer opportunities to explore students' test-taking processes, but manually labeling them is hard to scale. We proposed a human-centered AI approach to make sense of process data to contextualize performance, and to help identify sources of low performance and generate rich feedback.

#### **Generative AI Chatbot for Large-Scale Assessment Data Analysis**

*Sinan Yavuz, American Institutes for Research; Blue Webb, American Institutes for Research; Paul Bailey, American Institutes for Research; Ting Zhang, American Institutes for Research; Yuqi Liao, American Institutes for Research; Emmanuel Sikali, NCES*

We fine-tuned the generative AI model and developed specific chatbots to help coding for large-scale assessment (LSA) data analyses. The chatbot focuses on helping with NCES LSA data analysis programming with the R package EdSurvey. We focused on methods that write the codes for analyses and debug the written codes.

#### **Tensor-Based Scoring For Cognitively Diagnostic Models**

*Russell G Almond, Florida State University*

The basic scoring operations for Cognitively Diagnostic models are expressed in terms of tensors, multidimensional arrays. This allows the use of computing optimization tools such as tensorflow or PyTorch in scoring and calibration of CDMs.

### 034. Rater Models

#### Paper Session

11:25 am to 12:25 pm

Convention Center: Level 1 - 120B

Chair:

*Xingyao Xiao, Berkeley School of Education, University of California, Berkeley*

Participants:

#### **Evaluating Ratings in a Small-Sample Speaking Assessment: A Hierarchical Rater Model Approach**

*Mirai Nagasawa, The University of Alabama; Stefanie A. Wind, University of Alabama*

We examined rater effects in a small-sample speaking performance assessment using a Bayesian Hierarchical Rater Model (HRM) approach. Results suggested that the HRM is effective for evaluating ratings with small sample sizes and nested structures and provides insight into rater effects related to rubric criteria.

#### **Rating Scale Model in Diagnostic Classification Modeling**

*Ren Liu, UC Merced; Alexandra Levin Kislyonkova, UC Merced*

The study introduces a new Diagnostic Classification Model (DCM) suitable for rating scalelike instruments. The model uses the Rating Scale Model and accommodates it to the assumptions of the DCM framework. The new model is tested in three studies: operational study, simulation study, and model comparison study.

#### **Modelling the Complexity of Competency Assessment Using Facet Analysis**

*Cuc Thi Kim Nguyen, University of Melbourne; Zhonghua Zhang, University of Melbourne*

This presentation addresses the challenge of aggregating and calibrating student abilities for complex competency assessment data involving variety of real-life and divert contexts. The study employs Australian data for assessment of the competency “knowing how to learn” in 2021 as a case study.

Discussant:

*Jodi M Casabianca, ETS*

### 035. Reliability Estimation

#### Paper Session

11:25 am to 12:25 pm

Convention Center: Level 1 - 120C

Chair:

*Xia Mao, NBOME*

Participants:

#### **A Simple procedure for Evaluating the Tau Equivalence Assumption for Coefficient Alpha**

*Fen Fan, NBME; Chunyan Liu, National Board of Medical Examiners; Peter Baldwin, National Board of Medical Examiners; Brian Clauser, National Board of Medical Examiners*

In this paper we will present a simple procedure to evaluate the extent to which violation of the tau equivalent assumption of coefficient alpha inflates the estimated reliability. The procedure is appropriate for use whenever there is systematic nesting of test items (e.g., reading comprehension, science experiment, or medical case).

#### **Extent of Reliability Overestimation for Self-reports when Ignoring Sources of Measurement Error**

*Walter Vispoel, University of Iowa; Hyeri Hong, California State University, Fresno; Hyeryung Lee; Tingting Chen, University of Iowa*

We illustrate how conventional reliability coefficients for self-reports routinely overestimate overall consistency of scores and describe effective techniques to address these issues. We also direct readers to online resources that would enable them to apply and extend those techniques.

### **Recursive IRT-based Reliability Estimation** Tim Moses, College Board;

*Young Koung Kim, College Board*

IRT-based reliabilities are traditionally estimated from theoretical assumptions. This proposal considers alternatives obtained through modifications of the Lord-Wingersky recursion algorithm, that more directly reflect observed scores and distributions, and sampling errors. The proposed method is evaluated over common IRT models and several large-volume tests. Additional applications will also be described.

Discussant:

*Howard Everson, Graduate Center, City University of New York*

### **036. Surviving and Understanding the US Department of Education's State Assessment Peer Review**

#### **Organized Discussion**

**11:25 am to 12:25 pm**

**Convention Center: Level 1 - 121A**

The US Department of Education's peer review of state assessments is much more than an exercise in "checking off the boxes and providing evidence." This organized discussion will present recommendations for moving the current process towards greater efficiency and benefit. With improved understanding and collaboration, the peer review process can serve as a means of discovering and monitoring student inequities. Panelists will provide an overview of the peer review process and leave attendees with a visual takeaway that outlines the tasks, timelines, roles, and responsibilities for peer review. Peer review experienced panelists will provide advice for providing evidence and surviving the challenges of peer review. The session will dive into the depths of the peer review process by addressing the following critical questions: 1) How can we leverage the peer review process to provide evidence that reduces educational inequities for students with disabilities and our historically marginalized populations? 2) What role and responsibility does the assessment vendor/provider have in the peer review process? 3) What does a more valuable and effective state assessment peer review process look like? 4) How can collaboration with assessment designers, equity champions, and local test administrators help improve the process?

Session Organizer:

*Regina A. Lewis, Maine Department of Education*

Presenters:

*Donald Peasley, US Department of Education*

*Andrew J. Middlestead, Michigan Department of Education*

*Zachary Warner, New York State Education Department*

*Darin Kelberlau, Millard Public Schools*

*Trent Workman, Pearson*

Discussant:

*Ellen Forte, edCount, LLC*

### **037. Longitudinal Credentialing**

#### **Paper Session**

**11:25 am to 12:25 pm**

**Convention Center: Level 1 - 121B**

Chair:

*Seongeun Kim, University of North Carolina at Greensboro*

Participants:

#### **Detecting Compromised Items in Longitudinal Assessments through Sequential Procedure**

*Chansoon Lee, American Board of Internal Medicine*

This study employs an IRT-based sequential procedure to detect compromised items in longitudinal assessments. The approach proves effective in identifying potentially compromised items and supporting test security in a longitudinal assessment program.

**Equity and Longitudinal Assessments: Perspectives from Physician Assistants/Associates (PAs) Participating in PANRE-LA**

*Andrzej Kozikowski, National Commission on Certification of Physician Assistants (NCCPA); Joshua Goodman, National Commission on Certification of Physician Assistants (NCCPA); Andrew Dallas, National Commission on Certification of Physician Assistants (NCCPA)*

Ensuring assessments are fair and equitable is a fundamental validity issue. Longitudinal assessments (LAs) may offer more flexibility and unique opportunities to enhance equity. We evaluated potential differences in perspectives and experiences of Physician Assistants/Associates participating in an LA program by a variety of demographic and practice characteristics.

**Response Time Patterns and Physician Performance in Longitudinal Assessment for Continuing Certification**

*Youngjun Lee, The American Board of Anesthesiology; Emily Karen Toutkoushian, The American Board of Anesthesiology; Huaping Sun, American Board of Anesthesiologists*

This study aims to explore relationships among response time, item difficulty, and physician performance in longitudinal assessment for continuing certification. Shorter response time is associated with better physician performance. Four distinct physician groups of test-taking behaviors are identified from cluster analysis. Implications about the “less-engaged group” are discussed.

Discussant:

*Irina Grabovsky, NBME*

**038. Computational Thinking and Problem Solving****Paper Session**

**11:25 am to 12:25 pm**

**Convention Center: Level 1 - 121C**

Chair:

*Natasha Jayne Williams, Curriculum Associates*

Participants:

**Assessing Computational Thinking for 9th Grade Students in Rural High Schools**

*Kathleen C Haynie, Haynie Research and Evaluation; Jerry L. Gorham, Tropical Gulf Measurement*

The Computational Thinking Assessment for High School (CTA-HS) was developed through the federally-supported Pathways for Alabama Computer Science (PACS) project. CTA-HS is utilized twice in rural schools across Alabama: with 9th graders as a Year 1 outcome, and with those same students in 11th grade as a Year 3 outcome.

**Measuring Computational Thinking Instruction: Reliability Of A New Classroom Observation Instrument**

*Maria-Paz Fernandez, UCLA; Imelda Nava, UCLA; Jose Felipe Martinez, UCLA - School of Education and Information Studies; Deborah La Torre, UCLA; Leticia Perez, WestEd; Justin Betzelberger, UCLA*

The study investigates the reliability of an observation rubric intended to assess and provide formative feedback to improve instructional practices in Computational Thinking. We present the results of our examination of the reliability of ratings of teaching practice generated using the rubric and discuss the various sources of measurement error.

**Designing Collaborative Problem Solving Assessment Tasks to Elicit Valuable Workplace Skills**

*Jessica Andrews Todd, ETS; Carol McGregor Forsyth; Emily Kerzabi, ETS*

Collaborative problem solving (CPS) skills are important for workplace success, but little research targets developing research-driven tools that support the development of CPS skills with formative assessment tasks in higher education contexts. We describe a CPS training program and evaluate the extent to which formative assessments tasks elicit targeted skills.

Discussant:

*Wayne J. Camara, LSAC*

### 039. IRT Calibrations

#### Paper Session

11:25 am to 12:25 pm

Convention Center: Level 1 - 122A

Chair:

*Yi-Fang Wu, Cambium Assessment, Inc.*

Participants:

#### “Don’t Know About ‘I Don’t Know’ ”?: Handling Such Responses in Likert Scales

*Teresa Ober, ETS; Caitlin Tenison, ETS; Jesse R. Sparks, ETS*

We examine different methods for scoring “I don’t know” (IDK) responses in Likert self-report measures of prior knowledge of a topic relevant to a scenario-based assessment. Our findings indicate IDK responses may be treated as missing or neutral responses though additional information is gained by accounting for IDK response tendencies.

#### CAT Item Bank Calibration with New Structured Matrix Completion Based Methods

*Yawei Shen, Pearson; Shiyu Wang, University of Georgia*

To address the challenges of calibrating an item bank with small sample sizes, this study proposed new methods based on structured matrix completion without assumptions for the missing mechanism. From a simulation study, the proposed methods led to better item parameter estimations than the baseline methods.

Discussant:

*Ruben Castaneda, College Board*

### 040. School Accountability Models in the USA and Abroad: Past, Present, Future

#### Coordinated Paper Session

11:25 am to 12:25 pm

Convention Center: Level 1 - 122B

Persistent achievement gaps, exacerbated by COVID shutdowns call to question the effectiveness of ESSA mandated school accountability models. Critics claim they have “underachieved,” and may actually do more harm than good. The strongest criticisms are they are too rigid and summative assessment heavy and are indifferent to the differential history, culture, and values of states, schools and the diverse communities they serve. Many claim they are biased and unfair, particularly to schools with diverse, under-represented student populations. To reimagine a fair, robust, and flexible accountability model that fits the needs of all students and schools, this coordinated paper session begins by looking backwards: how did USA school accountability get to this point? Paper 1 examines the history of school accountability in the USA. Paper 2 contrasts school accountability models across several English-speaking countries, including Australia, Canada and England. Several states are in the process of revising both their assessment and accountability models, post-COVID, and with the goal of greater inclusion, fairness and flexibility. Paper 3 examines these trends relative to the shortcomings of current models and promise of future success. Paper 4 examines the complex process a particular state used to build a new inclusive school accountability system.

Session Organizer:

*Stanley N Rabinowitz, EdMetric LLC*

Participants:

#### History of School Accountability in the USA

*Jami-Jon Pearson, Measurement Incorporated*

#### New Ideas for School Accountability

*Hillary Michaels, HumRRO*

#### Lessons Learned Designing an Inclusive State Accountability Model

*Charity Smith, Fetterman & Associates*

Discussant:

*Karla Egan, EdMetric LLC*

### 041. Response Times

#### Paper Session

11:25 am to 12:25 pm

Convention Center: Level 1 - 123

Chair:

*Tony Albano, University of California, Davis*

Participants:

#### **A Conceptual Psychometric Model Explaining Residual Dependencies Between Response Accuracy and Time**

*Weicong Lyu, University of Washington, Seattle; Daniel Bolt, University of Wisconsin, Madison*

We illustrate how consistent empirical patterns in the dependencies between response accuracy and response time residuals can be explained by presence of exogenous item specific factors. One important consequence is that psychometric differences associated with fast and slow responses may be artifacts and not manifestations of varying psychological response processes.

#### **Cognitive Processes on Achievement Tests: Intra-Person Differences in Item Response Times Correlations**

*Susan Embretson, Georgia Institute of Technology*

Several item response theory (IRT) models incorporate response times (RT) to increase interpretability and precision of trait level scores. If examinees differ in strategies for using RT, the joint model assumptions may not be plausible. In the current study, mathematical achievement tests are examined for within-examinee differences in RT relationships.

#### **Using Response Times in Answer Similarity Analysis**

*Kylie Gorney, Michigan State University; James Wollack, University of Wisconsin*

We expand the definition of answer similarity to include not only the item scores/responses, but also the item response times. Using detailed simulations and an experimental data set, we show that the new statistics are much more powerful, on average, than existing statistics at detecting several types of simulated collusion.

Discussant:

*Ben Domingue, Stanford University*

### 042. Applications of AI

#### Paper Session

11:25 am to 12:25 pm

Convention Center: Level 1 - 124

Chair:

*Yanyan Fu, GMAC*

Participants:

#### **Fine-Tuned Word Embeddings for Enemy Item Detection**

*Liu Liu, University of Washington; Marcus Walker, National Commission on Certification of Physician Assistants*

This study evaluates GloVe 6B and Clinical BERT for detecting enemy items in a certification item bank. Using word embeddings and machine learning, we aim to improve test precision and validity. Fine-tuned embeddings on specialty corpora contain context relevant to the specialty area with the potential for improved prediction accuracy.

#### **Training AI to Generate Human-Like Item Responses for Field-Testing**

*Hotaka Maeda, Smarter Balanced*

We propose a method of using large-language models to generate human-like item responses by training the model on simulated item response data. Results showed moderate levels of success in using AI-generated responses to recover item parameters. If refined further, AI could potentially replace human test-taker responses for field-testing new items.

# FULL SCHEDULE

FRIDAY, APRIL 12

## Using OpenAI GPT to Generate Reading Comprehension Items

*Ayfer Sayin, Gazi University; Mark J Gierl, University of Alberta*

The purpose of this study is to introduce and evaluate a method for generating reading comprehension items using automatic item generation. The model generates reading comprehension items where the student is required to identify one irrelevant sentence in passage that contains 5 sentences. The passages were generated using OpenAI GPT-3.

Discussant:

*Luke Weisman Miratrix, Harvard Graduate School of Education*

## 043. Situational Judgment Tests

### Paper Session

**11:25 am to 12:25 pm**

**Convention Center: Level 1 - 125**

Chair:

*Yuan Ge, The College Board*

Participants:

### Advancing the Conceptualization and Measurement of Coachability through Novel Situational Judgment Tasks

*Kevin Williams, ETS; Teresa Ober, ETS; Devon Kinsey, ETS; Wyman Brantley, ETS*

Coachability, which describes how one seeks, engages with, processes, and reacts to feedback and other learning opportunities, is associated with academic, professional, and personal growth. We conducted two studies evaluating a novel coachability situational judgment task, in which we expanded coachability's conceptualization to include growth mindset and reactions to praise.

### Evaluating Rater Bias Effects In Constructed-Response Scoring In Situational Judgment Tests

*Jinnie Shin, University of Florida; Zeyuan Jing, University of Florida; Alexander MacIntosh, Acuity Insights*

We analyzed 10,536 responses to situational judgment test items using the generalized many-facet Rasch model. This study aimed to identify common argument structures and evaluate rater biases in scoring, considering traits like consistency, severity, and restrictiveness. The research provides insights into refining assessment methodologies.

### Evaluating Argumentation Effectiveness Using DeBERTa Token Classification Approach

*Lingchen Kong, University of Florida; Alexander MacIntosh, Acuity Insights; Jinnie Shin, University of Florida*

This study implements DeBERTa model for evaluating the quality of argument elements in Situational Judgment Tests. Utilizing transfer learning techniques, the model demonstrated high accuracy and agreement in assessing the quality key argumentative elements like "Position" and "Claim." The approach has significant implications for enhancing the evaluation processes of SJTs.

Discussant:

*Patrick Charles Kyllonen, ETS*

## 044. NCME Career Award

### Invited Session

**11:25 am to 12:25 pm**

**Convention Center: Level 4 - Terrace Ballroom III**

Moderator:

*Derek Christian Briggs, University of Colorado Boulder*

Presenter:

*Henry Braun, Boston College*

Discussant:

*Sandip Sinharay, ETS*

### 045. GSIC Poster Session I

#### Graduate Electronic Board Session

11:25 am to 12:25 pm

Convention Center: Level 4 - Terrace Ballroom IV

Participants:

**1. Adapting a Scale to Measure Perceived Father and Mother Autonomy Support Separately**

*Yuqing Zou, University of Iowa; Yichong Cao, University of Iowa; Chunrui Zou, University of Pennsylvania*

We adapted a well-established scale to measure perceived father and mother autonomy support (FAS and MAS) separately and validated it. The factor analysis results showed that perceived FAS and MAS are distinct factors with adequate validity, reliability, model fit, and model invariance. These findings have implications for empirical measurement practices.

**2. A MFRM Study Comparing Human and Machine Raters in Scoring Oral Reading**

*Meng-Hsun Lee, University of Toronto*

The study assessed the consistency of scoring students' oral readings between human and machine raters. MFRM results indicated an acceptable fit for all raters, with infit and outfit values ranging from 0.50 to 1.50. The machine rater showed significant bias when rating English-L1 students.

**3. Assessment Alchemy: Lessons from Global South on Blending Culture, Participation and Validity**

*Ketan, University of Massachusetts Amherst*

This study delves into cultural responsiveness in educational assessments, drawing from Participatory Action Research (PAR) and Multicultural Validity. It examines ways to integrate PAR and Multicultural Validity principles and derives insights from India's ASER survey to offer actionable recommendations for educators and policymakers striving for equitable, inclusive education.

**4. Comparing CGS and PME on Estimating DCM Attribute Mastery Profiles**

*Jingyang Li, University of Georgia; Matthew James Madison, University of Georgia; Zhenqiu Laura Lu, University of Georgia*

Diagnostic Classification Models have emerged as valuable tools for assessing individuals' skills and attributes in various fields. In estimating attribute mastery profiles in DCMs, Collapsed Gibbs Sampling and Posterior Mode Estimation are two reliable methods. This research compares the performance of CGS and PME by conducting simulation studies under various conditions.

**5. Complementing Borich Needs Assessment with Text Mining by Natural Language Processing**

*Yewon Kim, Ewha Womans University; Ja-yoon Lee, Seoul National University; Jaehee Noh, Seoul National University; John Park, United Nations University*

This study explored a possibility of complementing results of Borich needs assessment model by implementing results of text mining by Natural Language Processing followed with principal component analysis, and co-occurrence analysis. The results show that NLP can provide complementary explanations that lead to better understandings of Borich needs assessment model.

**6. Continuous Effect Loglinear Diagnostic Classification Model: Incorporating General Continuous Effects in LCDMs**

*Sergio Haab, University of Iowa*

In this proposal, a new model incorporating a continuous effect into the log-linear cognitive diagnosis model (LCDM) framework is introduced. The results from the new model suggest that most researchers looking at real data would benefit by acknowledging an underlying continuous effect.

**7. Diagnostic Accuracy Of French Screeners With English Criterion Measures For Fi Bilingual Students**

*Songtao Wang, OISE/University of Toronto; Krystina Raymond, University of Toronto; Becky Chen, University of Toronto; Kathleen Hipfner-Boucher, Research Officer*

This study assessed phonological awareness in Grade 1 Canadian French immersion students, emphasizing its role in bilingual reading. Using TFHFL screenings, it found demographic variations in identifying at-risk readers, particularly among ELLs and indigenous students. The findings underscore the need for tailored assessment tools in diverse educational settings.

- 8. Enhancing Precision in the Generalized Graded Unfolding Model (GGUM) Using Multiple Responses**  
*Na Liu, Georgia Institute of Technology; James S. Roberts, Georgia Institute of Technology*  
This research presents a hybrid model combining the GGUM and method of successive intervals (MSI) in order to reduce sample size demands of the GGUM. Both real and simulated data are employed to validate the efficiency and applicability of this hybrid model in evaluating non-cognitive psychological constructs.
- 9. Evaluating Subgroup Analysis Indices and Guidelines for Automated Scoring Algorithm**  
*Yichi Zhang, University of Southern California; Edward W Wolfe, Pearson*  
There is limited research regarding operational use of agreement indices to detect subgroup differences between automated and human scores. Our study evaluates the accuracy of this application and proposed guidelines based on results of a Monte Carlo simulation.
- 10. Exploring the Machine Learning Approach to Evaluate Validities of A Game-Based Assessment**  
*Xinyue Deng, Fordham University; Heining Cham, Fordham University*  
This study uses machine learning techniques to assess a cognitive GBA's validities, utilizing data from Landers et al. (2021). We used random forest models and variable importance calculations to support the GBA's construct and criterion-related validities. The results showed the promising potential of utilizing machine learning in the GBA field.
- 11. Impact of Panel Designs in Multidimensional MST with Within-item Multidimensional Structure**  
*Xi Wang, The University of Iowa; Anthony D. Fina, University of Iowa*  
This study investigates the implementation of a multidimensional-MST with a within-item multidimensional structure. Specifically, simulations are used to examine the correlation between dimensions, the routing stage length, and the number of stages within a panel, as well as the effects of these factors on measurement precision and error.
- 12. Improving Automated Scoring of Prosody of ORF with Between-Word Silence Features**  
*Kuo Wang, Southern Methodist University; Akihito Kamata, Southern Methodist University; Yusuf Kara, Southern Methodist University; Joseph F. T. Nese, University of Oregon*  
This study aimed to use between-word silence times and acoustic features for the automated scoring of prosody in oral reading fluency assessment. The results demonstrated that features related to between-word silence time helped improve the classifier's performance with Elastic Net regularization.
- 13. Local Independence and Q3 Values: How much Matters?**  
*Kelsey Nason, James Madison University; Christine DeMars, James Madison University*  
A value of .2 on Yen's Q3 (1993) is often used to identify violations of local independence large enough to be of concern. Preliminary results suggest that even Q3 values below .2 can yield biased reliability and parameter estimates, with impacts slightly more severe in larger groups of dependent items.
- 14. Mediation Effect of ICT Attitude on Academic Performance: TIMSS 2019 Grade 8**  
*Dukjae Lee, University of Virginia; Lisa Keller, UMASS Amherst*  
Exploring three Nordic countries, this study found that students' more frequent usage of digital devices was associated with positive attitudes when using them (ICT Attitude), but negatively affected their academic performances in TIMSS 2019 Grade 8. How the survey questions regarding ICT Attitude written might be one reason.
- 15. On Some New and Improved Measures for Item Analysis**  
*Rachel Lee, Columbia University*  
Classical item analysis (CIA) has a long history and is still used today by testing companies, in classrooms, and in online-testing software. A new perspective on CIA is provided by signal detection theory, which suggests simple improvements to existing measures and yields some new and improved measures for item analysis.

### 16. Priors and Evidence: Introducing Bayesian Reasoning to Standard Setting

*Sarah Alahmadi, James Madison University; Chad W. Buckendahl, ACS Ventures, LLC*

Setting performance standards is critical for the validity of the intended interpretation and use of test scores. Various qualitative concerns with current standard setting methods have been raised by measurement and content experts. We propose a new approach applying the mathematics and reasoning of Bayes' Theorem to standard setting.

### 17. Reconceptualization of Educational Measurement in Light of Intersectionality

*Anna Yinqi Zhang, Penn State University*

The study draws on literature on intersectionality and its application in measurement to address diverse learners' needs. We analyzed theoretical and empirical studies from five databases and argue that integrating the theoretical framework of intersectionality with existing measurement theory and methodology is possible and make recommendations for research and practice.

### 18. Synergizing Educators' Data Attitudes and Digital Teaching Proficiency in Higher Education

*Kuixi Du, University of Southern Mississippi; Thomas J Lipscomb, The University of Southern Mississippi; Kyna J Shelley, The University of Southern Mississippi*

This research investigates how Mississippi educators' attitudes toward data utilization impact student capacity development. It analyzes the interplay between data skills, digital teaching, collaboration, and support to enhance post-pandemic education using SEM modeling. Our results aim to strengthen educators' data literacy, integrating technology to foster student success in the educational landscape.

### 19. Test Length's Impact on Ability Estimates: Considering Cognitive Fatigue and Speededness

*Tram-Anh Tran Nguyen, University of Massachusetts Amherst; Lisa Keller, UMASS Amherst*

While increasing test length can offer broader content coverage and provide more accurate estimates of students' proficiency, it may also induce cognitive fatigue and speededness. By using simulation, this study examines the effects of increased test length on proficiency estimates while considering the potential influence of cognitive fatigue and speededness.

### 20. The Impact of Different MST Calibration Designs on Parameter Estimation

*Jing Ma, The University of Iowa; Xi Wang, The University of Iowa; Mubarak O Mojoyinola, The University of Iowa; Ahmed Bediwy, The University of Iowa; Anthony D. Fina, University of Iowa*

This simulation study investigates the impact of various calibration methods on item parameter estimation in multistage testing (MST). Four calibration methods are examined, revealing differences in parameter recovery performance across two different ability distributions and panels.

### 21. Uncovering Test-Takers' Problem-Solving Patterns through Hidden Markov Models and Network Analysis

*Xiaoxiao Liu, University of Alberta; Okan Bulut, University of Alberta; Ying Cui, University of Alberta; Yizhu Gao, Government of Alberta*

Process data unveils hidden cognition processes during problem-solving. This study used the Programme for the International Assessment of Adult Competencies (PIAAC) data to identify examinees' problem-solving patterns. Utilizing Hidden Markov models and Gaussian graphical analyses, we found coherent patterns in the correct group, contrasting erratic patterns in the incorrect group.

### 22. Using Network Psychometrics to Analyze Student Misconceptions in Academic Tests

*Jacinta Olson, University of Iowa; Jonathan Templin, University of Iowa*

Typical analyses of educational assessments only use the information of whether a student got a question correct, losing all information told by which options they did not choose. In this proposal, we introduce the use of network analysis to investigate how option-level information can be examined.

**23. Tracing Students' Practicing: Effects of a Math Learning Program on Anxiety-Performance Link**  
*Anna Hilz, Leibniz Institute for Science and Mathematics Education; Abe Hofman, University of Amsterdam; Brenda Jansen, University of Amsterdam; Karen Aldrup, IPN - Leibniz Institute for Science and Mathematics Education*

This study (N=880 students) aims to investigate if math anxiety's negative effect on performance is attributed to avoidance behavior with a math learning program. When tracing students' practicing over 45 weeks, math anxious students practiced equally. Increased practice resulted in better performance, suggesting math anxious students benefit from such programs.

**052. Testing Standards Revision Updates and Membership Input (Joint AERA/NCME Session)**  
**Organized Discussion**

**11:25 am to 12:25 pm**

**Convention Center: Second Floor - 201C**

The Standards for Educational and Psychological Testing (Standards) Management Committee (SMC) representing AERA, APA, and NCME, has begun the process to revise the Standards. At the NCME/AERA joint session on the Standards in April 2023, Kristen Huff and Michael Rodriguez, the SMC representatives from NCME and AERA respectively, presented results of the preliminary survey of NCME and AERA members regarding priorities for revision of the Standards. Since then, the SMC appointed the joint committee (JC) co-chairs and members. The JC has been charged to undertake the revision and they began that process early in 2024. The JC co-chairs will lead a brief presentation on the revision process and summary of their work to date. Members of the JC will join them and open the session inviting input from members of NCME and AERA.

Presenter:

*Michael C. Rodriguez, University of Minnesota*

Presenter:

*Kristen Huff, Curriculum Associates*

**046. Towards A New Framework For Measurement In Higher Education Admissions Decision-Making**

**Coordinated Paper Session**

**1:15 to 2:45 pm**

**Convention Center: Level 1 - 120B**

Postsecondary admissions is a process whereby schools determine what they value (i.e., the desired skills and traits of their student body), and measure applicants accordingly. In the U.S., the admissions process is experiencing two major paradigm shifts affecting how students are evaluated; the recent Supreme Court ruling on affirmative action joins the pandemic and accompanying escalation of test-optional and holistic approaches to admissions in greatly changing what schools can measure about applicants, and how. With these shifts comes the opportunity to reflect on the changes that are occurring and to consider the best way forward for higher education admissions – how can schools measure in ways that meet their goals fairly and equitably? In this session, we explore student and admissions professionals' perspectives on testing in admissions and examine the evidence that has been collected so far on the effects of test-optional admissions. We also explore the well-used but less well-defined term “holistic admissions,” bringing this concept and associated practices into focus and laying out best practices for the holistic measurement of applicants to higher education institutions. Discussion will center on measurement insights for admissions practices, and potential innovations for the new admissions landscape.

Session Organizer:

*Marisol Kevelson, ETS*

Chair:

*Sara Haviland, ETS*

# FULL SCHEDULE

FRIDAY, APRIL 12

Participants:

**Graduate School Admissions Professional, Applicant, and Faculty Perspectives on Admissions Tests**  
*Sara Haviland, ETS; Marisol Kevelson, ETS; Michael E. Walker; Jose de Jesus Sotelo, ETS*

**Is Test-Optional Enough for Equity?**

*Reginald M Gooch, ETS; Vinetha Belur, vbelur@ets.org; Sara Haviland, ETS; Ou Lydia Liu, ETS*

**Unlocking Holistic Admissions: Scan Of Criteria And Procedures Used In Graduate Degree Programs**

*Teresa Ober, ETS; Olasumbo Oluwalana, ETS; Huidi Yang, NYU; Reginald M Gooch, ETS*

**In Theory And Practice: Exploring How Admissions Officers Define And Experience Holistic Admissions**

*Sara Haviland, ETS; Vinetha Belur, vbelur@ets.org*

Discussant:

*Joseph Paris, West Chester University*

## 047. Innovations in the Measurement of 21st Century Skills

**Coordinated Paper Session**

**1:15 to 2:45 pm**

**Convention Center: Level 1 - 122B**

Developing and measuring skills beyond the traditional core domains in education, such as 21st Century Skills (21CS), can enhance diversity and equity in education and the workforce by offering a wider range of strengths for individuals to capitalize on in pursuit of their best possible future. It is crucial for students to develop and master 21CS, and assessment plays a vital role in providing insights into their proficiency and progress. Our session delves into cutting-edge advancements in the evaluation of 21CS, both for individual learners and groups. We begin by broadening our perspective on 21CS in the context of generative AI, contemplating current frameworks and the possible integration of additional skills, as well as measurement considerations. Next, we examine a new multidimensional approach for measuring and nurturing creative thinking skills in a play-based learning environment. We then investigate innovative findings on the stability and generalizability of individual collaborative problem-solving skills across various task types. Additionally, we explore how Cognitive Diagnostic Modeling can be employed to gain insights into student mastery of 21CS within existing assessments. Lastly, we will discuss the potential of AI in streamlining the labor-intensive and time-consuming aspects of assessment development, particularly in generating critical thinking test items.

Session Organizer:

*Kristin Lansing-Stoeffler, ACT, INC.*

Chair:

*Alina A. von Davier, Duolingo*

Participants:

**Reexamining 21st Century Skills in the Era of Generative**

*AI John Weiner, Lifelong Learner Holdings*

**A Multidimensional Method For Measuring And Fostering Creative Thinking Skills In Learning-Through-Play Setting**

*Yigal Rosen, BrainPOP; Garrett Jaeger, LEGO Foundation; Michelle Newstadt, BrainPOP; Sara Bakken, BrainPop; Ilia Rushkin, BrainPOP*

**Characterizing Individual Collaborative Problem-Solving Through Conditional Transition Profile: Stability Across Tasks And Teams**

*Jiangang Hao, ETS; Patrick Kyllonen, ETS, Wenju Cui, ETS, Emily Kerzabi, ETS*

**Hiding in Plain Sight: 21st Century Skills in Existing Assessments**

*Kristin Lansing-Stoeffler, ACT, INC.; Andries Van der Ark, RESEARCH INSTITUTE OF CHILD DE; Tessa van Schijndel, University of Amsterdam; Jeffrey Steedle, Curriculum Associates; Alina A. von Davier, Duolingo*

# FULL SCHEDULE

FRIDAY, APRIL 12

## **Harnessing Artificial Intelligence for Generating Items in Critical Thinking Tests**

*Alina A. von Davier, Duolingo; Hyo Jeong Shin, Sogang University; Helen Baron, Saville & Holdsworth LTD; Ji Hoon Ryoo, Yonsei University*

Discussant:

*Ahmad Rahimi, University of Florida*

## **048. Challenges and Innovations in Creating Interactive Reports of Student Progress and Growth Coordinated Paper Session**

**1:15 to 2:45 pm**

**Convention Center: Level 1 - 121A**

Over the past decade the transition of large-scale educational assessment from a largely paper and pencil environment to a digital, internet-based environment have opened the door to new possibilities in the way that stakeholders are able to interact with score reports in ways that better facilitate meaningful interpretations about student learning. The presentations in this session focus on research that attempts to innovate within this space, and some lessons that are being learned in the process. A common question they address is whether score reporting practices in the context of student growth are sufficiently attentive to principles related to the effective graphical visualization of quantitative data in educational measurement, and/or whether these principles are themselves a moving target. Conclusions and recommendations are informed by data from and interactions with both the conveyers of score reports (e.g., state and school district leadership) and the target audience (e.g., teachers and parents).

Session Organizer:

*Derek Christian Briggs, University of Colorado Boulder*

Chair:

*Laurie Davis, Curriculum Associates*

*Participants:*

### **Designing a Reporting Interface with User Experience in Mind**

*Karen Horowitz, Curriculum Associates; Tricia Maas, Curriculum Associates*

### **Content-Referenced Growth: Combining Item Mapping with a Learning Progression**

*Erik Whitfield, University of Colorado Boulder; Derek Christian Briggs, University of Colorado Boulder; Olivia Cox, University of Colorado Boulder; Sanford R Student, University of Delaware; Nathan Minchen, Curriculum Associates, LLC; Laurie Davis, Curriculum Associates*

### **Outcomes Reporting: Tracking Growth and Intervention Implementation**

*Amy Dray, Spencer Foundation; Sonya Powers, Edmentum, Inc.*

### **Assessment Reporting in Support of Student Learning**

*Damian Betebenner, National Center for the Improvement of Educational Assessment; Adam Vanlwaarden, Center for Assessment*

Discussants:

*Diego Zapata-Rivera, ETS*

*Katherine Furgol Castellano, ETS*

## **049. Navigating Challenges in Operational Psychometric Environments**

**Coordinated Paper Session**

**1:15 to 2:45 pm**

**Convention Center: Level 1 - 121B**

Operational psychometric work presents unique challenges due to the intricate demands for accuracy and fairness across diverse assessments. This coordinated session convenes experts, practitioners,

and researchers from various organizations to collectively address these pressing challenges. Our goal is to identify and share practical strategies and solutions that can be expediently deployed to mitigate these challenges. The first paper leverages natural language processing techniques to predict the response time in adaptive Next Generation Science Standards tests and proposes to integrate it into the item development process. Subsequently, the second and third papers tackle challenges of local independence violations in item response theory. Specifically, they address two distinct situations: discrepancies in scoring rubrics within writing tests, and uncontrolled speed at the item-level in timed adaptive exams. Both papers aim to achieve precise item and ability estimates, offering unique solutions for each scenario. The last two papers address the complexities of collaborating with subject matter experts, such as careful management between psychometricians and subject matter subjects, and test construction review, with the objective of optimizing the development of assessment by merging the strengths of both domains.

Session Organizer:  
*Yawei Shen, Pearson*

Chair:  
*Siyu Wan, ABIM*

Participants:

### **Predicting Response Times for NGSS-Aligned Item Development**

*Mina Lee, Cambium Assessment; Wanchen Chang, Cambium Assessment; Wei Schneider, Cambium Assessment; Zebing Wu, Cambium Assessment; Suhwa Han, University of Texas at Austin*

### **Adjusting Item Parameter Inflation Caused by Local Item Dependence in Writing Items**

*Hongyu Diao, ETS; Larissa Smith, ETS*

### **The Effects of Uncontrolled Speededness on IRT Item and Ability Estimation**

*Jaime Malatesta, GMAC; Paulius Satkus, GMAC; Kyung (Chris) T. Han, GMAC*

### **Collaborating with Subject Matter Experts**

*Janet Mee, NBME*

### **Best Practice for Review Test Form Constructed from Content Experts**

*Yang Lu, Pearson; Yawei Shen, Pearson; Mustafa Kuzey Bilir, Pearson*

Discussant:  
*Jiawei Xiong, Pearson*

## **050. Assessment on the AI: Challenges and Opportunities of Leveraging Large Language Model**

### **Coordinated Paper Session**

**1:15 to 2:45 pm**

**Convention Center: Level 1 - 121C**

In an era where Generative Artificial Intelligence (GAI) and Large Language Models (LLMs) like ChatGPT are becoming increasingly integrated into various sectors, this session aims to explore their transformative potential and challenges in the realms of educational and psychological assessment. Bringing together a diverse group of researchers and educators, the session will delve into five groundbreaking papers that offer innovative approaches, methodologies, and solutions for enhancing the reliability, quality, and adaptability of AI-generated outputs in sensitive and critical areas. The papers cover a wide array of topics, from improving the reliability of language models to automating the process of educational assessment and preventing cheating in online exams. Each paper not only presents novel findings but also discusses the implications for educators, researchers, policymakers, and AI developers. This session provides essential foundational elements and technologies required for leveraging GAI effectively. Whether you are an educator looking to implement AI in your curriculum, a researcher aiming to push the boundaries of what GAI can achieve, or a policymaker interested in the ethical and practical considerations of AI in education, this session offers the tools and insights you need to navigate the complex landscape of Generative Artificial Intelligence in educational and psychological assessments.

# FULL SCHEDULE

FRIDAY, APRIL 12

Session Organizer:  
*Jaehwa Choi, George Washington University*

Participants:  
**Realtime Intra-Self-Evaluation: Framework for Coping with Hallucinations of Large Language Models**  
*Dayeon Lee, University of Maryland - College Park*

**Assessment Reverse Engineering by AI: A Premiere**  
*Shonai Someshwar, UNC Greensboro; Eunji Lee, University of Georgia*

**Item Modeling for Preventing Assessment Cheating with Technology**  
*Eunji Lee, University of Georgia*

**Enhancing the Distractor Development Process by Leveraging AI Techniques**  
*Sunhyoung Lee, University of Nebraska-Lincoln*

**Potential Application of Generative AI in Adaptive Psychological Testing**  
*Yewon Kim, EWHA Womans University; Yejin Woo, Ewha Womans University; Youn-Jeng Choi, EWHA Womans University*

Discussant:  
*Jaehwa Choi, George Washington University*

## 051. Cognitive Reasoning and Complexity

**Paper Session**

**1:15 to 2:45 pm**

**Convention Center: Level 1 - 122A**

Chair:  
*Luciana Cancado, Curriculum Associates*

Participants:  
**Cognitively Complex Thinking in Large-Scale Assessment Items**  
*Jim Patterson, College Board; Dona Carling, College Board*

This session reports findings of a study of test questions from a large-scale assessment system using a think-aloud methodology. Analyzed quantitatively and qualitatively, students' verbal responses as they worked through the questions offer strong evidence that the questions elicited cognitively complex thinking.

**Developing Comparability of New Forms of a Cognitive Reasoning Test Internationally**  
*Thanos Patelis, Johns Hopkins U & University of Kansas*

The presentation involves the methods and results of developing new forms in Greek while ensuring comparability when the specifications were not clearly articulated. Workshops and alignments with bilingual experts were undertaken. Items were developed and field-tested. Extra analyses were undertaken. Activities enhanced expertise, produced detailed specifications and comparable scores.

**General Versus Situational Online Reasoning Among Graduates**

*Alice Laufer, Johannes Gutenberg University; Denis Federiakin, Johannes Gutenberg University Mainz; Olga Zlatkin-Troitschanskaia, Johannes Gutenberg University*

This study used Bayesian regressions to compare students' self-reported and observed online behavior, revealing no significant correlation. This indicates that general media online use is not necessarily associated with situational online behaviors and that it might depend on other individual factors or contextual factors such as time constraints or stress.

**Mental Rotation Performance: Contribution Of Item Features To Difficulties And Functional Adaptation**

*Mehdi Rajeb, University of Alabama; Andrew Krist, The University of Alabama; Qingzhou Shi, Northwestern University; Stefanie A. Wind, University of Alabama; Joni Lakin, University of Alabama*  
Linear logistic test models (LLTM) are used to explore attributes of test items that contribute to

difficulty. Using these attributes, researchers can explore sources of cognitive complexity. This paper goes further to explore attribute invariance for sub-groups, extending our understanding of LLTM for exploring test fairness.

Discussant:  
*Steve Ferrara, HumRRO*

### **053. Revolutionizing Alignment and Tagging Using AI: Results from the Front Lines** **Organized Discussion** **1:15 to 2:45 pm** **Convention Center: Level 1 - 123**

Everyone is talking about the potential for AI to revolutionize the test development pipeline, but many have questions about what it can really do and how to get started. In this information-packed session, attendees will hear from experts on the ground who have developed and implemented a state of the art AI system for tagging assessment items and learning content. Attendees are assured to receive actionable information as this session will contain practical advice about using AI technology from those who have been doing so for the past several years. An AI scientist and a Chief Assessment and Learning Officer will describe how it works, the amazing benefits to be had, and potential pitfalls to be aware of. A director of content development will describe how they implemented AI technology for tagging learning modules, and they will share data and results from this work. Part of the session will be devoted to a lively discussion with the audience. Attendees are assured to leave this session better informed about how AI tagging works, and they will better understand next steps toward using it with their own item banks.

Session Organizer:  
*Brad Bolender, Finetune*

Presenters:  
*Brad Bolender, Finetunze*  
*Christine Mills, Ascend Learning*  
*Nick Koprowicz, Sr. AI and Machine Learning Scientist at Prometric*

### **054. Tackling Unfairness in AI Scoring from Multiple Angles** **Coordinated Paper Session** **1:15 to 2:45 pm** **Convention Center: Level 1 - 124**

The use of artificial intelligence (AI) to score constructed responses is an area of educational measurement undergoing rapid development. There are new use contexts introduced each day and new AI technology such as generative AI is being incorporated into scoring systems. New use contexts and AI capabilities challenge the “standard practices” for how to best build and evaluate automated scoring models. One leading concern is the fairness of AI scores. While AI affords more agile applications of testing and learning solutions, its major limitation is measurement and algorithmic bias (Johnson et al., 2022). New use contexts allow for unexpected and unanticipated sources of bias to be introduced into scores. This coordinated session will have five papers summarizing a program of research that has investigated how to reduce unfairness in different ways, or from different angles. Two papers will propose methods for engine development and model building (Liu & Fauss; Flor), one paper will investigate methods for detecting subgroup bias (Casabianca), another paper will report on how to follow-up on traditional subgroup analyses by performing differential feature functioning analysis (Choi), and the last paper explores methods for explainable AI and how they can be used to improve transparency and fairness (Zhang).

Session Organizer:  
*Jodi M Casabianca, ETS*

Chair:  
*Daniel McCaffrey, ETS*

# FULL SCHEDULE

FRIDAY, APRIL 12

Participants:

## **A Bayesian Nonparametric Model for Flexible Automated Scoring**

*Xiang Liu, ETS; Michael Fauss, ETS*

## **Fairness Aspects Of Similarity-Based Automated Short-Answer Scoring**

*Michael Flor, ETS*

## **A Comparison of Fairness Evaluation Methods for AI Scores**

*Jodi M Casabianca, ETS; Daniel McCaffrey, ETS; Matthew S Johnson, ETS; Chen Li, ETS*

## **Examining Partial Derivatives to Identify Causes of Differential Prediction Bias in Automated Scores**

*Ikkyu Choi, ETS; Matthew S Johnson, ETS*

## **Explainable AI: Exploring Subgroup Differences in Short-Response Scoring**

*Mo Zhang, ETS; Chunyi Ruan, ETS; Matthew S Johnson, ETS*

## **055. Automated Scoring I**

### **Paper Session**

**1:15 to 2:45 pm**

**Convention Center: Level 1 - 125**

Chair:

*Alexander Kwako, Cambium Assessment*

Participants:

## **Automated Assessment of Finer-Grained Writing Conventions: A Case Study on Pronoun-Antecedent Agreement**

*Honeiah Karimi, University of California – Santa, Barbara; Amy Burkhardt, Cambium Assessment, Inc.; Susan Lottridge, Cambium Assessment, Inc*

We focus on the automatic detection of a fine-grained element that contributes to an essay's Conventions score: pronoun-antecedent agreement errors. We first implement a solution on a small set of test cases and then describe our lessons learned when applying it to a large set of student essays.

## **Automated Essay Scoring: Feature Selection with Scaling for Increased Performance and Interpretability**

*Christopher Jones, College Board*

Automated essay scoring has become dominated by black-box algorithms in recent years, failing to gain acceptance from skeptics. This study explores regression models around overshadowed readability indices with mean/sigma scaling. These models yield comparable results for reading and exceeded human agreement for writing essays while retaining interpretability.

## **The Effects of an Automated Writing Evaluation System on State Test Performance**

*Yue Huang, Measurement Incorporated; Joshua Wilson, University of Delaware*

This study explored the statewide implementation of an automated writing evaluation (AWE) system. A quasi-experimental design using the propensity score method was conducted to explore its impact on students' state test performance. Findings show that more AWE usage was causally associated with better outcome when compared to less/no usage.

## **Three Novel Aspects of the Statistical Evaluation of Automated Scoring Systems**

*Edmund Jones, Cambridge University Press & Assessment; Mark Gales, Enhanced Speech Technology; Shilin Gao, Cambridge University Press & Assessment*

Systems for automated scoring of constructed responses are evaluated using measures of the agreement between their scores and scores from expert human raters. We argue that additional statistical methods are needed, to address fairness for low- and high-ability candidates, compare the overall distributions, and show the accuracy of agreement measures.

Discussant:

*Mark D. Shermis, Performance Assessment Analytics, LLC*

**056. Toward Balanced and Equitable Assessment Systems: Considerations for Implementation and Use****Organized Discussion****1:15 to 2:45 pm****Convention Center: Level 4 - Terrace Ballroom III**

This session will engage participants in a discussion about the recently published National Academy of Education (NAEd) volume, *The Implementation and Use of Balanced Assessment Systems*. The volume's steering committee and chapter authors put forth a multi-faceted argument that balanced assessment systems must first and foremost support equitable and ambitious teaching. We discuss the system-level design, implementation, and policy levers needed to support these aims. In addition to defining these goals, this session will provide an analytic framework to examine the national-level policy logics, policy presses on states and local districts, and capacities of states and local districts to implement and use balanced assessment systems. This schema will better inform how states and districts understand the multi-dimensional idea of balanced assessment systems and therefore support its implementation and use. By centering on equitable and ambitious teaching and learning, this volume emphasizes the supportive and critical roles that states and districts need to play in enacting and using assessments.

Session Organizer:

*Scott Marion, National Center for the Improvement of Educational Assessment*

Presenters:

*James Pellegrino, University of Illinois at Chicago**Lorrie Shepard, University of Colorado Boulder**Guillermo Solano-Flores, Stanford University**Jennifer Randall, University of Michigan***057. Electronic Board Session I****Electronic Board Session****1:15 to 2:45 pm****Convention Center: Level 4 - Terrace Ballroom IV**

Participants:

**1. A Bayesian Logistic Regression Method to Calibrate Pretest Items in Multistage Testing***TsungHan Ho, ETS*

The Bayesian logistic regression (BLR: Ho, 2023) method was developed to address the response sparseness of linking items leading to concern about the quality of the link to the item bank in multistage testing. An optimized approach is proposed to improve the BLR performance evaluated using the simulation data.

**2. An Examination Of Factors That Relate To Item Parameter Drift: A Survival Analysis***Dongwei Wang, UMass Amherst; Carl Setzer, AICPA*

The purpose of this study is to identify item characteristics that relate to the likelihood of items being flagged for item parameter drift. Results indicate that certain item parameter estimates, revisions to content codes, and item skill levels impact the likelihood of being flagged.

**3. Assessing the Efficacy of Multilevel Structural Equation Modeling for Estimating Cross-Level Interactions***Hahyeong Kim, University of Illinois at Urbana-Champaign; Xinchang Zhou, UIUC*

This study aims to explore the effectiveness of multilevel structural equation modeling (MSEM) to assess cross-level interaction effects without creating proxy measures based on group means or group-centered variables which has been used in conventional multilevel modeling (MLM). To achieve the goal, this study includes Monte Carlo simulation.

**4. A Structural Change Analysis To Detect Item Preknowledge***Onur Demirkaya, Riverside Insights*

This research examines the application of structural change analysis with score-based tests for identifying item preknowledge in tests scored dichotomously. The study employed simulations and two linear versions of a computer-administered licensure exam to showcase the efficacy of the suggested test, especially when there's varying confidence regarding the compromised items.

- 5. Comparison of Human Rater and Automatic Scoring on Students' Ability Estimation**  
*Zhen Wang, Cambium Assessment; Xiaotong Yang, Florida State University*  
The purpose is to compare human rater with automatic scoring in terms of examinees' ability estimation with IRT-based rater model. Each speaking item is analyzed with both IRT models without rater-effect and with rater-effects. The effects of different rating design may substantially increase the bias in examinees' ability estimation.
- 6. Detecting Item Preknowledge On Re-Used Forms**  
*Aijun Wang, FSBPT; Yu Zhang, Federation of State Boards of Physical Therapy*  
This study proposes a simple and easy-to-implement approach to identify item compromise on re-used forms. It also examines the level of confidence in the validity of test scores based on the data collected from the reused forms.
- 7. Enhancing Test Security: Utilizing Parallel MST Pools for Effective MST Panel Assembly**  
*Hong Chen, The University of Iowa; YuLan Su, Center for Applied Linguistics*  
This study demonstrates the effectiveness of constructing parallel MST panels through unlimited parallel sub-pools while monitoring item exposure rate and overlap rate throughout the assembly process. Simulation studies reveal that multiple panels assembled by parallel sub-pools achieve comparable decision accuracy as a single panel assembled by the entire item pool.
- 8. Evaluating Approaches to Handling Dependent Effect Sizes in Meta-Analyses: A Simulation Study**  
*Ting Sun, University of Utah; Chuang Wang, University of North Carolina at Charlotte*  
This study evaluated the averaging method, robust variance estimation (RVE), and three-level meta-analysis (3LM) in handling dependent correlational effect sizes in meta-analyses. Data were simulated with varying levels of number of studies, effect sizes, and sample size. Results revealed that 3LM and RVE produced unbiased estimates of effect sizes.
- 9. Evaluating Subscore Interpretation for an Interactive Score Reporting Dashboard**  
*Francis O'Donnell, National Board of Medical Examiners; Rich Feinberg, National Board of Medical Examiners*  
Score reporting dashboards can present subscore information in novel ways, helping to meet growing demand for more detailed assessment results. We utilized a survey approach to uncover how student dashboard users intend to use subscore results, what features made the presentation of results confusing, and what factors contributed to misinterpretation.
- 10. Impact of Group Variability on Rasch Equating Under the Random Groups Design**  
*Chunyan Liu, National Board of Medical Examiners; Michael Kolen, The University of Iowa; Raja G Subhiyah, National Board of Medical Examiners*  
Group equivalence is assumed when performing random groups equating. However, in practice, proficiencies for examinees taking different forms may not be equivalent if random assignment or form spiraling doesn't work well. In this study, we will investigate the impact of various differences in group proficiencies on Rasch true-score equating.
- 11. Measuring Significantly Disabled English Language Learners' Language Proficiency Progression**  
*Kyoungwon Lee Bishop, WIDA at University of Wisconsin Madison; Hacer Karamese, WIDA at University of Wisconsin; Shangchao Min, Zhejiang University; Xin (Grace) Li, University of Wisconsin-Madison; Edynn Sato, Sato Education Consulting LLC*  
This research challenges historical notions of limited language development among ELs with significant cognitive disabilities. By merging IRT and CDM methodologies, we provide a comprehensive understanding of their language proficiency, offering vital insights for educators, policymakers, and practitioners.
- 12. DIF: Cross-country Functionality of a Large Scale Spanish Language College Entrance Exam**  
*Ruben Castaneda, College Board*  
The PAA is a Spanish language large-scale standardized test used for college admissions in Puerto Rico and Latin America. This study investigated the PAA for differential item functioning (DIF). We discuss methods of detecting DIF and identified country-level features that may lead to potential DIF.

**13. Exploring the Impact of Culturally-Relevant Assessment Material on Black Students' Reading Comprehension***Armani Michia Morris*

This study applies a diagnostic classification model (DCM) to investigate the impact of culturally-relevant assessment materials on Black students' reading scores. Results suggest that average-level readers with high prior knowledge of African-American culture have a higher chance of being masters on African-American-orientated texts than they do on other cultural texts.

**14. Informative IG Variance Priors for Bayesian Multilevel Item Response Model Precision***Liu Liu, University of Washington; Elizabeth A. Sanders, University of Washington, Seattle*

This study demonstrates how inverse gamma can be leveraged for informative variance priors in Bayesian multilevel item response models. Using a 20% subsample of 71 countries with PISA 2018 math data, credible intervals for posterior intercept estimates based on informative variance priors were more precise, while capturing full sample estimates.

**15. Modeling Between- and Within-Person Response Time-Response Dependency: A Comparison Between Two Approaches***He Ren, University of Washington; Chun Wang, University of Washington; Elizabeth A. Sanders, University of Washington, Seattle*

Response Time (RT) provides essential information about individuals' test-taking behaviors. Given that RT-response relationship differs at between- and within-person levels, this study aims to model and differentiate such dependency at these two levels simultaneously. Two models are evaluated and compared through an empirical study.

**16. Optimizing Test Length with Deep Reinforcement Learning***James Zoucha, University of Northern Colorado; Igor Himelfarb, NBCE; Nai-En Tang, National Board of Chiropractic Examiners; Bruce L. Shotts, NBCE*

Employing deep reinforcement learning, we investigated whether a reduced item subset adhering to structural constraints could match the precision of the complete set-in estimating student ability. Although combinations yielded similar precision, none fulfilled all additional constraints.

**17. Score Equity Assessment of Clinical Assessment Linking in Subgroups***Hsin-Ro Wei, Riverside Insights; Hyeon-Joo Oh, Riverside Insights; Youngmi Cho, Riverside Insights*

The purpose of this study is to examine the score equity assessment (SEA) in three subtests of the clinical assessment battery (e.g., Woodcock-Johnson test) in gender, race/ethnicity, and age groups. The preliminary results for the three subtests suggested that score equity was achieved with respect to gender groups.

**18. Using Process Data in Diagnostic Classification Modeling with More Flexible Distributional Assumptions***Xin Qiao, Shandong Academy of Sciences*

Jointly modeling of process data and item responses helps improve classification accuracy and provides more comprehensive diagnosis in cognitive diagnostic modeling with the premise of correct model specifications. This study proposes the use of the flexible Conway-Maxwell-Poisson (CMP) distribution for action counts in this context for more accurate statistical inference.

**19. A Recommendation System To Support Teachers' Continuous Formative Assessment In Blended Classrooms***Jinnie Choi, Savvas Learning Company*

Ideally, K-12 teachers in blended classrooms use digital assessment results to adjust instruction in a continuous and formative manner. In reality, challenges remain in quickly and accurately identifying optimal instructional resources aligned to students' assessment results. We describe an approach to instructional recommendation that aims to address these challenges.

**20. Examining Differences in Student Group Performance Using the QuantCrit Framework***Quintin Love, WestEd; Lauren White, Pearson; Daniel Murphy, WestEd; Sumbo Oluwalana, WestEd*

Using the QuantCrit framework, this paper critically examines the statistical analyses and associated outcomes and inferences to assess the degree to which the Massachusetts innovative

pilot assessment project is meeting the desired goals of minimizing measurement inequities. Results are interpreted using asset-based narratives and contextualized with student survey responses.

### **21. Process Mining for Mapping Response Sequences: NAEP Grade 4 Math Item**

*Radhika Kapoor, Stanford University; Juanita Hicks, AIR*

Process data collected in NAEP 2017, such as clicks and interaction with items, is analyzed for one Grade 4 Math item. This study uses a Fuzzy Miner algorithm to extract and distinguish action sequences of students who got the item correct, partially correct, or incorrect.

### **22. Evaluating Dimensionality in Non-cognitive Measures: Exploratory vs. Hierarchical Exploratory Graph Analysis**

*Tarid Wongvorachan, University of Alberta; Okan Bulut, University of Alberta; Guher Gorgun, University of Alberta*

Network psychometrics is a new approach to exploring dimensionality of a measurement tool using partial correlations among items. In this study, we compared the performance of exploratory graph analysis (EGA) and hierarchical EGA in detecting the dimensional structure of a non-cognitive measure.

### **23. A Micro-Longitudinal Approach to the Effects of Immediate Feedback on Test-Taker Emotions**

*Livia Kuklick, IPN Kiel, Germany; Ute Mertens, Leibniz Institute for Science and Mathematics Education; Sara Finney, James Madison University; Marlit Annalena Lindner, IWM - Leibniz Institut für Wissensmedien, University of Tübingen*

This study investigated the effects of implementing immediate, automated feedback in a computer-based low-stakes assessment. While students' positive emotions and performance decreased over the course of the assessment without feedback, they increased with feedback. Data imply that automated feedback may buffer undesirable item position effects in low-stakes contexts.

### **24. Proficiency is Still Not Enough**

*Pohai Kukea Shultz, University of Hawaii at Manoa; Kerry Englert, Seneca Consulting, LLC*

Typical deficit approaches to communicating assessment results are not useful or equitable for marginalized students. To identify better ways of data reporting, we must work toward a system that gives voice to our communities. Our presentation will focus on how Hawaii's KĀ'EO assessment is advancing data reporting through community engagement.

## **058. Comprehensive Assessment and Score Development: The NIH Infant and Toddler Toolbox Coordinated Paper Session**

**3:05 to 4:35 pm**

**Convention Center: Level 1 - 120B**

The NIH Infant and Toddler ("Baby") Toolbox (NBT) is a groundbreaking measurement tool designed for children aged 1 through 42 months. The NBT leverages the capabilities of an iPad, incorporating built-in eye-tracking technology and brief video clips, to facilitate precise, objective assessments of children's developmental functioning. The administration of the NBT, along with data collection and scoring, is a central focus of this symposium. One key aspect of the NBT is its use of passive-viewing gaze measures, which are particularly suitable for children under 24 months of age. An administrator sets up a task on the iPad, which monitors where the child is looking as they view test stimuli on the screen. Crucially, the iPad's front-facing camera captures the child's gaze, allowing for a non-verbal expression of their comprehension and responses. Additionally, the development of normative data for these innovative assessments will be discussed, shedding light on how this tool can be used effectively to assess and understand the cognitive development of infants and toddlers. In sum, the NBT represents a cutting-edge approach to early childhood assessment, offering new insights into the developmental processes of our youngest learners.

Session Organizer:

*Lihua Yao, Northwestern University*

Chair:

*Daniel Lewis, Creative Measurement Solutions LLC*

# FULL SCHEDULE

FRIDAY, APRIL 12

Participants:

## **An Overview Of The Nih Baby Toolbox: Introduction, Validation, Scoring, And Norming**

*Lihua Yao, Northwestern University; Vitali Ustinovich, Northwestern University; Aaron J. Kaat, Northwestern University; Miriam A. Novack, Northwestern University; Rachel Hanrahan, Northwestern University; Paula H. Sievert, Helium Foot Software; Cindy J. Nowinski, Northwestern University; Courtney Blackwell, Northwestern University; Rachel Flynn, San Francisco State University; Maria Varela Diaz, Northwestern University; Sarah Pila-Leiderman, Northwestern University; Richard C. Gershon, Northwestern University*

## **Models for Social Functioning/Behavior Assessment**

*Maxwell Mansolf, Northwestern University; Lihua Yao, Northwestern University; Hugh Adam, Northwestern University; Aaron Kaat, Northwestern University*

## **Validation of the Baby Toolbox**

*Miriam A. Novack, Northwestern University; Y. Catherine Han, Northwestern University; Elizabeth M. Dworak, Northwestern University; Aaron J. Kaat, Northwestern University*

## **Norm Development of the Baby Toolbox**

*Y. Catherine Han, Northwestern University; Aaron Kaat, Northwestern University; Lihua Yao, Northwestern University; Miriam A. Novack, Northwestern University; Elizabeth M. Dworak, Northwestern University; Richard C. Gershon, Northwestern University; Kay Savio, Sago*

Discussant:

*Richard Patz, University of California, Berkeley*

## **059. Randomly Parallel Tests: Theory, Use, and Benefits**

### **Coordinated Paper Session**

**3:05 to 4:35 pm**

**Convention Center: Level 1 - 120C**

This session will explore theoretical and practical perspectives of this old/new innovation of creating and using randomly parallel tests (including the large pools of items) or the resulting randomly parallel test forms. Presentations will cover the history and theory of randomly parallel tests, including the closely related topic of stratified randomly parallel tests. The definitions of sources of measurement error, both random and systematic, and for true scores are revisited. Also presented are recommended procedures that quantify various psychometric properties of such testing, such as reliability. Procedures for efficiently building such tests today will be showcased. Finally, simulations and empirical research will provide evidence for the use of randomly parallel tests in operational testing programs.

Session Organizer:

*David Foster, Caveon Test Security*

Participants:

## **Randomly Parallel Forms: Some History, Theory, and Challenges**

*Robert Brennan, University of Iowa*

## **Psychometric Framework and Properties of Randomly Parallel Forms**

*Won-Chan Lee, University of Iowa; Stella Kim, University of North Carolina at Charlotte*

## **What is the Generalizability of Randomly Parallel Forms?**

*Craig Wells, UMass Amherst; David Foster, Caveon Test Security; Jungwon Kyung, UMass Amherst*

## **Evaluating the Test Security Benefits of Randomly Parallel Tests: Protecting Against Pre-Knowledge**

*Sergio Araneda, Caveon; David Foster, Caveon Test Security; Andrew Marder, Caveon*

Discussant:

*Wayne J. Camara, LSAC*

### **060. Furthering research underpinning OECD Programme for International Student Assessment - PISA**

**Coordinated Paper Session**

**3:05 to 4:35 pm**

**Convention Center: Level 1 - 121A**

This session presents a collection of papers that each aim to contribute to the research base that feeds into the ongoing improvement of the OECD PISA study. As a large-scale assessment that places a significant priority on maintaining cross-sectional trends, it is important that research balances the needs to maintain the underpinning constructs, with innovation that improves the efficiency of the assessment and integrates new approaches from the literature. This is particularly important in the context of PISA, a study that is growing in size: both in the number of participants, but also in the growing size of item pools and questionnaire forms. The focus in this session is embedded in the context of the now concluded 2022 cycle, and forthcoming 2025 cycle. In 2025, more than 90 countries will participate in PISA, all core assessment domains (Reading, Mathematics, and Science) will transition to be adaptive tests, and the questionnaire item pool will continue to grow and develop. In turn, the papers presented here interrogate how consistent international measurement and comparisons can be maintained, how we can efficiently assemble adaptive tests, how we can ensure contextual questionnaire material is relevant, and how to better interpret group-level differences in contextual constructs.

Session Organizer:

*Alejandra Osses-Vargas, Australian Council for Educational Research*

Chair:

*Goran Lazendic, ACER*

Participants:

#### **Considering the Balance Between Measurement Non-Invariance and Construct Validity**

*Daniel Cloney, Australian Council for Educational Research; Alla Berezner, Australian Council for Educational Research; Raymond J Adams, Australian Council for Educational Research; Bakir Haryanto, Australian Council for Educational Research*

#### **Implication of Modelling Choices for Indicators of Economic, Cultural, and Social Status**

*Alejandra Osses-Vargas, Australian Council for Educational Research; Raymond J Adams, Australian Council for Educational Research; Ursula Schwantner, Australian Council for Educational Research*

#### **Unpacking Automated Test Assembly: New Findings and Directions for the Future**

*Ling Tan, Australian Council for Educational Research; Charlotte Waters, Australian Council for Educational Research; Fuchun Huang, Australian Council for Educational Research; Daniel Cloney, Australian Council for Educational Research*

#### **Developing Empirical Growth Benchmarks to Interpret Group-Level Differences in PISA**

*Francesco Avvisati, OECD*

### **061. Linguistically Equitable Reading Screening: Bias Analyses and Universal Item Design**

**Coordinated Paper Session**

**3:05 to 4:35 pm**

**Convention Center: Level 1 - 121B**

The majority of US states mandate universal reading screening, which calls for thoughtful solutions to ensure screening remains equitable for the ever-growing bilingual student population. The Multitudes project is developing a free K-1 reading screener, focused on early identification of reading problems in California's linguistically diverse population. We therefore incorporate multilingual learners throughout the entire assessment development process. In this session, we describe the development and

evaluation of a pilot version of the screener using a suite of existing English reading measures and evaluate it with a focus on linguistic fairness (paper 1). Subsequently, we present the development and evaluation of a suite of English and Spanish tasks, culturally and linguistically tailored to California's student population. Papers 2 and 3 discuss the development of a sentence repetition and word reading task, respectively, and investigate the roles of language proficiency and linguistic features of the stimuli on item difficulty. Paper 4 provides a novel way of screening Spanish/English bilingual students regardless of their language profile by using conceptual scoring for an expressive vocabulary task. Overall, in this session, we sketch out the process of developing a linguistically equitable reading screener and share valuable insights into appropriate design and analysis considerations.

Session Organizer:

*Julian Maximilian Siebert, Stanford University*

Participants:

### **Universal Reading Screening for California: Linguistically Sensitive Task Selection and Prediction Evaluation**

*Julian Maximilian Siebert, Stanford University; Hugh Catts, Florida State University; Yaacov Petscher, Florida Center for Reading Research; Francesca Pei, University of California, San Francisco; Marilu Gorno Tempini, University of California, San Francisco*

### **Item Difficulty in English/Spanish Sentence Repetition Tasks and Correlations with English Proficiency**

*Amy Pratt, University of Cincinnati; Javier Jasso, The Ohio State University; Lillian Duran, University of Oregon; Mónica Zegers, University of California, San Francisco; Cengiz Zopluoglu, University of Oregon*

### **Spanish Word Reading: Considerations for Item Selection, Stimulus Analysis, and Item Difficulty**

*Mónica Zegers, University of California, San Francisco; Julian Maximilian Siebert, Stanford University; Cengiz Zopluoglu, University of Oregon*

### **Conceptual Scoring in Universal Screening: An Example Using Expressive Vocabulary**

*Lillian Duran, University of Oregon; Cengiz Zopluoglu, University of Oregon; Mónica Zegers, University of California, San Francisco; Julian Maximilian Siebert, Stanford University*

Discussant:

*Ben Domingue, Stanford University*

## **062. Using Assessment Engineering and Item Complexity Modeling to Enhance Score Interpretation** **Coordinated Paper Session** **3:05 to 4:35 pm** **Convention Center: Level 1 - 121C**

Subject matter experts (SMEs) used assessment engineering techniques to define complexity design layers (CDLs) and develop complexity scoring protocols (CSPs) for MAP Growth math and reading assessments. Psychometricians used the ratings and derived complexity covariates (DCCs) to predict item difficulty. Purposes of the work include the development of resources to facilitate score interpretation and automatic item generation. Math and reading SMEs followed a similar process to define the CDLs and CSPs. They drew from existing literature and content standards to define the initial versions. SMEs then went through two rounds of rating a sample of 50 items and revising the protocols. After a final round of revisions, SMEs rated a sample of 100 items in each subject. Psychometricians analyzed the ratings. R-squared was 0.79 for the math analysis and it was 0.69 for the reading analysis. The analysis explored a variety of machine learning methods for producing DCCs and predicting item difficulty. Two papers in this coordinated session describe work by the SMEs to define and refine the CDLs and CSPs in math and reading. The other two papers describe item difficulty modeling results using regression and a variety of machine learning methods.

Session Organizer:

*Patrick Meyer, NWEA*

Chair:

*Richard Melvin Luecht, University of North Carolina at Greensboro*

Participants:

### **Developing Complexity Design Layers and Scoring Protocols for Math Items**

*Sarah Whitney, NWEA; Alane Ferland, NWEA; Mark Lawler, NWEA; Desiree Spikings, NWEA; Mary Resanovich, NWEA*

### **Math Item Difficulty Modeling Using Complexity Design Layers and Item Metadata**

*Xueming Li, NWEA*

### **Defining Text Complexity, Item Complexity, and the Text-by-Item Interaction for Reading Assessment**

*Gina Wilmurth, NWEA; Julie Richardson, NWEA; Adam Withycombe, NWEA; Janice Lee Johnson, NWEA*

### **Reading Item Difficulty Modeling with Various Machine Learning Methods**

*Ann Hu, NWEA; Patrick Meyer, NWEA*

Discussant:

*Steve Ferrara, HumRRO*

## **063. Bayesian Techniques**

### **Paper Session**

**3:05 to 4:35 pm**

**Convention Center: Level 1 - 122A**

Chair:

*Klint Kanopka, New York University*

Participants:

### **Integrating Bayesian Explanatory IRT Models and CAT Algorithms for Small Sample Settings**

*Ae Kyong Jung, University of Iowa; Jonathan Templin, University of Iowa; Anne Corinne Huggins-Manley, University of Florida; Tae Yeon Kwon, University of Florida; Jiehan Li*

This study investigates the integration of Bayesian explanatory item response models and computerized adaptive testing algorithms to address limited sample size challenges in classroom assessments. The study introduces novel algorithms for item selection and ability estimation with a fully Bayesian approach. Explanatory IRT models are utilized to efficiently estimate item properties.

### **Item Wording Effects within Variations of ESEM and BSEM Factor Models**

*Hyeri Hong, California State University, Fresno; Earvin Balderama, California State University, Fresno*  
Method effects can cause misinterpretation of results in self-report measures if not properly controlled. We investigated controls for negative and positive item wording effects within Exploratory and Bayesian structural equation modeling procedures. The results suggest that ESEM WLSMV and informative Bayesian factor analytic techniques reduce the item wording effects.

### **Performance of Non-parametric Approaches to Model Selection in Mixture Psychometric Models**

*Clifford Erhardt Hauenstein, Johns Hopkins School of Medicine; Eunbee Kim, Georgia Institute of Technology*

A pseudo non-parametric method for model selection in mixture modeling applications has recently been proposed as an improved alternative to fully non-parametric methods. We evaluate performance of this recent method for mixture and hidden Markov IRT models. Additionally, we discuss its utility over traditional model selection approaches.

### **Sensitivity Of Fit Measures To Misspecification In Ordinal Bayesian Confirmatory Factor Analysis**

*Ejike Edeh, University of Arkansas; Xinya Liang, Department of Rehabilitation, Human Resources and Communication Disorders, University of Arkansas*

This study investigates the sensitivity of various fit measures in ordinal BCFA, including posterior predictive p-value, BIC, DIC, WIC, and LOO, to the misspecification of dimensionality and cross-loadings. The findings shed light on how each fit measure behaves under different prior specifications and ordinal distributions.

Discussant:

*Wes Bonifay, University of Missouri*

### **064. What Are Important Quality Criteria For Assessments In A Competency-Based Educational Program?**

**Organized Discussion**

**3:05 to 4:35 pm**

**Convention Center: Level 4 - Terrace 3**

There continues to be an increase and interest in competency-based education programs in both K-12 and higher education. The definitions and descriptions of competency-based education programs have assessments as the foundation. But there are various types of assessments that are used. As we know, we must gather evidence to support the quality of these assessments. Unfortunately, many programs start quickly, have limited staff and funding, and are forced to stage the various aspects of evidence over time. Because there are various types of assessments and various types of evidence needed, decisions need to be made as to what evidence should be prioritized first, second, third, etc. In this session, three presenters with experience and expertise in assessment and competency-based education will offer their suggestions for what is the most important criterion for assessments. After each of the presenters have made their suggestions, the chair of this session will ask the presenters whether they would change their recommendation based on what the other presenters indicated. Additionally, the chair will turn to the audience for comments and questions. This session should offer those involved in implementing competency-based education programs and/or responsible for the assessments some actionable ideas.

Session Organizer:

*Thanos Patelis, Johns Hopkins University & University of Kansas*

Moderator:

*Thanos Patelis, Johns Hopkins University & University of Kansas*

Presenters:

*Neal Kingston, University of Kansas*

*Katie McClarty, McGraw Hill Education*

*Scott Marion, National Center for the Improvement of Educational Assessment*

### **065. Subscores: A Practical Guide for Their Production and Consumption**

**Coordinated Paper Session**

**3:05 to 4:35 pm**

**Convention Center: Level 1 - 123**

This coordinated session highlights the contents of the just-published book entitled, “Subscores: A Practical Guide for Their Production and Consumption”. The book synthesizes research on the most salient topics and questions related to subscores so that a broad range of stakeholders can have a coherent and complete source of practical advice when considering test design and score interpretation decisions. In this session, we will provide an overview of the main themes of the book including: (1) historical context and application with the U.S. Census and Armed Services Vocational Aptitude Battery (ASVAB) as illustrative examples, (2) what subscores look like in practice and considerations for score report design, (3) how to tell which subscores add value and hence are worth reporting, and (4) how often subscores add value in practice and what credible options are available when they don’t. We will also include recommendations for communicating subscore challenges and opportunities to facilitate a productive dialogue in the likely scenario that subscores lack sufficient psychometric properties to

# FULL SCHEDULE

FRIDAY, APRIL 12

support their desired purposes. Beyond the book content, this session also includes survey feedback from parents, educators, and leaders that provide insight into how subscore information is used and factors contributing to increased demand.

Session Organizers:

*Rich Feinberg, National Board of Medical Examiners*  
*Shelby Joel Haberman*

Participants:

**Historical Context For Subscore Use With The U.s. Census As An Illustrative Example**

*Howard Wainer, National Board of Medical Examiners*

**How Do Various Stakeholders Use Subscore Information?**

*Chris Domaleski, Center for Assessment*

**How Subscores Are Reported, Design Considerations, And How To Quantify Subscore Value**

*Rich Feinberg, National Board of Medical Examiners*

**How Often Subscores Add Value And What Can Be Done When They Don't**

*Sandip Sinharay, ETS*

Discussant:

*David Thissen, University of North Carolina*

## **066. Unpacking Test-Taker (Dis)Engagement: Innovative Approaches to Enhance Validity in Educational Assessment**

**Coordinated Paper Session**

**3:05 to 4:35 pm**

**Convention Center: Level 1 - 122B**

The validity of inferences drawn from educational assessments is contingent not only on the quality of the test items but also on the level of engagement exhibited by the test takers. Disengagement can distort scores, potentially underestimating a student's abilities. This session delves into the multifaceted nature of test (dis)engagement, the methodologies to detect it, and address its effects. Presentation 1 examines the impact of disengagement-adjusted scoring, recommending score-adjustment procedures consider both rapid guessing and performance change information. Presentation 2 uses Monte Carlo simulations to quantify how disengagement affects the recovery of latent growth curve model parameter estimates and exemplifies how to mitigate related issues. Presentation 3 uses Monte Carlo simulations to examine the parameter recovery of the multidimensional Nested Logit Model, which measures ability and test-taking effort jointly to gain a deeper comprehension of test-takers' response behaviors. Presentation 4 employs AI algorithms to analyze students' navigation patterns in the NAEP assessments, offering insights into test engagement. Lastly, Presentation 5 introduces and validates the Cognitive Modeling Method, a theory-driven approach to detect not-fully-effortful responses to further improve inferences drawn from assessments. Together, these presentations highlight the importance of (dis)engagement in assessments and provide novel approaches to ensure valid score interpretation.

Session Organizer:

*Burcu Arslan, ETS Global B.V.*

Chair:

*Burcu Arslan, ETS Global B.V.*

Participants:

**The Impact of Disengagement-Adjusted Scoring on Test Score Validity**

*Steven Wise, Measurement Consultant; Megan Kuhfeld, NWEA; Jazmin Isaacs, NWEA*

**Quantifying the Effect of Test Disengagement on Student Growth Estimates**

*James Soland, University of Virginia*

### **Joint Modeling of Ability and Test-Taking Effort with Multidimensional Nested Logit Model**

*Okan Bulut, University of Alberta; Kylie Gorney, Michigan State University; Seyma N. Yildirim-Erbasli, Concordia University of Edmonton*

### **Investigate Students' Engagement with the NAEP Assessment Through Item Navigation**

*Hongwen Guo, ETS; Matthew S Johnson, ETS; Luis Saldivia, ETS; Michelle Worthington, ETS*

### **Introducing and Evaluating a Theory-Driven Method to Detect Not-Fully-Effortful Responses**

*Burcu Arslan, ETS Global B.V.; Blair Lehman, ETS; Marlit Annalena Lindner, IWM - Leibniz Institut für Wissensmedien, University of Tübingen*

Discussant:

*Frank Goldhammer, DIPF | Leibniz Institute for Research and Information in Education, ZIB*

## **067. Cognitive Diagnostic Models I**

### **Paper Session**

**3:05 to 4:35 pm**

**Convention Center: Level 1 - 125**

Chair:

*Justin L. Kern, University of Illinois at Urbana-Champaign*

Participants:

### **A Polytomous Extension of the Higher-order, Hidden Markov Model with Hierarchical Trajectories**

*Danhui Zhang, Beijing Normal University; Shixiu Ren, Beijing Normal University; Xin Xu, Beijing Normal University; Xintong Shan, Beijing Normal University*

In this paper, a new method called PHO-HMM was suggested for analyzing polytomous response data to accurately track students' learning progress over time. The results of both simulation and empirical studies demonstrated that the PHO-HMM model can be effectively used for longitudinal polytomous response data.

### **Longitudinal CDMs With Small Samples: In Contrast to Nonparametric CD Methods**

*Hyunjee Oh; Chia-Yi Chiu, University of Minnesota*

The study aims to examine the effectiveness of the existing longitudinal CDMs in analyzing small samples, in contrast to the nonparametric classification methods. The findings of this study will serve as a practical guideline for analyzing longitudinal data in CD based on the sample size.

### **Nonparametric Cognitive Diagnosis When Attributes are Polytomous**

*Youn Seon Lim, University of Cincinnati*

Within cognitive diagnosis models, attributes are typically perceived as binary. However, polytomous attributes may yield higher precision in the assessment of examinees' attribute mastery. This study proposed a nonparametric method for cognitive diagnosis for use with polytomous attributes, called the "Nonparametric Polytomous Attributes Diagnostic Classification" (NPADC) method.

### **Q-Matrix Accuracy Indices for Cognitive Diagnosis Models**

*Rodrigo Schames Kreitchmann, Universidad Nacional de Educacion a Distancia; Jimmy de la Torre, University of Hong Kong*

This study introduces a family of indices to quantify the uncertainties associated with the provisional Q-matrix specifications. Based on the posterior probabilities of the q-vectors for each item, the kappa indices indicate the proportion of correctly specified q-entries. Simulation results demonstrate their effectiveness in determining the quality of Q-matrix specifications.

Discussant:

*Laine Bradshaw, Pearson*

**068. Fairness in the Next Generation of the Standards: A Discussion on Equity**  
**Organized Discussion (Joint AERA/NCME Session)**  
**3:05 to 4:35 pm**  
**Convention Center: Level 2 - 203AB**

This panel is jointly sponsored by NCME and AERA, and will discuss issues related to how the next version of the Standards for Educational and Psychological Testing should treat the role of fairness from the perspective of equity for traditionally marginalized racial-ethnic communities. In early 2023, NCME and AERA both surveyed membership on the issues that should be addressed in the next generation of the Standards, which are currently undergoing review for revision. In both surveys, the theme of fairness was prominent. There is much discussion in the field of educational measurement about how assessment purpose and use, design, psychometrics, analyses, and interpretation should evolve to better ensure equitable treatment of students from all populations, including and especially students from historically marginalized communities. The members of the panel have been carefully chosen to represent voices from a variety of perspectives in the field: academics, practitioners, and graduate students. NOTE: Charlotte Gilbar is a presenter, not a discussant, but it would only allow me to add 5 presenters.

Session Organizer:  
*Michael C. Rodriguez, University of Minnesota*

Presenters:  
*Michael E. Walker*  
*Pohai Kukea Shultz, University of Hawaii at Manoa*  
*Molly Faulkner-Bond, WestEd*  
*Montserrat B Valdivia Medinaceli, UC DAVIS*  
*Tyrone Howard, UCLA*  
*Kyndra Middleton, Howard University*

Discussant:  
*Charlotte Gilbar, Natrona County School District*

**069. Electronic Board Session II**  
**Electronic Board Session**  
**3:05 to 4:35 pm**  
**Convention Center: Level 4 - Terrace Ballroom IV**

Participants:

**1. An Evaluation Of Factors Influencing Inclusive Calibrations**

*Mingjia Ma, Cambium Assessment, Inc.; Elizabeth Ayers-Wright, Cambium Assessment, Inc.;*  
*Christina Schneider, Cambium Assessment, Inc.; Jocelynn Pittman, Cambium Assessment*

Items translated into a different language often share parameters with the original language version of the item. Inclusive calibrations occur when examinees taking the translated assessment are included with the original language population. This paper investigates how examinee sub-populations influence the scoring tables under inclusive calibration.

**2. Comparing CUSUMs using Z-scores and Model-Based Residuals for Detecting Aberrant Responses**

*Xuan Wang, National Conference of Bar Examiners; Mengyao Zhang, National Conference of Bar Examiners*

This study aims to compare the performance of two types of CUSUM procedures, traditional CUSUM using Z-scores and CUSUM using model-based residuals, and to evaluate how UB/LB thresholds, detection rates and detected outliers differ when applying these procedures under various conditions of test length and sample size.

**3. Enhancing Performance Prediction with Transformer-Based Deep Knowledge Tracing**

*Fang Peng, Houghton Mifflin Harcourt*

Deep Knowledge Tracing (DKT) is a research frontier that leverages deep learning techniques to

model and predict students' learning progress. This study explores a transformer-based DKT model to forecast student responses, unveil patterns in knowledge acquisition, and infer relationships among measured knowledge components from large-scale assessment data.

#### **4. Exploring Maximum Likelihood Estimation with Fence Method in Final Scoring in CAT**

*Chunxin Wang, ACT INC.; Jie Li; Yi He, Edmentum; Jin Zhang, ACT; Chi-Yu Huang, ACT, INC*

This study explores the scoring method of maximum likelihood estimation with fence (MLEF) in final scoring in CAT. The results will provide information on how the MLEF method can be implemented in practice to yield optimal results given the test length, selection method, and ability levels of examinees.

#### **5. Investigation of Measurement Invariance for Computational Thinking Assessment**

*Richard Brown, West Coast Analytics; Emily Anne Brown, University of North Texas*

Previous research has been limited regarding the measurement of computational thinking, particularly as a learning progression in K-12. This study investigates the measurement invariance of a newly developed measure of computational thinking (Toyama, et al., 2020; Wilson, et al., 2021).

#### **6. MIRT Linking for the Bifactor with Mixed Format Test**

*Sohee Kim, University of South Alabama; Yulim Kang, Yonsei University*

This study compared MIRT linking methods (CC and FIPC) for the bifactor model with mixed format tests. Different ratios of common item formats and group ability distributions were considered. The both methods well recovered the item parameters and they produced similar patterns based on the outcomes from simulation.

#### **7. Psychometric Properties of Illinois Bullying Scale For Youths in Restrictive School Settings**

*Wenchao Ma, University of Alabama; June Preast, University of Alabama; Kristine Jolivet, University of Alabama; Kimberly Odom, University of Alabama; Stephanie Shelton, The University of Alabama; Olivia Hester, University of Alabama; Sara Sanders, University of Alabama; Nicole Prewitt, University of Alabama*

The present study investigated the psychometric properties of the Illinois bullying/victimization scale within restrictive school settings. Exploratory structural equation modeling (ESEM) was used and was found to provide better model-data fit than confirmatory factor analysis. Some modifications to the scale were suggested based on ESEM analysis.

#### **8. Seed Mining: Using Intentional Seed Setting to P-hack Bootstrap Estimates**

*Wilson Hatcher, UNLV CREA*

We introduce "seed-mining" as a method of p-hacking bootstrap estimations. Seed mining is the systematic search for a seed that produces a confidence interval or p-value that helps publication likelihood. Including the seed in pre-registration of studies that use bootstrapping and similar methods can reduce the threat.

#### **9. Investigating The Impact Of Test Wiseness In Multiple Response Items: A Simulation Study**

*Jiayi Deng, University of Minnesota, Twin Cities; Kirk Becker, Pearson*

This simulation investigated ability estimation accuracy for multiple response (MR) items when manipulating: (a) item stem presentation; (b) item structure design; (c) number of keys; (d) scoring method; (e) test wiseness; and (f) test length. Results suggest test wiseness and item scoring decisions can bias ability estimation in MR items.

#### **10. Assessing Differential Item Functioning of the Children's Hope Scale (CHS)**

*Latif Kadir, The Ohio State University; Desmond Atitsogbey, The Ohio State University; Jerry Bean, The Ohio State University*

Differential Item Functioning is a common measurement issue that plagues measurement scales, rendering results from such scales biased as it yields an unfair advantage to a group of test takers. This study conducts DIF and DTF analyses on the CHS across gender and race. Results indicate consistency across both groups.

**11. Comparing Interrater Reliability Assessment Methods Using Simulation**

*Tianshu Pan, Pearson*

Intraclass correlations (ICC) are often applied to assess interrater reliability. Pan and Yin (2023) suggested using Theil-index ratios. This study compares their performances using simulation. The results show the Theil-index ratios tend to be close to true ICC values and their estimates, and ICC tend to underestimate the interrater reliability.

**12. Evaluating the Effect of Bootstrap-Resampling in Enhancing the Norming Accuracy**

*Tong Wu, Riverside Insights; Hyeon-Joo Oh, Riverside Insights; David Dailey, Dailey Data Group; Erica Laforte, Northwestern University*

Norm-referenced tests rely on precise normed scores to measure an examinee's relative standing, potentially impacting major life decisions. This study investigates bootstrap-resampling technique to enhance the accuracy of reference ability scores, a type of normed scores, when paired with different norming methods under a set of simulation conditions.

**13. Exploring Dynamic Relationships Within and Between Noncognitive Measures: A Psychometric Network Approach**

*Hatice Cigdem Bulut, Northern Alberta Institute of Technology; Betul Gokcen Dogan Lacin, Mustafa Kemal University*

This study explores the complex relationships among the items from two noncognitive measures using psychometric network models. The results reveal new insights into the relationships between noncognitive measures, distinct from traditional measurement models. Also, these models allow for investigating the invariance of the noncognitive measures based on external variables.

**14. How Many Is Enough to Calibrate Mixed Format of 3PL and GPCM?**

*Hye-Jeong Choi, Human Resources Research Organization; Dipendra Subedi, Pearson; Yufeng Berry, Minnesota Department of Education; Meng Fan, Human Resources Research Organization; Gerald Griph, Pearson; Changjiang Wang, Pearson; Yvette M Nemeth, HumRRO*

This study intends to investigate the effects of sample size on mixed IRT model parameter calibration (3PL and generalized partial credit model). The performance of three software packages (PROC IRT, IRTPRO, and mirt) will be compared. To do so, a simulation study will be conducted to evaluate recovery of parameters.

**15. Impact of Item Statistics on Item Evaluations by Content Experts**

*Jinah Choi, Edmentum, Inc.; Sonya Powers, Edmentum, Inc.; Sandra Olivarez-Durden, Edmentum, Inc.; Sally Rhodes, Edmentum, Inc.; Jennifer Rubel, Edmentum, Inc.; Lauren Hamilton, Edmentum, Inc.; Rebekah Boothe-Corley, Edmentum, Inc.; Roshi Matewere, Edmentum, Inc.*

This study examines the influence of assessment item statistics on the item review process. Statistics are provided for half of a set of items, randomly selected. Content experts use the same item evaluation criteria to evaluate both sets of items. Differences in evaluation outcomes across item sets are compared.

**16. Investigating Scientific Inquiry Processes in an Interactive Task with Sequential Analysis**

*Shuang Wang, Beijing Normal University; An Hu, State Key Laboratory for Artificial Microstructure and Mesoscopic Physics, School of Physics, Peking University; Tao Xin, Beijing Normal University; Wei Tian, Beijing Normal University*

In this study, we integrated log data and lag sequential analysis to extract behavioral characteristics of scientific inquiry processes in an interactive task for measuring scientific literacy. By analyzing the log data of 334 fourth-grade students, we identified and visualized behavioral patterns and cognitive strategies related to task performance.

**17. Developing and Validating a Short Digital Literacy Assessment for Grade 3 - 12 Students**

*Qianqian Pan, National Institute of Education, Nanyang Technological University; Yuxiao Zhang, Purdue University; Qianru Liang, Jinan University; Sisi Tao, The University of Hong Kong, Faculty of Education; Min Lan, Zhejiang Normal University; Cheng Yong Tan, The University of Hong Kong; Nancy Law, University of Hong Kong*

This study aims to validate a short digital literacy (DL) assessment based on a DL assessment

# FULL SCHEDULE

FRIDAY, APRIL 12

with strong psychometric properties. Analyses showed that this short DL assessment measures a unidimensional DL with good psychometric properties. Both long and short DL assessments exhibited comparable performance trends across gender and grades.

## 18. Latent Profile Analysis: Comparison of Achievement versus Ability-Derived Subgroups of Mathematical Skills

*Sharon Frey, Riverside Insights; Onur Demirkaya, Riverside Insights; Sid Sharairi, Riverside Insights*

This study compares latent profiles derived with student subgroups of varying levels of mathematical skills defined by achievement and ability assessment scores. Achievement and ability cut scores for identifying students at both ends of the mathematics spectrum were applied and the resulting latent profiles within each condition were compared.

## 19. Sparse Variational Approximations For Partially Confirmatory Factor Analysis

*Yi Jin, University of Hong Kong; Jinsong Chen, University of Hong Kong*

Different to MCMC-based methods for partially confirmatory factor analysis models (PCFA), this study proposes using sparse variational approximation approach as a compelling and faster alternative, which can achieve considerable accuracy in terms of computationally efficiency and accuracy-efficiency trade-offs, with scalability to large-scale problems.

## 20. Estimate Benchmark Assessment Cut Scores to Predict State Assessment Performance Levels

*Yi He, Edmentum, Inc.; Sonya Powers, Edmentum, Inc.*

One important purpose of benchmark assessments is to predict students' performance levels on the state summative assessments. This study compares three methods (equipercentile linking, LOESS regression, and ROC analysis) to estimate cut scores on benchmark assessments and investigates the impact of correlations and sample sizes on the effectiveness of prediction.

## 070. Voices from the Measured: The Role of Testing in an Equitable Society

**Invited Session**

**4:55 to 6:25 pm**

**Convention Center: Level 4 - Terrace Ballroom III**

Moderator:

*Michael E. Walker*

Panelists include:

*Leslie Calebrese, Rutgers University*

*Cheryl Logan, University of Pennsylvania*

*Hal Smith, National Urban League*

## NCME WELCOME RECEPTION

*(open to all attendees)*

**Friday, April 12 | 6:25 to 8 pm**

**Convention Center: Level 4 - Terrace Ballroom IV**

SATURDAY, APRIL 13

## 071. Hot Topics in Measurement - Roundtable Discussion I

**Organized Discussion**

**7:45 to 8:45 am**

**Convention Center: Level 4 - Terrace Ballroom IV**

Moderator:

*Jonathan Weeks, ETS*

### AWARDS CEREMONY AND BUSINESS MEETING Challenging Traditional Views of Measurement

9:00 am to 10:00 am

Convention Center: Level 4 - Terrace Ballroom III

### PRESIDENTIAL ADDRESS Michael Walker, NCME President

10:00 am to 11:00 am

Convention Center: Level 4 - Terrace Ballroom III

#### 072. Evaluating Fit

##### Paper Session

11:25 am to 12:25 pm

Convention Center: Level 1 - 120B

Chair:

*Lizzy Wu, University of Illinois Urbana-Champaign*

Participants:

##### Assessing Fit in Common Factor Models Using Empirical Moment Functions

*Youjin Sung, University of Maryland, College Park; Yang Liu, University of Maryland, College Park*

A comprehensive fit assessment framework is developed for common factor models. Residuals are defined using empirical moment functions, which estimate the conditional moments of response variables given common factors. Formal statistical tests are then constructed, and their performance is evaluated via simulations and real data examples.

##### Evaluating Model Replication Efforts Through Prior Predictive Similarity Checking

*Sonja D Winter, University of Missouri, Columbia; Wes Bonifay, University of Missouri; Hanamori Skoblow, University of Missouri; Ashley Watts, Vanderbilt University*

In model-based research, goodness-of-fit (GOF) of the original model (e.g., CFA) to the replication data is routinely provided as evidence of replication. We demonstrate that GOF obscures important differences between the original and replication studies. We present Bayesian prior predictive similarity checking: a tool for rigorously evaluating model-based replication.

##### Implied Probabilities of Polytomous Response Functions for Model-Based Prediction and Comparison

*Ben Domingue, Stanford University; Klint Kanopka, New York University; Esther Ulitzsch, University of Oslo; Lijin Zhang, Stanford Graduate School of Education*

Polytomous response models are typically motivated by assumptions about specific dichotomizations. We build on this by considering all possible dichotomizations. We enumerate such dichotomizations and show that many of these dichotomizations have dramatically different forms when computed for different models. These differences can be used to evaluate model fit.

Discussant:

*Holmes Finch, Ball State University*

#### 073. Reliability and Scoring

##### Paper Session

11:25 am to 12:25 pm

Convention Center: Level 1 - 120C

Chair:

*Cheng Hua, University of Montevallo*

Participants:

### **Designing Scoring Reliability and Instructional Support into Classroom-Based Math Assessments**

*Xingyao Xiao, Berkeley School of Education, University of California, Berkeley; Richard Patz, University of California, Berkeley; Mark Wilson, Berkeley School of Education, University of California, Berkeley*

This study examines how classroom assessments employing open-ended items can attend to scoring reliability in a manner both instructionally supportive and validity enhancing. The approach is implemented for a classroom-focused, math assessment. Data from human-scored student responses are analyzed, with implications for construct validity and score reliability discussed.

### **Examining the Psychometric Impact of Targeted and Random Double-Scoring in Mixed-Format Assessments**

*Yangmeng Xu, University of Alabama; Stefanie A. Wind, University of Alabama*

Double-scoring CR item responses is a common but costly practice in mixed-format assessments. We explored how targeted and random double-scoring impact the psychometric quality of examinee achievement estimates. We found that these two approaches have a similar impact on examinee achievement and fit. We consider implications for research and practice.

### **Extending Multivariate G-theory Designs to Incorporate Congeneric Relationships and Assess Subscale Added-Value**

*Walter Vispoel, University of Iowa; Hyeryung Lee, University of Iowa; Tingting Chen, University of Iowa*

We demonstrate how structural equation models can extend multivariate G-theory designs to allow for congeneric relationships among scores. Results revealed that generalizability coefficients and subscale added-value indices for persons  $\times$  items  $\times$  occasions congeneric designs consistently exceeded those for traditional multivariate generalizability theory designs at both composite and subscale levels.

Discussant:

*Igor Himelfarb, NCBE*

## **074. Longitudinal and Multilevel Models**

### **Paper Session**

**11:25 am to 12:25 pm**

**Convention Center: Level 1 - 121A**

Chair:

*Edgar Sanchez, ACT*

Participants:

### **Contextual Predictors Of Academic Success: Findings From The Pacific Islands**

*Tim Friedman, ACER; Seema Prasad, Educational Quality and Assessment Programme (EQAP)*

The study uses data from the only regional large-scale assessment in the Pacific Islands to investigate which contextual factors predict achievement in primary age students, using multi-level models. Results show that a range of student characteristics and school level factor contribute to models of achievement.

### **Monotonicity in Longitudinal Regressions**

*Judit Antal, College Board; Ying Lu, College Board*

Longitudinal regression models assess student growth over time. In tracking growth, it is reasonable to expect monotonic functions, even though they are not always attainable. This study investigates the use of growth models and provides recommendation for a technique that ensures monotonicity and provides reasonable fit.

### **Specialized Longitudinal Mixture IRT Model Measuring Differential Learning Progress in Target Domains**

*Minho Lee, University of California Los Angeles; Minjeong Jeon, UCLA*

We propose a specialized longitudinal mixture item response model that simultaneously assesses growth in a construct of interest and differential learning progress on target domains. Two empirical examples and a simulation study will be presented to demonstrate the usefulness of the proposed model over existing models.

Discussant:

*Lihua Yao, Northwestern University*

### **075. Considerations in Assessing Special Populations**

**Paper Session**

**11:25 am to 12:25 pm**

**Convention Center: Level 1 - 121B**

Chair:

*Heather Buzick, ETS*

Participants:

#### **Exploring Universal Text-to-Speech Use in Assessment from a Cognitive Demand Perspective**

*Ejike Edeh, University of Arkansas; Heather Buzick, ETS*

Text-to-speech (TTS) is increasingly offered to any student during testing, to expand opportunities for fairness and equity. There is a need to fully understand who uses and benefits from it, the cognitive demands associated with its use, and the factors that may inhibit its use.

#### **Policies and Practices that Support Making More Appropriate Assessment Participation Decisions**

*Sheryl Lazarus, National Center On Educational; Mari Quanbeck, University of Minnesota - Twin Cities*

States struggle to meet ESSA's 1.0% cap on student participation in the alternate assessment. This paper examines the relationship between AA-AAAS participation rates and placement rates in more restrictive educational settings, read aloud allowance on the general assessment, and using "time spent receiving special education services" in making participation decisions.

Discussant:

*Edynn Sato, Sato Education Consulting LLC*

### **076. Oral Reading Fluency Models**

**Paper Session**

**11:25 am to 12:25 pm**

**Convention Center: Level 1 - 121C**

Chair:

*Hyeon-Ah Kang, University of Texas at Austin*

Participants:

#### **A Sentence-Level Joint Testlet Model for Measuring Oral Reading Fluency**

*Yusuf Kara, Southern Methodist University; Akihito Kamata, Southern Methodist University; Cornelis Potgieter, Texas Christian University; Joseph F. T. Nese, University of Oregon; Xin Qiao, Shandong Academy of Sciences*

This study introduces a testlet-based model for measuring the speed, accuracy, and oral reading fluency of students. Similar to testlet response models, a binomial-lognormal hierarchical speed and accuracy model is extended to account for dependency among sentences within passages. The utility of the model is demonstrated on an empirical dataset.

#### **Enhancing Precision in Joint Count-Time Measurement Models: A Testlet Approach**

*Cornelis Potgieter, Texas Christian University; Yusuf Kara, Southern Methodist University; Akhito*

*Kamata, Southern Methodist University*

This study introduces a testlet model for jointly-observed count and time assessment data. Incorporating latent accuracy and speed dimensions, our approach captures item correlations within and between testlets. Method of Moments and Maximum Likelihood parameter estimation methods are developed. Methods are compared through simulation and empirical illustration.

### **Measuring Reading Fluency through Response Time and Accuracy**

*Mingfeng Xue, University of California - Berkeley; Qing Cai, University of California, Berkeley*

We propose a novel approach to measuring reading fluency by recoding responses according to the combination of response accuracy and response time. Then the GPCM is applied. Results show that our proposed method yields estimates with satisfactory reliability and superior validity compared to the joint model (van der Linden, 2007).

Discussant:

*Adam E Wyse, Renaissance*

## **077. Effort Responding**

### **Paper Session**

**11:25 am to 12:25 pm**

**Convention Center: Level 1 - 122A**

Chair:

*Carina McCormick, Southeast Community College Board of Governors*

Participants:

### **Detecting Low-Effort Response Patterns in Computerized Adaptive Tests via a Nonparametric Approach**

*Dandan Chen, Pearson; Elizabeth Adele Patton, Curriculum Associates; Logan Rome, Curriculum Associates*

We applied a nonparametric approach, hierarchical clustering, to group item response time data from 146,519 examinees in Grades 1-8 who completed the i-Ready Diagnostic, a large-scale CAT. We unveiled patterns of examinees' low effort throughout CAT, which has important implications for future test development and effort-moderated response modeling.

### **Impact of Student Engagement within Summative Assessments**

*Marc W Julian, DRC; Litong Zhang, DRC; Xiao Zhang, DRC; Daisy Ye, Houghton Mifflin Harcourt*

Item response times were used as an indicator of examinee engagement during a large-scale test administration. The degree to which engagement and associated impact of disengagement varies across different test administrations, regions, schools, classrooms, and other demographic information is explored to better understand our systems of measurement.

### **Response Strategies In Multiple-Choice Tests: An Option-Position-Based Strategies Inventory**

*Séverin Lions, CIAE - Universidad de Chile; María Paz Blanco, Center for Advanced Research in Education (FB0003), Institute of Education, Universidad de Chile, Santiago, Chile; Gabriel Ortega, Center for Advanced Research in Education (FB0003), Institute of Education, Universidad de Chile, Santiago, Chile; Pablo Dartnell, Center for Advanced Research in Education (FB0003), Institute of Education; Center for Mathematical Modeling (ACE210010 and FB210005); and Department of Mathematical Engineering; all from Universidad de Chile, Santiago, Chile; Consuelo Andrea Soto*

Two hundred interviews about response strategies to solve multiple-choice test items were conducted to determine which option-position-based strategies students use and when. Results showed that most students use option-position-based strategies when they lack content knowledge or time. An inventory of these strategies is presented.

Discussant:

*Christine DeMars, James Madison University*

**078. Methodological Challenges and Solutions for Administering Large-Scale Assessments in India****Coordinated Paper Session****11:25 am to 12:25 pm****Convention Center: Level 1 - 122B**

This symposium presents our collaborative efforts with India's assessment governmental agency to improve assessment methodologies to promote evidence-based policies for equitable opportunities across genders and social classes. The first presentation introduces the National Achievement Survey (NAS), which is one of the key measures of student achievement in India. In the second study, we investigate differences between test designs that administer all subjects of interest to all students and test designs that administer a subset of these subjects. The goal here is to maximize the precision of group-level statistics while taking into account the constraints related to sample size and testing time. In the third study, proficiency estimates of subgroups derived from plausible values (PVs) are compared with those derived with weighted maximum likelihood estimates (WLEs). In the fourth study, principal components analysis is compared to principal components regression and partial least squares in reducing the large number of background variables often seen in IRT latent regression models used in LSA. The final study compares currently used operational ad-hoc procedures for treating missing responses with modern imputation-based approaches.

Session Organizer:

*Jonas Bertling, ETS*

Participants:

**The National Achievement Survey in India***Indrani Bhaduri, National Council of Educational Research and Training, India; Dinesh Prasad Saklani, National Council of Educational Research and Training, India; Jonas Bertling, ETS***Methodological Considerations in Measuring Single or Multiple Subjects in Large-Scale Assessment Design***Han-Hui Por, ETS; Daniel McCaffrey, ETS; Indrani Bhaduri, National Council of Educational Research and Training, India***Comparing Group Scores from Plausible Values and Weighted Maximum Likelihood Estimates***Parth Tusharbhay Soni, Indian Institute of Management Ahmedabad; Daniel McCaffrey, ETS; Indrani Bhaduri, National Council of Educational Research and Training, India***A Comparison of Data-reduction Techniques for IRT Latent Regression***Peter van Rijn, ETS GLOBAL; Jonas Bertling, ETS; Indrani Bhaduri, National Council of Educational Research and Training, India***Modern Techniques for Treatment of Missing Item Responses in Large-Scale Survey Assessments***Kamal Chawla, University of Delaware; Usama Ali, ETS; Peter van Rijn, ETS GLOBAL***079. Missing Data****Paper Session****11:25 am to 12:25 pm****Convention Center: Level 1 - 123**

Chair:

*Xin Qiao, Shandong Academy of Sciences*

Participants:

**Detection of Differential Item Functioning with Missing Values in the Grouping Variable***Isaac Li, American Board of Psychiatry and Neurology; David Shin, Pearson*

Licensure/certification exams that allow examinees to report their demographics voluntarily face missing values in the variables for manifest groups in DIF study. This research explores various approaches that can tackle this problem by comparing their performance in DIF detection in simulation. These methods are applied to a medical exam dataset.

### **Parameter Recovery from Higher Order Item Response Theory Models with Structural Missingness**

*Yale Quan, University of Washington; Chun Wang, University of Washington*

By combining HO-IRT models with structural missingness researchers can control instrument length while still measuring hierarchical general and subdomain constructs. This study provides insight into the application of HOIRT models with structural missingness by exploring how group design, sample size, and correlation influence parameter recovery.

### **Tackling Missing Data from Omitted and Not-reached Items: A Model-based Approach**

*Dongwei Wang, UMass Amherst; Craig Wells, UMass Amherst*

In low-stakes assessment, there is often a substantive amount of missing data observed from omitted and not-reached items. Missing item responses are often MNAR and lead to biased parameter estimates if not handled properly. This study evaluated and compared the accuracy of parameter estimates between classical approaches and model-based approaches.

## **080. Years of Learning: Relating Changes in Student Attainment to Time**

### **Coordinated Paper Session**

**11:25 am to 12:25 pm**

**Convention Center: Level 4 - Terrace Ballroom III**

One of the most prominent ways that educational assessment has entered into main stream media over the last several years is in discussions of the academic impact associated with the COVID-19 pandemic (i.e., learning loss). Numerous headlines have been written communicating the negative impacts of the pandemic on student academic achievement. In most cases, to facilitate communication with the general public, academic impact is reported as months or years of learning. For example, headlines declaring that students lost several months of learning were common. In terms of concise communication, there is a clear advantage to using years of learning as a measure of academic impact over more standard indicators like effect size. Current president of NCME Andrew Ho weighed in on the subject on Twitter declaring his support for reporting academic impact in terms of months of learning despite potential technical shortcomings to better communicate results to non-technical stakeholders. But do the benefits of simplicity outweigh the risks of misinterpretation? In this coordinated session we present papers that highlight the subtle psychometric and practical issues that underlie the conversion of changes in academic attainment to exposure time to learning.

Session Organizer:

*Damian Betebenner, National Center for the Improvement of Educational Assessment*

Participants:

### **Vertical Scale Design and Years of Learning**

*Sanford R. Student, University of Delaware; Derek Christian Briggs, University of Colorado Boulder; Laurie Davis, Curriculum Associates*

### **Using Growth Norms to (In)Validate Years of Learning**

*Damian Betebenner, National Center for the Improvement of Educational Assessment; Charles DePascale, Psychometric Confections, LLC*

### **Contrasting Years of Learning Approaches Using a Common Scale**

*Eric Stickney, Renaissance*

**081. GSIC Poster Session II****Graduate Electronic Board Session****11:25 am to 12:25 pm****Convention Center: Level 4 - Terrace Ballroom IV**

Participants:

**1. AI in Assessment: Revolutionizing Item Development for Professional Licensure Examinations***Oscar Rios, PSI Services LLC; Tony Albano, University of California Davis*

This proposal evaluates items developed using artificial intelligence (AI) within the licensure examination context. By combining empirical findings, established best practices, and theoretical insights, we aim to provide a comprehensive overview that aids in the understanding of the potential of AI-driven assessment tools.

**2. A Modified Bisquare Estimation Procedure for Identifying Partial Disengagement***Claudia J Ventura, University of Connecticut; Joseph A. Rios, University of Minnesota*

Evaluating examinees' true ability becomes notably challenging due to the presence of aberrant responses, especially those that are partially disengaged. In this study, we propose a new procedure to identify and downweight partial disengagement (PD), while addressing the limitations of previous methods.

**3. Comparative Analysis of Zero-Shot and Task-Specific Transformer Models in Argumentation Analysis***Nils-Jonathan Schaller, IPN – Leibniz Institute for Science and Mathematics Education*

In this paper, we present a comparison between a zero-shot language model and a task-specific transformer model specifically trained on the DARIUS corpus. Both models predict argumentative structures in German high school essays with the goal of providing adaptive feedback. We measure their performance using the same DARIUS testing corpus.

**4. Comparing the Performances of Anchor Item Replacement Procedures***Cihan Demir, Washington State University; Matthew Swain, American Board of Internal Medicine*

Drifted anchor item parameters can negatively affect equating processes. Replacement offers a potential solution. We assessed two replacement procedures — only b-parameter and b-parameter + content — utilizing real ABIM datasets. Results show effectiveness in maintaining psychometric quality; however, careful consideration is needed to maintain the mini-test property of anchor item sets.

**5. Concurrent ATA: Integrating MST Simulation Factors into Automated Test Assembly***Hong Chen, The University of Iowa; YuLan Su, Center for Applied Linguistics*

The proposal introduces a concurrent automated test assembly approach for assembling multiple multistage test (MST) panels using mixed-integer linear programming (MIP). Integrating design factors during panel assembly reduces the effect of confounding factors and facilitates conclusions in the simulation study. Moreover, valuable optimization suggestions for the ATA process are included.

**6. Demonstrating Lord and Wingersky 2.5 as a Method for Differential Subscale Functioning***Jared Block, UCLA; Steve Reise, UCLA*

The present research uses the Alcohol Use Disorder scale to show the effectiveness of Lord and Wingersky 2.5 to detect person fit. External correlates are also analyzed, demonstrating the potential use of LW 2.5 as a method of detection of DIF at the subscale level.

**7. DIF and 'Don't Know' Responses in the NFCS's Financial Literacy Questions***Horacio A Rocha, CUNY Graduate Center; Jay Verkuilen, CUNY Graduate Center*

Financial Literacy is an essential life skill in today's very complex world. This paper studies the impact of Don't Know responses on DIF and overall scoring of financial literacy questions on the 2021 NFCS survey. Preliminary conclusions suggests that DIF and responses are affected by these types of responses.

### 8. Evaluating Algorithm Efficacy of an Item-Level Computer Adaptive Test with Routing Stage

*Cassandra Griger, University of Iowa; Ahmed Bediwy, The University of Iowa; Jonathan Templin, University of Iowa*

This study aims to evaluate a hybrid computer-based testing algorithm, the routing stage preceding item-level Computerized Adaptive Test (rCAT), for its efficacy and accuracy in ability measurement. It compares rCAT with traditional CAT under varying conditions. The research aims to optimize algorithm design, potentially benefiting educational and clinical testing.

### 9. Exploring Students' Receptivity to Feedback: A Latent Profile Analysis

*Jan Luca Bahr, Leibniz Institute for Science and Mathematics Education; Lars Höft, Leibniz Institute for Science and Mathematics Education; Thorben Jansen, Leibniz Institute for Science and Mathematics Education*

We performed latent profile analysis on students' receptivity to instructional feedback (RIF) on a dataset of N=1800. We characterized three RIF profiles, predicted profile membership using established factors (cognitive ability, personality traits, prior achievement, age, and gender) that relate to feedback reception, and examined how profile membership influenced writing performance.

### 10. How can Large Language Models Revolutionize Learning Analytics?

*Elisabetta Mazzullo, University of Alberta; Okan Bulut, University of Alberta; Tarid Wongvorachan, University of Alberta; Bin Tan, University of Alberta*

Large language models continue to revolutionize all aspects of education, including learning analytics. In this study, we summarize the literature on the ongoing challenges in the design and application of LA in education, explain how large language models be leveraged to address these challenges, and offer directions for future research.

### 11. Impact Of Zero-Cell Correction Strategies On The Estimation Of Tetrachoric Correlation Matrices

*Jeongwon Choi, Vanderbilt University; Hao Wu, Vanderbilt University*

This study investigates different zero-cell correction strategies in estimating tetrachoric correlation matrices, including which value to add, whether to keep marginals, and where to add. Preliminary results show that the correction strategy significantly affects the positive-definiteness of the correlation matrix. Correlation size and threshold pattern moderate the results.

### 12. IRT Chained Equipercentile Equating

*Min Liang, University of Iowa; Won-Chan Lee, University of Iowa*

This study aims to propose a new IRT chained equipercentile (CE) equating method under the common-item nonequivalent groups (CINEG) design. The new method utilizes the IRT fitted score distributions instead of the observed score distributions in traditional CE equating. The effects of ability distribution on the equating results were also evaluated.

### 13. MST Calibration: Impact of Software and Priors on Parameter Recovery

*Mubarak O. Mojoyinola, The University of Iowa; Ahmed Bediwy, The University of Iowa; Jing Ma, The University of Iowa; Anthony D. Fina, University of Iowa*

Given the proliferation of commercial and open-source IRT software, with varying capabilities and supporting documentation, the evaluation and comparison of these software programs is critical. This study aims to evaluate item parameter recovery in an MST for the most popular programs. Examined factors include priors, sample size, and ability distributions.

### 14. Performance and Authentic Assessment Practices in Basic Schools in Ghana

*Francis Ankomah, Ohio University; Frank Agyemang Oppong, Ohio University*

The study examined the practice of performance/authentic assessment among teachers by adopting explanatory sequential mixed methods design. A sample of 1264 basic school teachers were engaged. It emerged that teachers had a negative perception, which influenced their willingness to practice. Recommendations were made to the Ministry of Education.

- 15. Psychometric Properties of Composite Theta Scores for Simple Structure MIRT Model**  
*Huan Liu, University of Iowa; Won-Chan Lee, University of Iowa*  
This study introduces analytic formulas and procedures for estimating CSEMs, reliability coefficients, and classification indices for the composite theta scores under the framework of SS-MIRT model. Preliminary results from two real data sets, one mixed-format assessment and one test spanning multiple content domains, are presented to illustrate the proposed methods.
- 16. Simulation Studies for Computer Adaptive Testing Algorithms in an English Vocabulary Test**  
*Haoyang Yu, Accessible Teaching, Learning, and Assessment Systems (ATLAS) at Achievement & Assessment Institute*  
This study simulates Computer Adaptive Testing (CAT) for ESL vocabulary placement using a 96-item bank and the Generalized Partial Credit Model. It finds no superior algorithm for measurement accuracy but identifies efficient ones that expedite testing, benefiting marginalized groups.
- 17. Teachers' and Administrators' Grading Perceptions**  
*Sarah Ruth Morris, University of Arkansas; Fabio Andres Parra-Martinez, University of Arkansas*  
We will explore teachers' and administrators' grading perceptions using a sample of Arkansas educators. In prior work, we find teachers who have liberal political ideologies are more likely to prefer grading equity practices. While we wait on results for the principals, we will measure how principals prefer grading equity practices.
- 18. The Changing Role of Testing in Admissions Decisions**  
*Yen Vo, University of Iowa; Jeongmin Ji, University of Iowa; Xiaoting Zhong, University of Iowa; Catherine Welch, University of Iowa; Stephen Dunbar, University of Iowa; Anthony D. Fina, University of Iowa*  
The impact of the pandemic and social inequality discussions have significantly changed the traditional college admission process. This study explores replacing traditional admission tests with statewide assessments that are accessible and aligned for all students. Results support consistent admission scores while also creating pathways for all students.
- 19. Towards Automated Formative Evaluation Of Teachers' Facilitation Of Simulated Science Discussions**  
*Michael Ilagan, ETS Canada; Jamie Mikeska, ETS; Beata Beigman Klebanov, ETS*  
Effectively facilitating students' discussions while encouraging them to engage with each other's ideas is a difficult skill to master for teachers. In the context of simulated science discussions for developing this skill, we developed an automated system to assess the teacher's performance as well as provide positive examples as feedback.
- 20. Using Large Language Models for Automatic Item Generation to Create Multiple-Choice Questions**  
*Nour Armoush, University of Alberta; Mark J Gierl, University of Alberta*  
This research-in-progress investigates an automatic item generation (AIG) methodology which utilizes subject matter expert (SME) model and large language model capabilities to create a large number of high quality items. We follow Gierl et al. (2017) three-step method in AIG. Items will be evaluated by textual features and by SME.
- 21. Using Process Data to Explain Gender Differences in PIRLS 2021**  
*Dihao Leng, Boston College; Matthias von Davier, Boston College*  
Girls consistently outperform boys on average in international large-scale reading assessments. This study categorized students into three types utilizing mixture models on PIRLS 2021 process data: "Rapid", "Middle", and "Slow". The "Rapid" type includes more boys. The study examines whether the response process types explain gender achievement gaps in reading.
- 22. Context-Rich Mathematics Item Generation Using Generative Language Models**  
*Olushola Olufemi Soyoye, University of Delaware*  
This project aims to investigate the viability of using a large language model to automatically generate appropriate contexts based on the target construct and numerical data. We fine-tuned a pre-trained language model to generate mathematics items. Results show that the fine-tuned model can achieve accuracy and control in automated item generation.

### 23. Validation Of Dass And The Brief RCOPE Inventory Among Ghanaian Population

*Enoch Tsey, University of Cape Coast; Regina Mawusi Nugba, University of Cape Coast*

Lack of item clarity on the DASS and the Brief RCOPE Scale possess a lot of validity and reliability threats to researchers/clinicians who often administer these scales to respondents in Ghana. This study seeks to validate the aforementioned scales among the Ghanaian population through the lenses of CTT and IRT.

### 082. Leveraging Experiences of Latin American and Muslim Women for Equitable Measurement Practices

**Organized Discussion**

**1:15 to 2:45 pm**

**Convention Center: Level 1 - 120B**

This panel discussion session moderated by Dr. Fiona Hinds aims to explore the narratives and experiences of Latin American and Muslim women in the field of educational measurement, and a liberation model for re-imagining an equitable system for the measurement industry that is inclusive of racial and gender equity, local culture, and language. The presenters: Dr. Shelia Lallmamode, Independent Assessment Researcher Dr. Xaviera Gonzalez-Wegener, Dr. Leslie Rosales and Maria Hamdani, MSc will illustrate the findings of qualitative research studies and provide insight into the intersectionality of identities and experiences of Latin American and Muslim women to uncover challenges and barriers. The researchers will provide insights to explore transformative viewpoints of creating a liberation model that addresses frameworks and strategies for overcoming racial, gender, cultural and language inequities in educational measurement.

Session Organizer:

*Fiona Hinds, Independent Consultant*

Presenters:

*Leslie Vanessa Rosales de Veliz, JML Measurement and Testing Services, LLC*

*Xaviera Gonzalez-Wegener, UCL Institute of Education*

*Sheila Lallmamode, Independent Assessment Researcher*

Discussant:

*Maria Hamdani, Center for Measurement Justice*

### 083. Reconceptualizing Diagnostic Classification Models: Applications and New Development Coordinated Paper Session

**1:15 to 2:45 pm**

**Convention Center: Level 1 - 120C**

Diagnostic paradigms have gained prominence in educational measurement following increase in demands for testing instruments that are closely aligned with instructional practices and can present actionable diagnostic feedback. Cognitive diagnostic models (CDMs) or diagnostic classification models (DCMs) have since emerged as a promising approach capable of identifying students' strengths and weaknesses on multiple fine-grained knowledge skills. Recently, research has been focused on extending classical models that exploit the wealth of student information available in modern assessments to more accurately reflect their knowledge states. The end goal is to make better informed diagnostic decisions and thus, enhance the learning process for each student. In this organized paper session, the speakers will discuss applied and methodological advances in CDMs working toward this purpose. Topics include leveraging process data using a subtask-based approach for diagnostic modeling via deep learning; and incorporating various response types under a novel paradigm of identifiable general-response diagnostic models; recognizing the effects of and accounting for differences in test-taking behaviors within CDMs such as rapid guessing and omission and assessing both intervention effectiveness and impact of covariates in longitudinal settings. An expert in diagnostic models will be the discussant for this coordinated paper session.

Session Organizer:

*Yon Soo Suh, NWEA*

# FULL SCHEDULE

SATURDAY, APRIL 13

Chair:

*Shiyu Wang, University of Georgia*

Participants:

**Diagnostic Assessment via Process Data based on Subtask Analysis**

*Susu Zhang, University of Illinois at Urbana-Champaign*

**New Paradigm Of Identifiable General-Response Diagnostic Models: Beyond Categorical Data**

*Seunghyun Lee, Columbia University; Yuqi Gu, Columbia University*

**A Motivation-based Cognitive Diagnostic Model for Insufficient Responses Detection**

*Yingshi Huang, University of California, Los Angeles; Shiyu Wang, University of Georgia; Yanfang Pan, Beijing Normal University; Xiangyu Lu; Ping Chen, Beijing Normal University*

**Considerations and Recommendations for Rapid Guessing with Examinee Motivation Filtering in DCMs**

*Nicolas Mireles, James Madison University; Yu Bao, James Madison University*

**Assessing Intervention Effectiveness and the Impact of Covariates Using Longitudinal Cognitive Diagnosis Models**

*Benjamin Kweku Lugu, University of Alabama; Ying Chen, University of Illinois at Chicago; Yue Yin, University of Illinois at Chicago; Wenchao Ma, University of Alabama*

Discussant:

*W. Jake Thompson, University of Kansas*

## 084. Investing in Research and Development to Foster Innovation in PISA

**Coordinated Paper Session**

**1:15 to 2:45 pm**

**Convention Center: Level 1 - 121A**

This session presents a collection of studies undertaken as part of the PISA Research, Development and Innovation (RDI) initiative, through which the PISA Governing Board aims to introduce conceptual, methodological and operational advances in how the OECD's PISA study is conducted. PISA faces a number of challenges to the adoption of frontier technologies and of new methods, chief among them the large number of languages (about 100) and the wide variation in national contexts in which the assessment is conducted. These challenges are shared with other international large-scale assessments, such as the IEA's TIMSS or PIRLS. All presentations in this session illustrate how these challenges play out throughout the assessment cycle, but also how they give rise to fertile research questions. The session will in particular present results in three areas: • the use of artificial intelligence and natural-language processing technologies for the automatic coding of open responses (2 presentations); • the measurement of students' socio-economic status; • the use of adaptive technologies to increase accessibility in an international context;

Session Organizer:

*Francesco Avvisati, OECD*

Participants:

**Improving the Measure of Socio-Economic Status: Lessons from the PISA 2022 Cycle**

*Francesco Avvisati, OECD; Celine Wuyts, OECD*

**Improving Accessibility in PISA for Students with Special Needs**

*Ava Guez, OECD; Ketan, University of Massachusetts; Mario Piacentini, OECD*

**Beyond Human Raters: Using AI and NLP in PISA's Creative Thinking Tests**

*Ricardo Primi, University of Sao Francisco; Roger E. Beaty, Penn State University; John D. Patterson, Penn State University; Mathias Benedek, University of Graz; Ivandre Paraboni, Universidade de São Paulo; Denis Dumas, University of Georgia; Peter Organisciak, University of Denver; Tiago A Calíço, Organization for Economic Cooperation and Development; Mario Piacentini, OECD*

**From Free-Form Job Descriptions to Standardized Occupational Codes: A Machine-Learning Approach**  
*Tiago A Caliço, Organization for Economic Cooperation and Development; Francesco Avvisati, OECD*

Discussant:  
*Emmanuel Sikali, NCES*

**085. Putting Research into Practice: Case Studies Developing Reporting Tools with Stakeholder Engagement**  
**Coordinated Paper Session**  
**1:15 to 2:45 pm**  
**Convention Center: Level 1 - 121B**

Although educational research is often motivated by an interest in improving achievement-score-related measures for stakeholder use, researchers are rarely involved in the translation of that research into practice. This session highlights four case studies in which researchers engaged with diverse groups of stakeholders in rigorous, iterative processes of development and engagement to create digestible reporting tools that avoid misinterpretations and are accessible to a variety of stakeholders. We demonstrate how partnerships with stakeholders via co-design, focus groups, usability studies, surveys, or stakeholder meetings are key to designing reporting tools with equitable access to all intended audiences. We highlight lessons learned about communicating complex concepts and measures to various audiences. The first presentation focuses on bridging the gap between assessment results and instructional action with classroom-level score reports for a through-year assessment program. The second highlights the process of creating school/district-level summaries of an improved aggregate-level growth score. The third provides a few examples of the development process for tools in the National Assessment of Educational Progress (NAEP) reporting ecosystem, including the recently developed NAEP achievement gap tool. The fourth illustrates the application of the iterative multistep framework in the development of parent/family reports for early learning assessments in one state.

Session Organizer:  
*Katherine Furgol Castellano, ETS*

Chair:  
*Katherine Furgol Castellano, ETS*

Participants:  
**What Does this Mean for My Next Unit? A Classroom Reporting Approach**  
*Nathan Dadey, Center for Assessment; E. Caroline Wylie, Center for Assessment*

**A State's Journey of Communicating School/District Summaries of Student Growth**  
*Joseph A. Martineau, ETS; Kimberly Mundhenk, California Department of Education; Katherine Furgol Castellano, ETS*

**How Stakeholder Needs Inform An Equitable And Meaningful Nation's Report Card**  
*Robert Finnegan, ETS; Madeline Goodman, ETS; Timothy Shaw, Forum One Communications*

**Need For Ongoing Monitoring After Operational Deployment: Indiana ISPROUT Reports**  
*Priya Kannan, WestEd; Matthew Brunetti, WestEd; Jaylin Nesbitt, WestEd; Lynn Schemel, Indiana Department of Education*

Discussant:  
*Karoline Jarr, Curriculum Associates*

**086. Tackling Challenges in Automated Scoring: Data Augmentation, Multilingualism, and Feedback Mechanisms****Coordinated Paper Session****1:15 to 2:45 pm****Convention Center: Level 1 - 121C**

Educational assessments have been undergoing a transformative evolution, with automated scoring systems emerging as a nexus between advanced technology and pedagogical needs. This session illuminates the innovations harmonizing current challenges in automated scoring, offering a unified vision for the future of scoring mechanisms. Central to this vision is data augmentation, a strategy that not only addresses challenges like class imbalances but also fortifies the foundation for more precise and inclusive scoring models. As education transcends borders, the imperative to adeptly handle multilingual responses also becomes more evident. This necessitates automated scoring systems that are linguistically versatile, ensuring fairness and accuracy across diverse linguistic landscapes. Beyond mere scoring, the session also delves into the potential of automating feedback and distractor generation, especially in assessments with multiple-choice items. This signifies a shift towards a more holistic, responsive, and adaptive assessment experience. Collectively, this session with five papers will present a cohesive narrative, emphasizing that the future of automated scoring transcends mere efficiency. Our session envisions a landscape marked by inclusivity, adaptability, and a holistic approach to education, where technology and pedagogy converge to enhance the assessment experience for all stakeholders.

Session Organizer:

*Xiaoming Zhai, University of Georgia*

Chairs:

*Xiaoming Zhai, University of Georgia**Hong Jiao, University of Maryland*

Participants:

**Data Augmentation for Class Imbalance in Developing Generic Models in Science Assessment***Hong Jiao, University of Maryland; Chandramani Fnu, University of Maryland; Xiaoming Zhai, University of Georgia***Text Augmentation for Enhancing the Accuracy of Automated Scoring in Low-Resource Languages***Tahereh Firoozi, University of Alberta; Okan Bulut, University of Alberta; Guher Gorgun, University of Alberta***Data Augmentation Using GPT-4 for Unbalanced Dataset in Automated Assessment***Luyang Fang, University of Georgia; Gyeong-Geon Lee, University of Georgia; Xiaoming Zhai, University of Georgia***AI-Based Automated Scoring Of Multilingual Responses In International Large-Scale Assessment***Ji Yoon Jung, Boston College; Lillian Tyack, Boston College; Matthias von Davier, Boston College***Exploring Automated Distractor And Feedback Generation For Math Multiple-Choice Questions***William McNichols, UMass Amherst; Wanyong Feng, UMass Amherst; Jake Lee, UMass Amherst; Alex Scarlatos, UMass Amherst; Andrew Lan, University of Massachusetts at Amherst; Digory Smith, Eedi; Simon Woodhead, Eedi*

Discussant:

*Susan Lottridge, Cambium Assessment*

### 087. Computer Adaptive Testing Applications

#### Paper Session

1:15 to 2:45 pm

Convention Center: Level 1 - 122A

Chair:

*Yuan-Ling Liaw, IEA Hamburg*

Participants:

#### **A Score-Based Method for Detecting Item Compromise and Preknowledge in CAT**

*Kylie Gorney, Michigan State University; Chansoon Lee, American Board of Internal Medicine; Jianshen Chen, College Board*

We develop a score-based method for detecting compromised items and examinees with preknowledge in computerized adaptive tests. A detailed simulation study reveals that the new method produces small false positive rates and large true positive rates across most conditions.

#### **Enhancing Global Efficiency in CAT: A Maximin Information Criterion for Item Selection**

*Hsiu-Yi Chao, Soochow University; Jyun-Hong Chen, National Cheng Kung University*

To enhance global test efficiency, this study applies maximin information criterion (MMI) for item selection. MMI tailors item selection to each examinee's provisional trait estimate, improving the efficiency of item administration. Simulation results indicate that MMI outperforms other ISRs by offering comparable RMSE and a more balanced item pool usage.

#### **Adaptive Testing for Multiple-Choice, Multiple-Attempt Test Items**

*Yikai Lu, University of Notre Dame; Ying Cheng, University of Notre Dame*

This study proposes the combination of multiple-choice, multiple-attempt test items and CAT (MM-CAT). Our small-scale simulation study demonstrated that CAT can benefit from allowing multiple attempts on items by either improving the accuracy of theta estimates in fixed-length CAT or reducing the test length of variable-length CAT.

#### **Measuring Test-Taking Engagement on Computerized Adaptive Test**

*Guanyu Chen, The University of British Columbia; Ching-Lin Shih, National Sun Yat-sen University; Yan Liu, Carleton University*

This study investigated test-taking engagement through RT-based and model-based approaches using CAT data, comparing the results of both approaches. It also explored the relationship between English proficiency and test-taking engagement, offering insights into the underlying mechanisms. These results lay the foundation for subsequent studies integrating demographic and psychological variables.

Discussant:

*Laurie Davis, Curriculum Associates*

### 088. A Conversation about Rebuilding Pathways to Educational Measurement

#### Organized Discussion

1:15 to 2:45 pm

Convention Center: Level 1 - 122B

The pathway from college to careers is changing rapidly, including the educational measurement pathway. For decades we have seen a slow decline in measurement graduate program enrollments. With a quickening pace in this trend over the last few years, and current professionals leaving traditional measurement positions, there is a need to consider what steps need to be taken to support the field in growing. The Educators of Measurement SIGIMIE spent the year gathering information on these issues. The session brings together a panel of diverse viewpoints to engage in conversation with the audience about how to build sustainable efforts for recruiting and retaining future measurement professionals. We will present what we have learned from the last year and pose questions to the audience via text message and our panel to cover a variety of topics on this issue. The conversation will lay the foundation for recommendations for next steps with NCME and our educational and industry partners.

# FULL SCHEDULE

**SATURDAY, APRIL 13**

Session Organizer:  
*Deborah Harris, University of Iowa*

Presenters:  
*Anne Corinne Huggins-Manley, University of Florida*  
*Brian French, Washington State University*  
*Justin L. Kern, University of Illinois at Urbana-Champaign*

## **089. Scaling, Linking, and Equating in Automatic Item Generation Organized Discussion 1:15 to 2:45 pm Convention Center: Level 1 - 123**

Automatic Item Generation (AIG) refers to the process of developing items from various inputs, such as item models, with the aid of computer algorithm. Although scaling, linking, and equating (SLE) is a common topic in the literature, there have been relatively few research discussing SLE in the context of AIG. Thus, the goal of this organized discussion session is to bring professionals from both academia and industry together to discuss the applications and issues of SLE in AIG.

Session Organizers:  
*Kyung Yong Kim, University of North Carolina Greenboro*  
*Min Wang, College Board*

Presenters:  
*Richard Melvin Luecht, University of North Carolina at Greensboro*  
*Ikkyu Choi, ETS*  
*Weiwei Cui, College Board*  
*Thomas Proctor, College Board*

## **090. Promoting Competency-Based Education: Competency Frameworks and Classroom Implementation Coordinated Paper Session 1:15 to 2:45 pm Convention Center: Level 1 - 124**

The 21st century is a skills century. An effective educational system needs to reflect students' nonlinear educational pathways and focus on their competencies, rather than fixed seat time. This coordinated session aims to shed light on the most salient aspects of competence-based education (CBE), including theoretical frameworks, skills definition, application of AI in skills extraction, and innovative district implementation. It includes five papers: • A comprehensive overview of competency frameworks and a set of durable skills that are cross-cutting and predictive of future success. • A discussion of how improvement science can be used to facilitate the implementation of CBE to enable teachers to offer personalized guidance and collaborate with a network of educators. • An illustration of the innovative approach the Cajon Valley School District in California adopts in implementing CBE through Learner Employment Records. • Insights on critical skills expected of employers and trends data in career trajectory and how that impacts CBE, and • The application of AI in extracting skills from unstructured data to enable the recognition of a wide variety of prior learning. The five presentations jointly provide a unique opportunity for the audience to gain insights into CBE, connecting researchers, practitioners, administrators, and employers.

Session Organizer:  
*Ou Lydia Liu, ETS*

Participants:  
**Skills for the Future: A New Vision for Skills-Based Assessment**  
*Ou Lydia Liu, ETS; Yuan Wang, ETS; Lei Liu, ETS; Guangming Ling, ETS*

# FULL SCHEDULE

SATURDAY, APRIL 13

## **Building Collective Capacity for Competency Based Education: The Promise of Networked Improvement**

*David Sherer, Carnegie Foundation; Paul LeMahieu, Carnegie Foundation*

## **Cajon Valley's Data Journey**

*Chris Collins, Cajon Valley Union School District; David Miyashiro, Cajon Valley Union School District; Karen Minshew, Cajon Valley Union School District; Bethany Schwappach, Cajon Valley Union School District; Jana Leonard, Cajon Valley Union School District*

## **From Classroom to Boardroom: Decoding the Skills of High Career**

*Arcs Yustina Saleh, The Burning Glass Institute*

## **AI Mapping Of Learning And Working Experience Into A Skill-Based Profile**

*Guangming Ling, ETS; Carol McGregor Forsyth; Blair Lehman, ETS; Andrew Hoang, ETS; Keith Kiser, ETS; Diego Zapata-Rivera, ETS*

Discussant:

*Anthony Bryk, Carnegie Foundation*

## **091. Classroom Assessment and Culturally Responsive Teaching**

### **Paper Session**

**1:15 to 2:45 pm**

**Convention Center: Level 1 - 125**

Chair:

*Roti Chakraborty, Georgia State University*

Participants:

### **A Discussion About Culturally Responsive Assessment for Classroom Assessment Practices**

*Janine Jackson, Morgan State University; Kyndra Middleton, Howard University*

This study presents practical considerations for implementing culturally responsive strategies into the teaching and learning cycle to facilitate alignment between teaching activities and classroom assessments in K-12 schools. We examine classroom assessment design, scoring, reporting, and use. Recommendations are discussed that consider teachers, administrators, parents, and students experiences and responses.

### **A Psychometric Argument for Culturally Responsive Assessment**

*Catherine Taylor, University of Washington; Steve Ferrara, Cognia/Steve Ferrara*

This presentation demonstrates how test development and psychometric methods guarantee standardized tests are not appropriate for all students. Development practices are dominated by ideas and backgrounds of middle-class White culture. The more students differ from mainstream White culture, the less likely they are to perform well on standardized tests.

### **Attending to Politics to Reduce Inequities in Assessment Policy and Practice**

*Sarah Beach, University of Virginia*

This study considers how politics play a key role in measurement practice through a one-state case study that is then discussed within the larger national narrative. Theoretical framing about how politics can be attended to in order to address the NCME theme of reducing inequities is considered.

### **Designing Assessments to Support Instruction**

*Brooke Nash, University of Kansas; Mary Majerus, Missouri Department of Elementary and Secondary Education; Shaun Bates, Missouri Department of Elementary and Secondary Education*

New assessment models that better support instruction are needed. Current models tend to create arbitrary checkpoints of student proficiency unrelated to instruction and learning. We describe instructionally embedded assessments designed to provide timely data and meet summative reporting needs. Findings from teacher focus groups on assessment features are presented.

Discussant:

*Joshua Sussman, University of California, Berkeley*

**092. Foundational Challenges to Foundational Competencies: Debating the Task Force Report  
Organized Discussion****1:15 to 2:45 pm****Convention Center: Level 4 - Terrace Ballroom III**

What are “foundational competencies in educational measurement”? What knowledge, skills, and abilities must modern students of educational measurement possess in order to continue learning in our field? In March of 2023, a 12-member NCME Presidential Task Force delivered a report answering these questions, identifying three competency domains and five competency subdomains, and demonstrating how educational measurement careers and curricula develop these competencies. In this symposium, presenters will put the Task Force Report to the test by challenging the framework along four substantive dimensions. First, Alina von Davier argues that the rapid rise of Artificial Intelligence demands greater emphasis on intersections between computational competencies and educational measurement competencies. Second, Pohai Kukea Shultz recommends a stronger emphasis on community engagement as a necessary condition for valid use. Third, Sue Brookhart shows how Task Force competencies compare to the 2015 Classroom Assessment Standards and argues that the frameworks should interact. Fourth, Kristen Huff discusses how the Task Force framework can advance the revision of the 2014 AERA/APA/NCME Standards for Educational and Psychological Testing, particularly with respect to issues of Fairness and Context. Derek Briggs, who chaired the Task Force when he was NCME President, will moderate discussion among presenters and the audience.

Session Organizer:

*Andrew Ho, Harvard Graduate School of Education*

Moderator:

*Derek Christian Briggs, University of Colorado Boulder*

Presenters:

*Alina A. von Davier, Duolingo**Pohai Kukea Shultz, University of Hawaii at Manoa**Susan Brookhart, Duquesne University**Kristen Huff, Curriculum Associates***093. Innovation Demonstration II****Electronic Board Session****1:15 to 2:45 pm****Convention Center: Level 4 - Terrace Ballroom IV**

Participants:

**1. Affordable Population Scale Assessments through Mobile Paper-Digital Integration***Nirmal Patel, Smart Paper*

Paper, when paired with computer vision, provides a cost-effective method for population-scale assessments, enabling data-informed instruction for all students, irrespective of digital access. This demonstration shows the elements of the mobile paper-digital integration system that solves the assessment scale issue, as shown in India at 4.5 million student scale.

**2. Comparison Of Free Annotation Packages For Data Labeling***Janet Mee, NBME; Milton Quranda, NBME; Andrew Houriet, NBME; Thai Quang Ong, National Board of Medical Examiners; Wenli Ouyang, National Board of Medical Examiners; Christopher Runyon, NBME; Victoria Yaneva, NBME*

ML and AI applications in assessment frequently require labeled or annotated data. Measurement professionals may struggle to obtain and manage annotations due to software constraints. This demonstration compares the features and functionality of three no-cost annotation solutions: Scale Studio, John Snow Labs NLP Lab, and BRAT.

### 3. Generative AI Chatbot As NAEP Helpdesk

*Ting Zhang, American Institutes for Research; Yuqi Liao, American Institutes for Research; Paul Bailey, American Institutes for Research; Blue Webb, American Institutes for Research; Sinan Yavuz, American Institutes for Research; Emmanuel Sikali, NCES*

We fine-tuned the generative AI model and developed a chatbot to help with NAEP data and methodology questions. Our goal is to improve accessibility, efficiency, and user satisfaction in accessing and understanding NAEP data by implementing a generative AI chatbot. We focused on methods that eliminate misinformation from chatbot output.

### 4. Reducing Data Usage Barriers: Using Bookdown for Flexible, Open-Source Data Documentation

*Karen Yi, American Institutes for Research; Grant Adams, American Institutes for Research; Juanita Hicks, American Institutes for Research; Ruhan Circi, American Institutes for Research*

Data releases can involve dozens of data files and accompanying documentation, which can quickly become confusing. To remove barriers to data usage, reduce time spent in user support, and promote collaboration among data users, we centralized documentation using bookdown, an R package that builds distributable, living books from R Markdown.

### 5. R Shiny Application for Regression Discontinuity Analysis with Measurement Models

*Muwon Kwon, University of Maryland - College Park; Youngjin Han, University of Maryland College Park; Youjin Sung, University of Maryland, College Park; Ji Seung Yang, University of Maryland; Yang Liu, University of Maryland, College Park*

We introduce an R shiny application for regression discontinuity analysis with latent running and/or outcome variables measured by categorical and/or continuous responses. The web-based application allows researchers to examine heterogeneity and generalizability of average treatment effect by adopting the novel approach with a straightforward graphical interface and produce publication-ready figures.

### 6. TDCM: An R Package for Estimating Longitudinal Diagnostic Classification Models

*Matthew James Madison, University of Georgia; Sergio Haab, University of Iowa; Minjeong Jeon, UCLA; Michael Cotterell, University of Georgia*

Longitudinal diagnostic classification models (DCMs) are psychometric tools for analyzing diagnostic assessments administered over multiple occasions. Different from traditional psychometric frameworks for growth, longitudinal DCMs provide categorical growth estimates with criterion-referenced interpretations. This brief hands-on training introduces and demonstrates a newly developed R package for estimating longitudinal DCMs.

### 7. Your AI-powered Exam Development Associate – Welcome to the World of Possibilities!

*Vinita Talreja, AWS; Jennifer Davis, Amazon Web Services (AWS)*

We will demonstrate a novel and user-friendly chatbot that uses the power of Large Language models to help with common tasks in the assessment development lifecycle, from Job Task Analysis to Item writing and review. We will discuss how we have used the chatbot for operational work and lessons learned.

### 8. Template-Based Automated Item Generation and Management System

*Yanyan Fu, GMAC*

The template-based automated item generation (TAIG) is a promising approach for effectively developing exam questions. However, it is often constrained by technology limitations because creating flexible and complex templates and managing the TAIG development process easily have not been achieved simultaneously. This demonstration offered a solution to the problem.

## 094. What a Cluster! Examining Local Item Dependence in Multidimensional Science Assessment

### Coordinated Paper Session

**3:05 to 4:35 pm**

**Convention Center: Level 1 - 120B**

In assessing multidimensional science standards using items related to phenomena, questions about item clusters are likely to arise. Will we meet the assumption of local item independence (LID) which is fundamental in Item Response Theory? The four papers presented in this session each address cluster-based assessment of multidimensional science standards: a large-scale testing vendor's findings from

## FULL SCHEDULE

SATURDAY, APRIL 13

the largest operational application of a bifactor model across state summative assessment programs; a university center's research questioning the traditional method of model comparison to examine testlet effects if the intended purpose is individual measurement rather than parameter estimation; a large scale testing vendor's analysis of two summative assessments where evidence of LID was not found; and, a research organization's findings from two unique state's summative assessments where comparisons between stand-alone items and cluster-based tasks could be made, and no evidence of LID was found. The session discussant worked on the NRC science framework, several state assessments, and the updated framework for NAEP science, and will engage the presenters and audience in lively academic dialogue around item dependency in cluster-based science assessments.

Session Organizer:

*Sarah Quesen, WestEd*

Participants:

### **Calibration Model for Three-Dimensional Science Assessments**

*Allison Boykin, Cambium Learning Group; Suhwa Han, University of Texas at Austin; Frank Rijmen, Cambium Assessment*

### **Examining Local Item Dependence Effects in Testlets: Practical and Statistical Significance**

*Perman Gochyyev, University of California, Berkeley; Mark Wilson, Berkeley School of Education, University of California, Berkeley*

### **Exploring LID in Two NGSS Large Scale State Testing Programs**

*Timothy O'Neil, Pearson*

### **Evaluating Item Clusters in Multidimensional Science Assessment**

*Sarah Quesen, WestEd; Marianne Perie, WestEd; Daniel Murphy, WestEd*

Discussant:

*James Pellegrino, University of Illinois at Chicago*

## **095. Equity by Design: New Models for Implementing Universal Design**

### **Coordinated Paper Session**

**3:05 to 4:35 pm**

**Convention Center: Level 1 - 120C**

Principles of universal design have the potential to improve the assessment of diverse populations by expanding the ways that students may engage with assessments and demonstrate their knowledge and skills (CAST, 2018). Despite its potential to address equity, universal design has a limited but growing evidence base on its application for assessment (Ketterlin-Geller, 2005; Preiser, 2008; Thompson et al., 2002). Our community can provide insights into new forms of measurement of engagement (process data), synthesize research findings, and propose new models for communicating levels of efficacy and rigor of evidence informed research. These new models can then inform the next generation of the Joint Standards for technical quality (AERA, NCME, APA Joint Standards) and integrate with existing standards for digital accessibility (W3C) and culturally responsive assessment (Montenegro & Jankowski, 2020). Stepping beyond high stakes assessments, this collection of papers will address how new ideas of universal design can be integrated into assessment practices. The papers include a range of topics: a new framework, alignment considerations, student choice, and adaptations to improve accessibility. A discussant will respond on how the ideas in these papers can be used to support more equitable and universally designed assessments.

Session Organizer:

*Melissa L. Gholson, ATLAS, University of Kansas*

Participants:

### **A Case for Reimagining Universal Design in Assessment Systems**

*Cara C Laitusis, Center for Assessment; Meagan Karvonen, University of Kansas*

### **Systems Universal Design for Standards Alignment**

*Anne H Davidson, EdMetric LLC*

### **Studying the Role of Choice in Assessment Tasks: Process Evidence and Impacts**

*Meagan Karvonen, University of Kansas; Brooke Nash, University of Kansas; Robert Dolan, Diverse Learners Consulting*

### **Designing Assessments to Improve Access: A Case from Early Mathematics**

*Leanne Ketterlin Geller, Southern Methodist University; Christopher Johnstone, University of Minnesota; Anne Hayes, Inclusive Development Partners*

Discussant:

*Molly Faulkner-Bond, WestEd*

## **096. Based on True Story: Developing CAT-based High-stakes Exam with Latest Innovations**

### **Coordinated Paper Session**

**3:05 to 4:35 pm**

**Convention Center: Level 1 - 121A**

Since its inception in 1954, the Graduate Management Admission Test (GMAT) has been the leading assessment for admissions to 7700+ MBA and business-related programs across 2400+ global universities. With the evolving demography of test-takers, propelled by the geo-political dynamics and the emergence of the Generation Z cohort, coupled with the nuanced shifts in the considerations and process for admission decision making, and the concomitant advancements in information technology, it became crucial for the GMAT exam to undergo continuous evolution to maintain its relevance in the market and effectively cater to the needs of its users and stakeholders. Recently unveiled in 2023, the latest iteration of the GMAT, aptly named GMAT Focus, represents a monumental transformation of the exam, marking the most significant overhaul in its 70-year history. This comprehensive redesigning, underpinned by a decade of research and three years of meticulous development and execution, epitomizes the commitment to progress. In this coordinated session, the psychometricians intricately involved in the evolution of GMAT Focus unveil novel innovations integrated into the new test format. Moreover, they candidly share insights gleaned from real-world experiences, alongside the challenges surmounted during the course of its conceptualization and implementation.

Session Organizer:

*Kyung (Chris) T. Han, GMAC*

Participants:

### **New CAT Design for Optimizing Measurement Efficiency and Test Security**

*Kyung (Chris) T. Han, GMAC; Yanyan Fu, GMAC; Sung-Hyuck Lee, GMAC*

### **Improving Test-Taking Experience of CAT with Review & Edit Feature**

*Paulius Satkus, GMAC; Kyung (Chris) T. Han, GMAC*

### **Designing Score Scales and Student Reports in Presence of Test Construct Changes**

*Jaime Malatesta, GMAC; Hwanggyu Lim, GMAC; Kyung (Chris) T. Han, GMAC*

### **Conducting Triangulated Predictive Validity Study Using Test Data from Past**

*Yanyan Fu, Graduate Management Admission Council; Kyung (Chris) T. Han, GMAC*

### **Streamlining Test Pool Production Using Honeycomb Framework**

*Hwanggyu Lim, GMAC; Kyung (Chris) T. Han, GMAC*

Discussant:

*Kirk Becker, Pearson*

### 097. Approaches to Evaluating Item Quality

#### Coordinated Paper Session

3:05 to 4:35 pm

Convention Center: Level 1 - 121B

The old aphorism, “If you want to know what a test measures, look at the items,” rings true because it is impossible to have a high-quality test without high-quality items. This session presents the results of three efforts to examine item quality. The first presentation discusses a project designed to assess item quality outside the standard content and bias reviews typically conducted by vendors. The evaluation criteria, which are publicly available, can be applied to a broad array of assessment types and designs including state summative, interim, and curriculum-embedded assessments. The second presentation describes a tool and process designed to support an independent evaluation of commercial interim assessments. The process includes the review of operational test items by teams of educators against a detailed set of item quality expectations. Finally, we discuss a protocol developed to help districts make informed assessment procurement decisions. In this case, item quality is evaluated through the review of vendor documentation supporting claims of item quality and appropriateness. In sharing these efforts, we hope to motivate more extensive discussions about the quality of items we offer to students to support inferences about student learning.

Session Organizer:

*Erika Landl, Center for Assessment*

Participants:

#### Extending the Evaluation of Item Quality in New Hampshire

*Erika Landl, Center for Assessment; Brian Gong, Center for Assessment*

#### Evaluating the Implementation of New Hampshire’s Item Quality Criteria

*Rebecca Kelley, University of New Hampshire; Erika Landl, Center for Assessment; Scott Marion, National Center for the Improvement of Educational Assessment*

#### Piloting the District Assessment Procurement Protocol

*Susan Lyons, Lyons Assessment Consulting; Erika Landl, Center for Assessment*

#### Evaluating Item Quality within the Interim Assessment Evaluation Toolkit

*Michael Briscoe, Ed Reports*

Discussants:

*Kristen Huff, Curriculum Associates*

*Kristen Crawford, New Hampshire Department of Education*

### 098. Automatic Item Generation

#### Paper Session

3:05 to 4:35 pm

Convention Center: Level 1 - 121C

Chair:

*Dandan Chen, Pearson*

Participants:

#### Leveraging Pre-Trained Large Language Models for Automatic Item Generation: A Scoping Review

*Bin Tan, University of Alberta; Nour Armoush, University of Alberta; Okan Bulut, University of Alberta; Elisabetta Mazzullo, University of Alberta; Mark Gierl, University of Alberta*

This scoping review examined 60 research studies that utilized pre-trained language models (PLMs) and large language models (LLMs) for automatic item generation (AIG). We summarized popular models, their specific usage in AIG, model performance, and the characteristics of items generated, underscoring the significant potential of PLMs and LLMs in AIG.

### **Producing Item Sets in Automatic Item Generation**

*Hollis Lai, University of Alberta; Mark J Gierl, University of Alberta; Veronica Vele, Australian Dental Council*

Our paper demonstrates a new method for generating item sets--a set of items related to a common clinical scenario--through the existing template-based item generation approach. We provide an overview of the method, the modifications required, and a demonstration of the method in the content area of dentistry.

### **Transforming Item Generation in Modern Assessment Systems with Generative AI**

*Yunting Liu, UC Berkeley; Shreya Bhandari, University of California, Berkeley; Zachary A. Pardos, University of California, Berkeley*

This study examines how generative AI can aid the process of item generation. Items generated by prompting the ChatGPT 4.0 model are compared with items generated by humans in terms of costs and psychometric properties. Meticulous psychometric linking design is applied to ensure the comparability of the item parameters.

### **Predicting Expert Judgment on the Quality of Automatically Generated Items**

*Guher Gorgun, University of Alberta; Ikkyu Choi, ETS; Jiyun Zu, ETS*

The goal of this study is to predict expert judgment of the quality of automatically generated items. We employed two approaches: a feature-engineering approach using random forest and gradient boosting classifiers and a representation learning approach using BERT. The best performing models were achieved with the feature engineering approach.

Discussant:

*Tony Albano, University of California Davis*

## **099. Validity and Test Use**

### **Paper Session**

**3:05 to 4:35 pm**

**Convention Center: Level 1 - 122A**

Chair:

*Anne Traynor, Purdue University*

Participants:

### **Appropriate Use of Test Scores from A Large-Scale Assessment**

*Salih Binici, Florida Department of Education; Yachen Luo, Human Resources Research Organization*

This study examines consequences of model misfit and measurement error on reporting outcomes for a large-scale assessment. It investigates whether ignoring model misspecification and measurement error has any practical impact on reported scale scores for parents and teachers, also their secondary use in statistical analyses to inform policy makers.

### **College Grading Trends Pre- and Post Covid**

*Paul Westrick, College Board; Emily Shaw, College Board; Emily Angehr, College Board; Jessica Marini, College Board*

Utilizing course grade data from 47 four-year higher education institutions, this study highlights the increase in first-year grade point average (FYGPA) and domain-specific college GPAs between the 2017-2018 and 2021-2022 academic years. High school GPA also rose, while SAT scores remained relatively unchanged. Implications for validity research are discussed.

### **Examining External Validity through a Cultural Lens**

*Kerry Englert, Seneca Consulting, LLC; Pohai Kukea Shultz, University of Hawaii at Manoa*

Developing culturally appropriate assessments calls for a process that integrates cultural and community input. This presentation highlights how language, culture, and worldview play integral roles in development. We will provide examples of how validity studies must evolve to integrate community perspectives that simultaneously build technical quality.

### Understanding Parents' Interpretation and Use of Individual Student Score Reports

*Brian Gane, Accessible Teaching, Learning, and Assessment Systems ATLAS University of Kansas; Amy Clark, ATLAS: University of Kansas; Deborah Adkins, Accessible Teaching, Learning, and Assessment Systems ATLAS*

Score report feedback provides important consequential evidence. We collected and analyzed parent feedback on individual student score report interpretation and use. Feedback revealed successes and challenges around effective interpretation and use of the current score report design and suggested features for revision or redesign.

Discussant:

*Michael C. Rodriguez, University of Minnesota*

### 100. Strengthening Partnerships to Improve Classroom Assessment Research, Practice, and Policy (Joint AERA/NCME Session)

**Organized Discussion**

**3:05 to 4:35 pm**

**Convention Center: Level 1 - 122B**

According to Mark Wilson's 2016 remarks, educational measurement's core activity is to help in the educational progress of each student as they learn; the topic of classroom assessment is indeed crucial for the entire educational enterprise and should be seen as the most likely pathway for educational measurement to make a positive and central contribution to education (<https://www.ncme.org/community/ncme-committees/classroom-assessment/classroom-assessment129>). Since these remarks, a taskforce was formed to elevate the position of classroom assessment within NCME. Now a committee, we seek to engage all members and conference attendees in a organized discussion about Committee initiatives and goals in order to gather feedback on alignment with NCME's broader mission and vision.

Session Organizer:

*Jade Caines Lee, University of Kansas*

Moderator:

*Dustin Van Orman, Western Washington University*

Presenters:

*James McMillan, Virginia Commonwealth University  
Derek Christian Briggs, University of Colorado Boulder  
Michele Carney, Boise State University  
Yoav Bergner, New York University  
Bryan Drost, Rocky River City School District*

### 101. Pipeline Trends in Certification and Licensure: Psychometric Implications

**Organized Discussion**

**3:05 to 4:35 pm**

**Convention Center: Level 1 - 123**

Like trends observed in higher education with more institutions moving to test-optional policies, licensing and certification programs, are seeing shifts in their student and/or candidate pipeline. These changes may be due to declining college enrollments or generational shifts in student sentiment, learning preferences, and priorities, deregulation and anti-credentialing (e.g., licensure, certification) movements, Covid-19, and economic issues such as increased costs, to name a few. These shifts are not necessarily (or even likely) to be driven by psychometrics-focused issues, but there may be some interaction – and the nature of the work of psychometricians on these exams will surely be impacted

over time. This session highlights similarities and differences among three licensure and certification programs in distinct sectors – financial services, healthcare, and legal services. Presenters will summarize the trends observed in their programs’ exam taking and performance. They will respond to questions about perceived drivers of these changes, the shorter and longer-term impacts to the respective professions, and what have been the actions and reactions to respond to these trends. The session will conclude with a discussant synthesizing how trends in these programs are being interpreted in the broader credentialing community and the implications for psychometric training programs and practitioners.

Session Organizer:  
*Joshua Stopek, AICPA*

Moderator:  
*Chad W. Buckendahl, ACS Ventures, LLC*

Presenters:  
*Matthew Schultz, AICPA*  
*Rosemary Reshetar, National Conference of Bar Examiners*  
*Mengyao Zhang, National Conference of Bar Examiners*  
*Drew Weiner, National Conference of Bar Examiners*  
*Joshua Goodman, National Commission on Certification of Physician Assistants (NCCPA)*

### **102. Rooted in Community: Equitable Design and Practice in Assessment Organized Discussion 3:05 to 4:35 pm Convention Center: Level 1 - 124**

The education field needs examples of what equitable assessment systems look like in practice. Assessment and practice are closely intertwined. Data is a powerful tool for informing instruction and policy. Ideally, what we value in practice should drive what we measure, however, too often decisions about what to measure and how to measure it are made apart from practice. The field lacks knowledge and proofs of concepts about how the assessment process can: reflect the values, priorities, perspectives and goals of local communities, particularly Black, Brown, Indigenous and other People of Color communities (BBIPOC); provide systems data that go beyond individual student outcomes data and towards a more holistic, contextualized picture of schools and the student, family and community experience. Changing these systems requires new processes, approaches and tools. This session will share findings from a landscape scan of district appetite for equitable practice in assessment and enabling conditions identified by district leaders. Panelists will share tools and case examples of leading efforts across the country to advance equitable assessment practices in measuring academic progress, school learning conditions and social and emotional learning.

Session Organizer:  
*Ila Deshmukh Towery, Education First*

Presenters:  
*Christina Cipriano, Education Collaboratory at Yale*  
*Sasha Rabkin, Equal Opportunity Schools*

Discussant:  
*Laura Hamilton, American Institutes for Research*

### 103. Multistage Testing

#### Paper Session

3:05 to 4:35 pm

Convention Center: Level 1 - 125

Chair:

*Kristin M. Morrison, Curriculum Associates*

Participants:

#### Explore Feasibility Of Using Multiple Imputation For Dimensionality Analyses In Multistage Testing

*Danqi Zhu, Fordham University; Louis Roussos, Cognia; Han Yi Kim, Cognia; Sandra Sweeney, Cognia; Qi Qin, Cognia*

This study examines the feasibility of using multiple imputation by chained equations (MICE) for dimensionality analyses in multistage test. We compare four MICE based methods and evaluate using DIMTEST and DETECT indexes. Results show that MICE with random forest has highest accuracy to recover the true factor structure.

#### Mitigating Preknowledge Bias in Ability Estimates in Multistage Testing

*Merve Sarac, College Board; James Wollack, University of Wisconsin*

We investigated a mitigation strategy in a multistage test (MST) to reduce score bias when item preknowledge occurs, which downweights observations on potentially compromised items during interim ability estimation to route examinees to modules of appropriate difficulty in successive stages. Results showed that preknowledge-weighting produced lower bias in ability estimates.

#### The Effect of Theta Distribution on Multistage Simulations

*Elizabeth Ayers-Wright, Cambium Assessment, Inc.; Tao Jiang, Cambium Assessment, Inc.; Christina Schneider, Cambium Assessment, Inc.; Hongwook Suh, Cambium Assessment, Inc.*

Increasingly, states are considering adaptive through-year assessments. Previous research has examined the effect of theta initialization on simulations when assuming an underlying Normal distribution of student ability. However, observed ability distributions are often non-Normal. This paper extends previous research to look at the effect of ability distribution on multistage simulations.

#### Vertical Scaling in Multistage Adaptive Language Assessment

*Lee Bishop, WIDA at University of Wisconsin Madison; Hacer Karamese, WIDA at University of Wisconsin; Xin (Grace) Li, University of Wisconsin-Madison; Edynn Sato, Sato Education Consulting LLC*

This proposal outlines a study aimed at exploring various psychometric methods to establish vertical scaling within a multistage adaptive system (MST) for a large-scale language assessment.

Discussant:

*Jonathan Weeks, ETS*

### 104. One Year After GPT-4: How AI Changed Assessments?

#### Organized Discussion

3:05 to 4:35 pm

Convention Center: Level 4 - Terrace Ballroom III

At NCME 2023, just a few weeks after the GPT-4 was released, we organized a panel discussion around the potential implications of GPT-4 and other advanced AI tools in educational assessment. With a year having elapsed since the last well-attended session, we propose a new panel discussion to reflect on how the landscape of assessments has been shaped and elevated by the transformative influence of powerful AIs ever since. Our discussion will center around the five key areas of assessment, skills and constructs, item and task development, psychometric methodologies, reimagining assessment administration, and upskilling of psychometric practitioners. We hope this panel discussion can catalyze the exchange of ideas and insights, steering us collectively toward a more cohesive understanding of the transformative potential of AI in assessment.

Session Organizer:  
*Jiangang Hao, ETS*

Moderator:  
*Alina A. von Davier, Duolingo*

Presenters:  
*Victoria Yaneva, NBME*  
*Susan Lottridge, Cambium Assessment*  
*Matthias von Davier, Boston College*  
*Deborah Harris, University of Iowa*

### **105. Electronic Board Session III** **Electronic Board Session** **3:05 to 4:35 pm** **Convention Center: Level 4 - Terrace Ballroom IV**

Participants:

#### **1. ddSimilar: A Joint Statistical Cluster Analysis Of Test Collusion**

*Luyao Peng, Awada Tech LLC; Sandip Sinharay, ETS*

To identify test collusion, previous studies are either model-based or machine learning-based. Model-based methods require model assumptions, and all those methods may not work well when the test data are small. We proposed ddSimilar, which uses a Bayesian non-parametric prior to enhance the similarity index to detect group-level test collusions.

#### **2. Dynamic Learning Models Within Cognitive Diagnosis Framework to Quantify Heterogeneous Learning Effectiveness**

*Yanyan Tan, Pearson; Shiyu Wang, University of Georgia*

This study proposed two dynamic learning models to quantify heterogeneous contributions from different learning materials. We also explored the appropriate learning design considering the effect of different factors on proposed models' performance. The preliminary results indicate the compressed model fits better and materials measuring a single skill are more effective.

#### **3. Incorporating CUSUM Procedure with Machine Learning to Detect Proxy Testing**

*Guangyun Liu, University of Iowa; Xuechun Zhou, Ascend Learning*

This study aims to detect proxy testers in repeated testing using machine learning techniques incorporating a statistical process control procedure, cumulative sum (CUSUM) with weighted residuals between observed and expected item scores. Results show that SVM and NN with CUSUM incorporated have a higher sensitivity and lower false positive rate.

#### **4. Integrating Multiple Perspectives in Validating Measures of Culturally Responsive Schooling**

*Lisa Keller, UMASS Amherst; Stella Kim, University of North Carolina at Charlotte; Jonathan Supovitz, University of Pennsylvania; Constance Lindsay, University of North Carolina at Chapel Hill*

This paper investigates validating measures of culturally responsive schooling using data collected from multiple stakeholder perspectives. A discussion of how to integrate the various perspectives, along with the data themselves is provided. Data come from over 2000 stakeholders across 120 schools. The results indicate that multiple perspectives are needed.

#### **5. Investigating Test-Taking Behaviors of Multilingual Learners in PISA 2018**

*Jung Yeon Park, George Mason University; Zikun Li, George Mason University; Sean Joo, University of Kansas; Hyejin Yoon, Michigan State University*

This study investigates links between multilingual learners' test-taking behaviors and performance in math, science, and reading tests to uncover unique learning needs of different home language groups. Specifically, PISA 2018 digital data was employed to conduct measurement invariance, differential item functioning, and response time functioning.

### 6. Item Exposure Control with Reverse Cumulative Sum of Exposure Counts in CAT

*Jyun-Hong Chen, National Cheng Kung University; Hsiu-Yi Chao, Soochow University*

To maintain measurement precision while controlling item exposure, we propose the Seq-SH procedure to emphasize exposure control early on and prioritize ability estimation later in CAT. The results indicated that Seq-SH can efficiently improve trait estimation by dynamically adjusting exposure control parameters and facilitating the use of popular items.

### 7. The Effect of Estimation Methods on Bifactor Models with Ordinal Responses

*Xinchang Zhou, University of Illinois at Urbana-Champaign; Bo Zhang, University of Illinois at Urbana-Champaign; Yan Xia, University of Illinois at Urbana-Champaign*

This simulation study aims to systematically examine the statistical performance of estimators in bifactor models when ordered categorical responses are present. Results show that WLSMV outperformed ULSMV and MLR in statistical power, but MLR resulted in higher convergence rates and smaller bias. Keywords: Bifactor model, WLSMV, ULSMV, MLR, ordinal data

### 8. The Effect of Patient Gender on ChatGPT's Success When Answering Medical MCQs

*Peter Baldwin, National Board of Medical Examiners; Nina James Braum-Bharti, Germantown Friends School; Miles Somay Baldwin, Harriton*

One concern about AI systems is their potential to reproduce biases present in their training data. In this paper, we investigate one possible manifestation of such bias by asking ChatGPT to answer gender-neutral medical MCQs rewritten to include patient gender. Results suggest that gender affects ChatGPT's success on these items.

### 9. Using Displacement for Anchor Item Stability Check in Rasch Model Linking

*Youngmi Cho, Riverside Assessments; Hsin-Ro Wei, Riverside Insights; Hyeon-Joo Oh, Riverside Insights*

Simulation and real data studies are used to investigate the appropriateness of using the displacement values provided in the Winsteps for anchor item stability check. The results are compared against four scale transformation methods under different conditions of the length of anchor item set and item response missingness.

### 10. Advancing Graduate-Level Statistics Education: A Computerized Adaptive E-Learning Platform

*Yu Wang, University of Minnesota - Twin Cities; Chia-Yi Chiu, University of Minnesota*

This project crafted an e-learning platform for education graduate students. Integrating cognitive diagnostic models and the nonparametric item selection method, this platform can tailor learning materials for students. Preliminary survey results show students hold positive attitudes toward the platform, and the feedback provided by it is informative.

### 11. An Application of Bi-factor MIRT Model to MST with Intersectional Routing

*Hyesung Shin; Guemin Lee, Yonsei University*

This study investigates the effectiveness of intersectional routing (ISR), with bi-factor MIRT model for mixed-format MST composed of MC and FR items. Simulation results show that ISR holds potential as an effective approach for accurate ability estimation, especially when test length is short and correlation between sections are high.

### 12. Developing And Considering An Item Response Model Which Captures Inattention Fluctuations

*Yuki Shimizu; Hidetoki Ishii, Nagoya University*

This study proposed an item response model which eliminates the effects of inattentive responses in self-report surveys with anchoring vignettes. Simulation studies showed that the proposed model estimated item discrimination and high trait values better than existing models. A numerical example also delivered practical guidelines to design such questionnaires.

### 13. Developing Dual-objective CD-CAT Algorithms for College Gate-way STEM Courses

*Xiuxiu Tang, Purdue University; Yuxiao Zhang, Purdue University; Hua-Hua Chang, Purdue University*

The study proposes dual-objective cognitive diagnostic computerized adaptive testing (CD-CAT) algorithms for college gate-way STEM courses, which will use more information from the estimation of examinees' overall ability at the beginning of the test and switch the assessment focus to the estimation of examinees' skill mastery as the test proceed.

### 14. Developing Indicators of Code Quality For Block-Based Programming

*Tianying Feng, UCLA CRESST; Gregory Chung, UCLA CRESST; Charles Parks, UCLA CRESST*

We present an automated approach to analyzing code structures of block-based programs and developing indicators of code quality: complexity, content, and structure. Using sample responses for debugging tasks from the Scratch repository, we validate how our indicators reflect a priori categorizations of four debugging patterns and present preliminary findings.

### 15. Examining Test Preparation, Test-Taker Attitudes, and High-Stakes Outcomes

*Steven Nydick, Duolingo; Ramsey Cardwell, Duolingo; Jill Burstein, Duolingo*

High-stakes assessment outcomes may profoundly impact test takers' educational and professional goals. Test preparation supports item familiarity, but may also impact a test-taker's attitudes and assessment outcomes. Using an attitudes survey, we examined relationships between test-taker attitudes, their practice test use for a high-stakes language test, and high-stakes outcomes.

### 16. Efficiency of CAT with a Cognitively Designed Item

*Bank Hao Luo, Department of Educational Psychology, Faculty of Education, East China Normal University; Xiangdong Yang, Department of Educational Psychology, Faculty of Education, East China Normal University*

The efficiency of CAT was compared when using different models to calibrate item bank based on cognitive design system approach. Research has shown that the loss of CAT efficiency caused by the use of predicted difficulty can be compensated by improving the explanatory rate or adding several items.

### 17. Stratified AIG Item Selection for Form Assembly

*Marcus Walker, National Commission on Certification of Physician Assistants; Yanlin Jiang, National Commission on Certification of Physician Assistants; Andrew Dallas, National Commission on Certification of Physician Assistants (NCCPA)*

The study proposed a stratified item selection approach to sample a group of items with overall control on content distribution from AIG item pool. Sampled items have been evaluated from text and psychometric perspectives. The results show the selected items possessed reasonable text dissimilarity and item statistics convergence.

### 18. Investigating Method Effect of Wording Valence: A Latent Space IRT Approach

*Biao Zeng, Beijing Normal University; Minjeong Jeon, UCLA*

This study employed an LSIRT model to investigate method effects, revealing its ability to accurately estimate valence-induced method effects and their strength through interaction maps. Importantly, this approach goes beyond the established theoretical framework of Confirmatory Factor Analysis, enabling simultaneous consideration of hypothesized method effects and other potential confounding effects.

### 19. Maximum Marginal Likelihood Estimation of Multidimensional Generalized Partial Preference Model

*Jianbin Fu, ETS*

The multidimensional generalized partial preference model (MGPPM) is a newly proposed IRT model for forced-choice questionnaires. The current study develops a maximum marginal likelihood estimation with an expectation-maximization algorithm for the MGPPM. Simulated and real data will be used to study its parameter recovery and demonstrate its use in practice.

### 20. Research Trends in Process Data using Topic Modeling

*Hanna Choi, Ewha Womans University; Eunjeong Jeon, Ewha Womans University; Youn-Jeng Choi, Ewha Womans University*

This study identifies research trends in process data using text mining. We analyzed more than 600 journal and dissertation abstracts using the Topicmodels R package. Implications for the use of process data can be obtained through this study.

**106. Validity of Socioculturally Responsive and Culturally Sustaining Assessments: Issues and Practice****Coordinated Paper Session****4:55 to 6:25 pm****Convention Center: Level 1 - 120B**

Our diverse student population learns in a range of ways; however, assessments tend to require demonstrations of learning in relatively narrow, prescribed ways. There are implications for how students from different cultures engage with and apply their knowledge and skills to assessment tasks that either are not relevant or aligned with their culture. Additionally, expectations and criteria for evaluating performance on these tasks tend to privilege the culture of the test developer and may not acknowledge ways of knowing in other cultures. Culture is a critical consideration vis-a-vis students' ways of knowing, thinking, learning, and doing, and culture ought to be considered in assessment design and development. In this coordinated paper session, issues and practices related to the validity of socioculturally responsive and culturally sustaining assessments will be discussed. Presenters address the role of culture as foundational to assessment development and (re)evaluate considerations and evidence related to establishing assessment validity. Since assessment can shape and drive needed educational reforms and improve student achievement, the papers in this session are intended to be resources for indigenous communities, state departments of education, policymakers, test developers, and the research community to support such efforts and to promote social justice, fairness, and equity.

Session Organizer:

*Edynn Sato, Sato Education Consulting LLC*

Moderator:

*Edynn Sato, Sato Education Consulting LLC*

Participants:

**Integrating Cognitive and Psychometric Models to Produce Socioculturally Responsive Assessments***Edynn Sato, Sato Education Consulting LLC; Grace Li, WIDA at the University of Wisconsin – Madison;**Justin Kelly, Center for Applied Linguistics***Assessment of First Nations Languages: A Review of Literature***Lorena Alarcon, University of Illinois at Urbana-Champaign***Culturally Sustaining and Valid Alaska Native Language Assessment***Rosalie Grant, WIDA at the University of Wisconsin – Madison***Reconceptualizing Assessment for Equity and Social Justice: Native language Assessment in Hawaii***Pohai Kukea Shultz, University of Hawaii at Manoa; Kerry Englert, Seneca Consulting, LLC*

Discussant:

*Sally Samson, University of Alaska, Fairbanks***107. Findings from a Pilot Study of Classroom-based NGSS Aligned Prototype Assessments****Coordinated Paper Session****4:55 to 6:25 pm****Convention Center: Level 1 - 120C**

This coordinated paper session addressed the challenge of designing and developing prototype NGSS aligned assessments for use in elementary school science classrooms. The papers in this session describe various aspects of the Stackable, Instructionally-embedded, Portable Science project (SIPS), a federally funded collaborative effort to address simultaneously states' needs for large-scale science assessments and the needs of educators, parents, and students for resources that support science learning throughout the school year. Collectively, the five papers will: (a) provide an overview of the project's overall design logic for the integration of curriculum, instruction, and assessment at each of Grades 5 and 8; (b) illustrate the end-of-unit design approaches, and application of methods and the resulting instructional resources for content area units at each grade level; (c) provide a summary of the findings from a multi-state pilot study of the administration of the instructionally aligned end-of-unit

# FULL SCHEDULE

**SATURDAY, APRIL 13**

assessments; (d) describe the methodology used to establish cut scores for an integrated set of SIPS through-course science assessments in grades 5 and 8; and (e) summarize the work of the SIPS project and speak to NCME members who are interested in addressing states' need for quality, standards-aligned science assessments that generate meaningful, interpretable, and actionable results.

Session Organizer:

*Howard Everson, Graduate Center, City University of New York*

Participants:

**SIPS Project Overview**

*Ellen Forte, edCount, LLC*

**Design of Prototype End-of-Unit Assessment Tasks**

*Daisy Wise Rutstein, edCount, LLC*

**Overview of the Pilot Study Results**

*Howard Everson, Graduate Center, City University of New York*

**Establishing Science Learning Standards**

*Daniel Lewis, Creative Measurement Solutions LLC*

**Summary of Findings from the SIPS Project**

*James Pellegrino, University of Illinois at Chicago*

Discussant:

*Kristen Huff, Curriculum Associates*

## **108. Unveiling Narratives: Advancing Equity in Psychometric Testing Through Black Scholars' Perspectives**

**Coordinated Paper Session**

**4:55 to 6:25 pm**

**Convention Center: Level 1 - 121A**

This coordinated paper session illuminates the multifaceted landscape of educational measurement, emphasizing the importance of acknowledging the past while charting a course towards a more equitable future. Comprising four papers, this session embarks on a journey that spans the past, present, and future of bias in testing, guided by emerging Black scholars in psychometrics. The first paper provides a succinct historical overview of psychometrics, honing in on bias in testing, spanning from the 1890s to the present. The second paper offers an exploration of the groundbreaking work of Robert Lee Williams, highlighting his contributions and the challenges he faced. The third paper delves into contemporary practices of differential item functioning (DIF) analyses and their implications, illustrating how traditional conceptions of race and ethnicity continue to shape testing practices. Lastly, the session concludes by examining the narratives of present-day Black psychometricians, showcasing how they confront the weaponization of their identity in their daily work.

Session Organizer:

*Malitsitso Moteane, University of North Carolina at Greensboro*

Chair:

*Fiona Hinds, Independent Consultant*

Participants:

**Revisiting Psychometrics History: Black Scholars' New Perspective on Psychometric Development**

*Malitsitso Moteane, University of North Carolina at Greensboro; Teneka Steed; Janine Jackson, Morgan State University*

**An Ode to the Contributions of Robert Lee Williams to Educational Testing**

*Brianna Hooks-Singleton, UNCG; Malitsitso Moteane, University of North Carolina at Greensboro;*

*Henry Makinde, Student; Jaylin N. Nesbitt, James Madison University*

### **Critically Exploring Race and Ethnicity as Grouping Variables in DIF Studies**

*Malitsitso Moteane, University of North Carolina at Greensboro*

### **The Lived Experiences of Black Psychometricians in Academia and Industry**

*Janine Jackson, Morgan State University; Teneka Steed; Brianna Hooks-Singleton, University of North Carolina at Greensboro*

Discussants:

*Micheline Chalhoub-Deville, University of North Carolina, GRE*

*Damon Bryant*

## **109. Is That a Mode Difference? Demystifying Mode Effect Detection and Adjustment Methods**

### **Coordinated Paper Session**

**4:55 to 6:25 pm**

**Convention Center: Level 1 - 121B**

Mode differences are often examined through mode studies using experimental or quasi-experimental designs and, if needed, adjusted for using equating methodology. The purpose of this session is to invite researchers and practitioners into a thorough re-evaluation and discussion of commonly used mode effect detection methods and linking designs. The first two studies presented in this session evaluate the effectiveness of various statistical methods to obtain equivalent groups without random assignment, one for multiple-choice tests and one for a writing test. The investigations are based on manipulations of operational data from paper testing so that the true mode differences are known. The third study highlights advantages and pitfalls of all three linking designs through demonstrations using simulated data. The study is motivated by common misconceptions and misunderstandings and seeks to provide enhanced perspectives to researchers and practitioners. The fourth study deals with practical situations in which true mode effects are unknown. In this study, results obtained through chained linking with a random groups design and those from a direct linking through propensity score matching are compared. Insights and information gained from this session can inform practical decisions about mode effect detection methods and mode linking designs.

Session Organizer:

*Dongmei Li, ACT*

Chair:

*Jeffrey Steedle, Curriculum Associates*

Participants:

### **Finding Optimal Solutions for Mode Effect Detections Without Random Assignment**

*Scott Woods, ACT, Inc.; Shalini Kapoor, ACT, INC.; Jeffrey Steedle, Curriculum Associates*

### **Finding Optimal Solutions for Detecting and Adjusting Writing Mode Differences**

*YoungWoo Cho, ACT; Dongmei Li, ACT; Ann Arthur, ACT*

### **Showcasing Advantages and Pitfalls in Mode Detection and Adjustment Methods**

*Dongmei Li, ACT; Ann Arthur, ACT; Hongling Wang, ACT, INC.*

### **Is That a Real Mode Difference?**

*Shalini Kapoor, ACT, INC.; Chen Qiu, ACT, INC.; Dongmei Li, ACT*

Discussant:

*Deborah Harris, University of Iowa*

### 110. Item Response Theory

#### Paper Session

4:55 to 6:25 pm

Convention Center: Level 1 - 121C

Chair:

*Yi-Fang Wu, Cambium Assessment, Inc.*

Participants:

#### **A Fused SDT/IRT Model for Multiple-Choice and Open-Ended Items**

*Lawrence T. DeCarlo, Teachers College, Columbia University*

Exams often consist of different types of items, such as multiple-choice and open-ended items, which raises questions about which of a variety of item response theory (IRT) models should be used. The present study shows practical advantages of a choice model based on signal detection theory.

#### **Combining Teacher Observations With Test Data Using A Mixed Fixed/Random Psychometric Model**

*Mark Wilson, Berkeley School of Education, University of California, Berkeley; Perman Gochyyev, University of California, Berkeley*

We present findings from a multi-year study, where we merge teachers' in-class observations (treated as random items), with item responses to pre- and post-tests (treated as fixed items). For content/interpretation this is based on a construct maps across the two modes. For analysis we use the mixed fixed/random-item IRT model.

#### **Unipolar IRT and Measurement of Vocabulary Knowledge**

*Qi Huang, University of Wisconsin - Madison; Daniel Bolt, University of Wisconsin, Madison*

Unipolar IRT models (e.g., log normal distribution of proficiency, cumulative log-logistic measurement model) use positive latent trait metrics having an absolute zero to represent absence of the construct. We apply unipolar IRT to a vocabulary test and demonstrate how empirically arbitrary IRT modeling choices significantly affect quantification of item/test information.

Discussant:

*Stephen G Sireci, University of Massachusetts, Amherst*

### 111. Alternative Ways of Estimating Performance in Large Scale Assessments

#### Paper Session

4:55 to 6:25 pm

Convention Center: Level 1 - 122A

Chair:

*Jianshen Chen, College Board*

Participants:

#### **Estimating Student Achievement in Large-Scale Assessments: Can Latent Regression Substitute Plausible Values?**

*Danielle Siegel, University of California, Davis; Ting Zhang, American Institutes for Research; Paul Bailey, American Institutes for Research; Sinan Yavuz, American Institutes for Research*

Direct estimation (DE) is a latent regression approach alternative to the use of plausible values to analyze large-scale assessment data. Findings about how well DE performs are mixed. A comparison with simulated data suggests that DE is a suitable alternative to the plausible value approach.

#### **Exploring Ways To Facilitate Estimation of Mean Naep Achievement of Schools**

*David Bamat, American Institutes for Research; Markus Broer, American Institutes for Research*

Motivated by the prospect of facilitating school-centric instead of student-centric reporting with NAEP data, our study finds that a technique involving the use of school-level plausible values generates mean achievement estimates for groups of schools that approximate benchmark values of achievement estimated through multilevel regression modeling

### **Forgoing Invariance Assumptions for Better Measurement in Comparative International Education Survey Research**

*Anne Traynor, Purdue University; Ming-Min Cheng, Purdue University; Yimei Xiong, Purdue University; Rumeysa Gorgulu, Purdue University; Amirreza Mehrabi, Purdue University*

In this theory project, we argue that forgoing measurement invariance assumptions to focus on optimal representation of participant attributes (e.g., students' attitudes, educators' perceptions) by the survey items in a comparative international education study may produce better measurement within countries, and more trustworthy comparisons of explanatory model results across countries.

### **Improving School Aggregate Scores by Accounting for Missing Student Achievement Data**

*Scott Monroe, UMASS Amherst*

For ESSA reporting, states routinely exclude nonparticipating students when calculating school aggregates of student achievement scores. However, depending on the nature of the nonparticipation (e.g., student opt-out), the aggregates may be biased. This study proposes multiple imputation to account for missing student data and improve the accuracy of school aggregates.

Discussant:

*Matthias von Davier, Boston College*

## **112. Re-envisioning Fairness: Far-reaching Implications for Measurement and Assessment Research and Practice (Joint AERA/NCME Session)**

### **Organized Discussion**

**4:55 to 6:25 pm**

**Convention Center: Level 1 - 122B**

The educational measurement community is experiencing heightened concerns for fairness and educational assessment faces increased criticism of its role as a perpetuator of system inequities. The proposed organized discussion will address conference themes — reducing inequities and dismantling racial injustices — through examining the treatment of fairness in measurement and assessment with implications for future research and practice. The anticipated revision of the current edition of the Standards for Educational and Psychological Testing (AERA, APA, NCME, 2014) could not be more timely as the current Standards position fairness as one of three key foundations. Along with validity and reliability, fairness undergirds assessment design, development, validation, and use. This notion of fairness deals with technical concerns that are largely under the control of measurement professionals, whereas issues related to equality of opportunity and the distribution of resources and privilege in our society are currently excluded from the Standards. Five scholars, whose work has been deeply impactful in the areas of assessment and measurement (across large-scale and classroom assessment) will focus on the meaning of fairness in contemporary measurement and assessment, address the conceptual and methodological challenges in adopting new approaches to fairness, and consider implications for research and practice.

Session Organizer:

*Alison Bailey, UCLA*

Moderator:

*Jose Felipe Martinez, UCLA - School of Education and Information Studies*

Presenters:

*Jade Caines Lee, University of Kansas*

*Guillermo Solano-Flores, Stanford University*

*Rochelle Michel, Smarter Balanced*

*Randy Bennett, ETS*

*Ye Tong, NBME*

### **113. Next-Gen Assessments: Probing Higher Cognitive Dimensions in the Common Core Era** **Organized Discussion** **4:55 to 6:25 pm** **Convention Center: Level 1 - 123**

Modern education is rapidly evolving, necessitating the move towards assessing higher cognitive dimensions congruent with Common Core Standards. While these standards typically resonate with elevated cognition levels, past assessment methods largely remained at rudimentary taxonomy levels. As technology advances, so does our capability to creatively gauge quintessential skills. This discussion aims to explore how the intersection of technology and evolving educational paradigms can foster the development of assessment tools that genuinely reflect the depth and breadth of today's educational objectives, aligning seamlessly with NAAD's commitment to equity, diversity, and best practices in measurement.

Session Organizer:  
*Matt Raimondi, School District U-46*

Presenters:  
*Arthur Thacker, HumRRO*  
*Patrick Meyer, NWEA*  
*Luis Saldivia, ETS*

### **114. Writing Assessments and Rater Effects** **Paper Session** **4:55 to 6:25 pm** **Convention Center: Level 1 - 124**

Chair:  
*Ye Yuan, University of Georgia*

Participants:

#### **Questioning the Copying of Text: Validity of Lower Elementary Source-Based Writing Rubric**

*Laura Boynton Hauerwas, Providence College; Elizabeth L. Adams, American Institutes for Research*

The paper explores the validity of a new source-based writing rubric for lower elementary students that was developed in a larger study. The analysis of survey and focus group data provides information about content and response process validity of the assessment and explores how raters address text copying developmentally.

#### **Automated Rater Routing Algorithm for Essays using Unfolding Scale and Topic Models**

*Jordan M. Wheeler, University of Georgia; George Engelhard, UGA; Jue Wang, University of Science and Technology of China*

This study proposes and demonstrates a new algorithm that predicts the location of essays on an unfolding scale. The predicted unfolding location of an essay is then used to assign raters that are more likely to accurately score the essay. This framework increases the validity and fairness of human-scored essays.

#### **Rater Effects vs Robo-Effects: Profiling Automated Classroom Evaluation**

*Michael Hardy, Stanford University*

Could a microphone improve the reliability of classroom observation ratings? This paper employs both Generalizability Theory and IRT rater models to contrast human rater effects with the effects observed in automated ratings produced by custom NLP models trained on classroom transcripts to investigate how they might support human observers.

### **Pursuing Fair Writing Assessment: Halo Effects in Primary School EFL Writing**

*Ruth Trüb, FHNW School of Education; Julian Lohmann, Kiel University; Jens Möller, Kiel University; Stefan Daniel Keller, Zurich University of Teacher Education*

Using a corpus of young EFL learners' texts, we analyzed which aspects of text quality influence pre-service teachers' assessment of other criteria (halo effects). Orthography, vocabulary, and punctuation revealed significant halo effects on various criteria. Communicative effect, complexity and correctness of syntax and grammar are determined by other text qualities.

Discussant:

*Paul Deane, ETS*

## **115. Test Security and Test Taking Behavior**

### **Paper Session**

**4:55 to 6:25 pm**

**Convention Center: Level 1 - 125**

Chair:

*Kylie Gorney, Michigan State University*

Participants:

### **An Experimental Manipulation of Test Consequences on Test-Taking Behaviors**

*Michalis Michaelides, University of Cyprus; Despoina Chatzidimitriou, University of Cyprus; Myria Chatzimichail, University of Cyprus*

Examinee test-taking behaviors may differ depending on the stakes of an assessment, and thus compromise score validity. The study examined self-reported test-taking effort, importance and anxiety in a survey experiment in Cyprus. Students responded differently on anxiety, but not on effort or importance perceptions, across high-, low-, and no-stakes conditions.

### **Detecting Illicit Collaboration Using Machine Learning**

*Danqi Zhu, Fordham University; Joseph Grochowalski, College Board; Amy H. Hendrickson, College Board*

This study proposes a framework that combines IRT modeling and machine learning for detecting illicit group collaboration among test takers. In preliminary empirical and simulation studies, this approach achieved high detection rate, which shows promise for efficient pre-screening in educational assessments.

### **Spectral Clustering of Mixed-Type Data for Collusion Detection**

*Soo Ingrisone, Pearson; James Ingrisone, Pearson VUE*

The effectiveness of spectral clustering (SC) on mixed type-data for detecting collusion groups on Computer-Based Tests is explored. Real testing data from a certification exam are used. The level of overlap between the collusion cluster and exact response matches on incorrectly keyed items in a study guide are compared.

### **Test Security: A Multivariate Outlier Detection Study**

*Cecilia Alves, Medical Council of Canada; Qi Guo, Medical Council of Canada; Andrea Julie Gotzmann, Medical Council of Canada*

The COVID-19 pandemic disrupted traditional in-person testing models, leading to the rise of remote proctoring. This study utilized the isolation forest algorithm to analyze testing data from a medical licensure exam, aiming to identify both outlier items and candidates. Preliminary results indicate low percentages of anomaly items and candidates.

Discussant:

*James Wollack, University of Wisconsin*

## FULL SCHEDULE

SATURDAY, APRIL 13

### 116. Sparking a Debate on the Role of Artificial Intelligence in Educational Measurement

**Organized Discussion**

**4:55 to 6:25 pm**

**Convention Center: Level 4 - Terrace Ballroom III**

Artificial intelligence (AI) has proven a useful and powerful technology for applications across numerous disciplines and industries. The growth of AI tools and applications reflects the zeitgeist of our times—AI is increasingly deployed to gain efficiencies through automation and provide precise predictions. Accordingly, educational researchers, methodologists, and psychometricians are increasingly interested in the role of artificial intelligence (AI) in educational measurement. The practice of educational and psychological measurement is guided by principles and standards. The proliferation of AI has in many ways outpaced The Standards and the purpose of this symposium is to provide a forum for debate concerning the role of AI for educational measurement. Researchers at Duolingo recently were the first educational testing company to produce “Responsible AI Standards” (RAIS) for AI in educational assessment (Burstein, 2023). Through short presentations and moderated debate, several educational measurement leaders will attempt to achieve consensus or draw clear lines of disagreement about how or whether to use AI in educational measurement.

Session Organizer:

*Steven Culpepper, University of Illinois Urbana-Champaign*

Moderator:

*Brian Junker, Carnegie Mellon University*

Presenters:

*Alina A. von Davier, Duolingo*

*Andrew Ho, Harvard Graduate School of Education*

*Ezekiel Dixon-Roman, University of Pennsylvania*

*Hua-Hua Chang, Purdue University*

*Derek Christian Briggs, University of Colorado Boulder*

*Matthew S Johnson, ETS*

SUNDAY, APRIL 14

### 117. Recent Innovations in Automated Scoring

**Coordinated Paper Session**

**7:45 to 9:15 am**

**Convention Center: Level 1 - 120C**

This coordinated session illustrates five recent developments in automated scoring. Some of the developments are due to the incorporation of transformer models or their variants (Devlin et al., 2019; Vaswani et al., 2017), the refinement of topic identification strategies (Xiong et al., 2021), or technological breakthroughs (Saxton et al., 2019). The first paper shows the challenges and complications of trying to operationalize the scoring of math items. The second uses an mBERT transformer model to evaluate writing in Persian, but can easily be applied to a number of different languages. The third paper highlights how Latent Dirichlet Allocation has improved to help with identifying the topic of writing. The fourth explains the advances in the automated scoring of spoken responses. The last paper is an example of how a traditional operational scoring engine was improved by adopting variations of BERT transformer models. Taken as a whole, these developments highlight that inroads in one area of automated scoring often impact other areas in a positive way. Ultimately, these recent innovations in automated scoring contribute to the broader goal of more equitable and effective educational measurement.

Session Organizer:

*Mark D. Shermis, Performance Assessment Analytics, LLC*

Moderator:

*Joshua Wilson, University of Delaware*

Participants:

### **Automated Scoring of Math Constructed-Response Items**

*Scott Hellman, Pearson; Luis Alejandro Andrade-Lotero, Pearson; Kyle Habermehl, Pearson; Alicia Bouy, Pearson; Lee Becker, Pearson*

### **Scoring Essays Written in Persian Using a Transformer-Based Model: Implications for Multilingual AES**

*Tahereh Firoozi, University of Alberta; Mark J Gierl, University of Alberta*

### **Latent Dirichlet Allocation of Essays**

*Jordan Wheeler, University of Nebraska Lincoln; Shiyu Wang, University of Georgia; Allan Cohen, University of Georgia*

### **Automated Scoring and Feedback for Spoken Language**

*Klaus Zechner, ETS; Ching-Ni Hsieh, ETS*

### **Redesigning Automated Scoring Engines to include Transformer-based Deep Learning Models**

*Susan Lottridge, Cambium Assessment, Inc; Christopher Michael Ormerod, Cambium Assessment; Milan Patel, Cambium Assessment, Inc.*

Discussant:

*Mark D. Shermis, Performance Assessment Analytics, LLC*

## **118. Assessment and Instruction Coherence for Dynamic Measurement of Student Proficiency Coordinated Paper Session**

**7:45 to 9:15 am**

**Convention Center: Level 1 - 121A**

Interactions with digital lessons represent great opportunities for measurement; this data could be used to update reports, track student growth, and evaluate the effectiveness of instruction, all within the context of student learning. In this session, we will share research connecting data from a large-scale digital assessment to data from digital lessons. The first paper will examine the content alignment between assessment items and lesson quiz items in two domains and grade levels. The second and third papers explore the various context effects surrounding the lesson quiz items and seek to understand how these effects contribute to item performance. For the second paper, we administered lessons to students under a variety of conditions, some with instruction and some without, to examine the effect of instruction and feedback on estimation of IRT parameters. The third study attempts to estimate item parameters from lessons as they are administered naturally in students' lesson path and attempts to account for the effects of preceding instruction and restriction of range in the sample in estimating parameters. Finally, pulling together the previous papers, we apply a tracking algorithm (Glicko) to students' assessment and lesson data to track students' learning in mathematics throughout the academic year.

Session Organizer:

*Logan Rome, Curriculum Associates*

Participants:

### **Alignment of Digital Assessment and Instruction Items**

*Amanda Brice, Curriculum Associates*

### **The Effect of Instruction and Feedback on IRT Parameter Estimation**

*Erin Banjanovic, Curriculum Associates; Nathan Minchen, Curriculum Associates, LLC; Ted Daisher, Curriculum Associates; Logan Rome, Curriculum Associates*

### Item Parameter Estimation from Students' Lesson Paths

*Sanford R. Student, University of Delaware; Erin Banjanovic, Curriculum Associates; Logan Rome, Curriculum Associates*

### Tracking Student Learning Through Digital Instruction

*Ted Daisher, Curriculum Associates; Erin Banjanovic, Curriculum Associates; Matthew Myers, University of Delaware; Sanford R Student, University of Delaware; Logan Rome, Curriculum Associates*

Discussant:

*Ying Cheng, University of Notre Dame*

## 119. Social Emotional Learning

### Paper Session

**7:45 to 9:15 am**

**Convention Center: Level 1 - 121B**

Chair:

*Tuba Gezer, UNC Charlotte*

Participants:

### Classifying Students and Schools Based on Social-Emotional Learning Trajectories: An Exploratory Study

*Jessica Elizabeth Schnittka Hoskins, UCLA*

This study used mixture modeling to classify students and schools based on middle school social-emotional learning trajectories. Focusing on one construct (social awareness) five classes of students and three classes of schools emerged. In this presentation, I will discuss potential uses of these classifications for promoting equity in schools.

### Cultural Adaptation of a Social and Emotional Assessment for Spanish Speaking Students

*Jennifer Robitaille, Aperture Education; Emily Hope Parker, Aperture Education; Matthew Buczek, Aperture Education; Evelyn Johnson, Aperture Education*

Best practice suggests that we not just translate but adapt assessments into other languages, ensuring cultural appropriateness and validity. We present our approach to creating cultural adaptations of a social and emotional assessment and share results of applying this process to the Spanish language version of the student self-report assessment.

### Response Anchoring and Student Disengagement on Self-Report Measures of Social-Emotional Learning

*Daniel Bolt, University of Wisconsin, Madison; Yang Caroline Wang, Education Analytics*

Response anchoring, the tendency for respondents to provide item responses equivalent (or in close proximity to) immediately preceding responses, often reflects disengaged responding. Using Grade 4-12 student responses to social-emotional learning measures, we examine student and item covariates that predict anchoring, and demonstrate score bias and survey design implications.

### Integrating Transformative SEL into Formative Classroom Assessment

*Jerome De Lisle, University of the West Indies; Reina Braithwaite, The University of the West Indies; Nalini Ramsawak-Jodha, University of the West Indies, St. Augustine, Trinidad; Nicola Mark-Worrell; Kristy Phillip, The University of the West Indies; Nirmala Ramnarine, The University of the West Indies; Patricia Claxton*

In promoting SEL everyday in schools, a useful strategy might be to incorporate SEL into formative classroom assessment. Transformative SEL enhances student capability and reduces inequity. We describe and illustrate theory and evidence-based protocols for classroom teachers to directly target SEL competencies through classroom assessment informed by different SEL frameworks.

Discussant:

*Marianne Perie, WestEd*

### 120. Automated Scoring II

**Paper Session**

**7:45 to 9:15 am**

**Convention Center: Level 1 - 121C**

Chair:

*Merve Sarac, College Board*

Participants:

#### **Comparing Measurement Properties of GPT and Human Scoring in Student Constructed Responses**

*Eun Hye Ham, Kongju National University*

The study investigates whether the psychometric properties of GPT scoring are comparable to those of human scoring when scoring university students' constructed responses in scenario-based assessment. We explore various GPT prompting methods and compare the factor structures and Rasch models fitted to GPT-scored data with those of human-scored data.

#### **Fair or Flawed? Analyzing Biases in Automatic Text-Response Scoring**

*Nico Andersen, DIPF | Leibniz Institute for Research and Information in Education; Fabian Zehner, DIPF | Leibniz Institute for Research and Information in Education, Centre f. Int. Student Assessm.; Julia Mang, TUM School of Education, Centre for International Student Assessment (ZIB), Technical University of Munich (TUM); Frank Goldhammer, DIPF | Leibniz Institute for Research and Information in Education, ZIB*

To measure the fairness of automatic scoring, models were trained on text-responses from the PISA 2015 reading comprehension tasks and results were compared between students of different genders and students who speak a different language at home. Scoring models systematically overestimate the performance of non-native speakers, giving them an advantage.

#### **The Integration of Automated Essay Scoring Models Using Hierarchical Rater Models**

*Aron Fink, Goethe University Frankfurt; Sebastian Gombert, Leibniz Institute for Research and Information in Education; Tuo Liu, Goethe University Frankfurt, Germany; Andreas Frey, Goethe University Frankfurt*

We propose a model-based method to integrate predicted scores from multiple automated essay scoring models (AESMs) and investigate its performance on data from an essay writing test. The proposed method achieves higher accuracy and is able to establish a higher level of measurement invariance compared to the individual AESMs.

#### **Use of Data Augmentation in Automated Essay Scoring Training Data**

*Justin O Barber, Pearson; Edward W Wolfe, Pearson*

Automated Essay Scoring (AES) requires costly human scored responses to train an engine. Our study explores how data augmentation (DA) method and insertion rate impacts AES model accuracy. Our results suggest that DA methods may improve model performance with as few as 10 generated examples.

Discussant:

*Hongwen Guo, ETS*

### 121. Differential Item Functioning I

**Paper Session**

**7:45 to 9:15 am**

**Convention Center: Level 1 - 122A**

Participants:

#### **Application of the IRT Residual DIF Framework to Ordinally Scored Polytomous Items**

*Hwanggyu Lim, GMAC; Jaime Malatesta, GMAC*

This study extends the IRT residual DIF detection framework (RDIF) for ordinally scored polytomous items. Preliminary simulation results indicate that the RDIF framework offers effective control of Type

I error and sufficient power compared to other methods, making it a practical tool for evaluating DIF in polytomously scored items.

### **Choice of the RMSD Cutoff for Multigroup-DIF Detection**

*Dandan Chen, Pearson; Jinming Zhang, University of Illinois Urbana-Champaign*

The root mean square deviation (RMSD) approach is commonly used with an arbitrary cutoff. We investigated the efficacy of various RMSD cutoffs and proposed a model-based approach to select an appropriate cutoff for RMSD for multi-group DIF detection. Researchers can use our proposed model to choose the right RMSD cutoff.

### **Comparison of CRUM and 2PL-CMIRT Models in the Context of DIF**

*Ahmet Guven, Augusta University; Yanyun Yang, Florida State University*

The two-parameter logistic compensatory MIRT model (2PL-CMIRT) and compensatory reparametrized unified model (CRUM) have a lot in common. The purpose of this study was to compare the performance of the Wald and likelihood ratio tests in both the 2PL-CMIRT model and the CRUM in detecting DIF when data were refitted.

### **Using a Rasch Differential Item Functioning Approach to Investigate Early Childhood Assessments**

*Toshiko Kamei, University of Melbourne; Nafisa Awwal, University of Melbourne; Lan Anh Nguyen Khoa, University of Melbourne; Jane Page, University of Melbourne*

This paper establishes the generalisability validity of eight assessments based on learning progressions for teachers of children aged 2 -6 in Victorian kindergarten programs. Through a Rasch differential functioning analysis (DIF), the study showed their applicability for all children, including those with a diagnosis of autism spectrum disorder (ASD).

Discussant:

*Jinghua Liu, National Board of Osteopathic Medical Examiners*

## **122. Large Scale Assessment**

### **Paper Session**

**7:45 to 9:15 am**

**Convention Center: Level 1 - 122B**

Chair:

*Tiago A Caliço, Organization for Economic Cooperation and Development*

Participants:

### **A New Strategy for TIMSS Quality Assurance**

*Henry Braun, Boston College; Matthias von Davier, Boston College; Jihang Chen, Boston College*

International large-scale assessments have become higher stakes for governments' education policies. This has led to an increased incentive to deviate from assessment guidelines and best practices in order to increase scores. This report describes strategies to identify unusual country-level results based on the estimated relationships between outcomes and background/contextual factors.

### **How Can PIRLS Take Advantage of Adaptive Testing?**

*Montserrat B. Valdivia Medinaceli, University of California Davis; Leslie Rutkowski, Indiana University*

This paper investigates the efficacy of Group Adaptive Testing (GAT) in international large-scale assessments (ILSAs), specifically the Progress in International Reading Literacy Study (PIRLS). GAT aims to enhance measurement precision at the ends of the proficiency continuum. This paper studies the challenges of the item pool alignment with proficiency distribution.

### **Linking NAEP and TIMSS Scores to Compare US States with TIMSS Countries**

*Paul Bailey, American Institutes for Research; Ting Zhang, American Institutes for Research; Huade Huo, AIR; Charles Blankenship, American Institutes for Research; Blue Webb, American Institutes for Research*

The study linked NAEP and TIMSS scale scores to compare US states with TIMSS countries. Building upon existing linking techniques, we enhanced the techniques by employing a Taylor series approach. This refinement enables us to make comparisons between US states and TIMSS countries, rendering the analyses feasible and replicable.

**123. Developing Curriculum-Agnostic, Instructionally Relevant Assessments****Coordinated Paper Session****7:45 to 9:15 am****Convention Center: Level 1 - 123**

Ongoing efforts to incorporate assessment systems in classrooms rely on assessments providing evidence of students' current understandings, skills, or reasoning patterns in actionable ways (Pellegrino, diBello, & Goldman, 2016). Such assessments must satisfy the need for instructional relevance through proximity to the learning environment while being broadly applicable and supported by sophisticated validation approaches often reserved for distal assessments (Ruiz-Primo et al., 2012). In this session, we bring five papers that illustrate evidence-centered approaches to designing instructionally relevant computer-based assessments in key domains of middle school math and science. We will conclude with comments from discussant Jim Pellegrino.

Session Organizer:

*Helena Connolly, Houghton Mifflin Harcourt*

Participants:

**Using Cognitive Models to Develop Instructionally Relevant Assessments***Helena Connolly, Houghton Mifflin Harcourt; Gavin Fulmer, HMH; Shiv Karunakaran, HMH; Susan Kowalski, HMH; Tyler Matta, NWEA***Developing an Assessment of Student Reasoning About Ratio***Helena Connolly, Houghton Mifflin Harcourt; Shiv Karunakaran, HMH; Colleen Oppenzato, HMH; Tyler Matta, NWEA***Developing an Assessment of Student Reasoning About Chemical Reactions***Katherine Lazenby, HMH; Gavin Fulmer, HMH; Susan Kowalski, HMH***Developing an Assessment of Student Understanding About Fraction Division***Colleen Oppenzato, HMH; Helena Connolly, Houghton Mifflin Harcourt; Erik Ruzek, HMH; Tyler Matta, NWEA***Putting the Cognitive in Cognitive Diagnostic Models***Daniel Katz, NWEA; Yon Soo Suh, NWEA; Meredith Langi, WestEd; Tyler Matta, NWEA*

Discussant:

*James Pellegrino, University of Illinois at Chicago***124. Student Growth****Paper Session****7:45 to 9:15 am****Convention Center: Level 1 - 124**

Chair:

*Paulius Satkus, GMAC*

Participants:

**Too Good to Be True: MAP Growth Reportage in Three Urban Districts***Paul Zavitkovsky, Center for Urban Education Leadership, University of Illinois at Chicago*

This ongoing study compares ten years of same-student data from NWEA and state testing in three urban districts. Preliminary z-score analysis reveals large and persistent anomalies between many comparable NWEA and state test distributions. This finding casts doubt on the claim that NWEA reportage consistently reflects statewide learning standards.

**Modeling Students' Growth in an Adaptive Learning System Using Deep-IRT***Tongtong Zou, Columbia University; Jinah Choi, Edmentum, Inc.; Sonya Powers, Edmentum, Inc.*

In this study, the Deep-IRT model is fitted to quiz items in an adaptive learning system. Preliminary

results showed good model prediction accuracy. The next step is to integrate latent knowledge state estimates for the prediction of students' growth on future assessments and to compare with traditional IRT methods' outputs.

### **Aggregating Interim Assessment Scores for Summative Decisions: Exploring Longitudinally-Weighted Proficiency Estimators**

*Aaron Myers, ABIM*

Competency-based assessment programs consisting of multiple, shorter, interim assessments rather than a longer summative assessment have garnered recent attention. Little consideration has been given to how interim scores are aggregated to produce summative scores. We propose and evaluate four IRT-based longitudinally-weighted proficiency estimators. Implications of weighted scoring are discussed.

### **A Comparison of Through-Year Cumulative Scoring Methods**

*Garron Gianopulos, Cambium Assessment; Chris Rozunick, TEA*

We compare eight methods for producing cumulative scores from a through-year assessment. We examine whether a cumulative score might perform as well as a single EOY summative score, both in terms of measuring short-term and long-term learning. We discuss the trade-offs of cumulative scores.

Discussant:

*Katherine Furgol Castellano, ETS*

## **125. Applications of AI II**

### **Paper Session**

**7:45 to 9:15 am**

**Convention Center: Level 1 - 125**

Chair:

*Jeongwon Choi, Vanderbilt University*

Participants:

### **Artificial Intelligence Approach to Classifying Oral Reading Fluency Performances**

*Gozde Sirganci, Southern Methodist University; Sarunya Somsong, Srinakharinwirot University & Southern Methodist University; Akihito Kamata, Southern Methodist University*

This study aims to identify meaningful classes for individuals with similar oral reading fluency scores by combining clustering and neural network methods. We analyze an empirical dataset to demonstrate how this approach can provide a clearer picture of the student's ORF level by understanding the characteristics of their group.

### **Effect of Administering Isomorphs on Ability Estimates in Multistage Testing**

*Hyung Jin Kim, University of Iowa; Won-Chan Lee, University of Iowa*

As computer-based testing becomes more prevalent, the need to have more available items has increased, leading to greater interest in automatic item generation. The study considers four approaches for administering and scoring isomorphs, evaluating their performance in terms of the accuracy in ability estimates.

### **Enhancing Automated Short-Answer Grading: The Impact of a Robust Scoring Rubric**

*Wallace Nascimento Pinto, University of Florida; Lodi Lipien, Florida Virtual School; April Fleetwood, Florida Virtual School; Jinnie Shin, University of Florida*

We examined the effect of a robust scoring rubric to mitigate scoring inconsistency on an automated short-answer grading system. Human scoring inconsistencies may have adverse impact on performance, reliability, and validity. We refined a Transformer-based model using the new scores generated from the robust rubric to improve the scoring performance.

### **Examining ChatGPT's Influence on Assisting Students with L2 Writing Tasks**

*Dan Song, University of Iowa; Alexander Tang, University of Hawaii at Mānoa*

This study reconceptualizes educational inequities by examining AI's role, notably ChatGPT, in second language learning. It explores its effect on L2 writing in a Mandarin context, comparing students' performances on AP Chinese writing tasks. The research focuses on differential outcomes and potential improvements with AI assistance in learning.

Discussant:

*Hollis Lai, University of Alberta*

### **126. Exploring Validity of NAEP: Multiple Approaches and Perspectives**

#### **Coordinated Paper Session**

**7:45 to 9:15 am**

**Convention Center: Level 4 - Terrace Ballroom III**

As the largest nationally representative, continuing evaluation of the condition of education in the United States, the primary purpose of the National Assessment of Educational Progress (NAEP) is to measure the educational achievement and progress of the nation's students at established grades and ages. To achieve this purpose, ensuring validity in NAEP is imperative. This coordinated session presents 4 recent research studies to provide validity evidence for NAEP. The first study explores potential differential functioning in NAEP at the test level, aiming to provide important new insights into the validity and fairness of NAEP. The second study investigates measurement invariance of NAEP noncognitive indices across assessment years, applying the updated guidelines by Wu & Estabrook and the region of measurement equivalence method. The third study examines speededness in NAEP, by evaluating the effects of speededness on IRT estimation and group ability estimates and explores empirical guidelines to evaluate test speededness. The last study addresses validity in the context of Socioculturally Responsive Assessment, an ongoing quest to advance equity and fairness in educational assessment, by exploring how a reading passage from NAEP that is culturally relevant for Hispanic students is related to the performance of Hispanic students and other student subgroups.

Session Organizer:

*Young Yee Kim, American Institutes for Research*

Chair:

*Markus Broer, American Institutes for Research*

Participants:

#### **Exploration of Differential Test Functioning in NAEP**

*Soo Lee, American Institutes for Research; Xiaying Zheng, American Institutes for Research*

#### **Investigating Measurement Invariance in NAEP Student Questionnaire Indices**

*Xiaying Zheng, American Institutes for Research; Young Yee Kim, American Institutes for Research; Soo Lee, American Institutes for Research; emma king, American Institutes for Research; Xiaoying Feng, American Institutes for Research*

#### **Are NAEP Assessments Speeded? How to Evaluate Speededness of NAEP?**

*Young Yee Kim, American Institutes for Research; Xiaying Zheng, American Institutes for Research; Xiaoying Feng, American Institutes for Research*

#### **A Potentially Culturally Relevant NAEP Reading Text Passage And Student Literary**

*Sakiko Ikoma, American Institutes for Research; Xiaying Zheng, American Institutes for Research; yuan zhang, American Institutes for Research; Yifan Bai, American Institutes for Research; Markus Broer, American Institutes for Research*

Discussant:

*Mark Reckase, Psychometric Solutions*

### 127. Innovation Demonstration III

#### Paper Session

9:35 to 11:05 am

Convention Center: Level 1 - 120B

Participants:

#### **A Needs Assessment to Guide Professional Development in Justice-Oriented Assessment**

*Brian C Leventhal, James Madison University; Sara Finney, James Madison University; Elaine Kaye, James Madison University; Nicole Wilson, James Madison University*

We showcase a Needs Assessment to determine a professional's values, confidence, and perceived barriers to engaging in work related to fairness, equity, and social justice in educational assessment. We share the resource, discuss our PhD program's experience using the assessment, and how our unit used results to guide professional development.

#### **EdTalk: Intelligent Conversational Agent for Informational Retrieval from Education Reports**

*Bhashithe Abeysinghe, American Institutes for Research; Abhinav Cheruvu, Texas A&M University; Ruhan Circi, American Institutes for Research*

EdTalk is a conversational agent that utilizes large language models to conduct document retrieval from education reports by referring to its knowledge base. It generates factual answers by addressing the issue of hallucination. EdTalk aims to make information more accessible to a wide range of users.

#### **The Benefits of a Center of Excellence for Assessment Research**

*Janet Mee, NBME; Kerbie Addis, NBME; Edie Bissel, NBME; Angelica Hutchinson, NBME; Daniel Jurich, National Board of Medical Examiners; Leta S Re-McGlynn, National Board of Medical Examiners; Amber Montanano, NBME; John Moore, NBME; Linette P. Ross, NBME; Kimberly Swygert, National Board of Medical Examiners; Ye Tong, NBME*

An Assessment Research Center of Excellence is a pivotal initiative to foster advancements in assessment methodologies, uphold standardization, and promote communication among measurement professionals at a testing company. This cross-unit structure provides a place for sharing and collaboration that focuses on improving research processes across an organization.

### 128. Test Taking Effort and Careless Responding

#### Paper Session

9:35 to 11:05 am

Convention Center: Level 1 - 120C

Chair:

*Hyeri Hong, California State University, Fresno*

Participants:

#### **A Mixture Unfolding Model for Insufficient Effort Responses**

*Kuan Yu Jin, HKEAA; Chen-Wei Liu, National Taiwan Normal University*

Insufficient effort responses, which can reduce reliability and validity, are often collected in surveys. Following the ideal-point approach, a mixture unfolding model is proposed that not only performs parameter estimation but identifies aberrant respondents. The simulation results showed good parameter recovery and high classification rates of the new model.

#### **Examining Noneffortful Responding In Survey Responses And SEL Factors**

*Rik Lamm, Bloomington Public Schools; Mohammed Abulela, University of Minnesota; Kyle Nickodem, University of Minnesota - Twin Cities; Michael C. Rodriguez, University of Minnesota*

Using a state wide survey of students for nonacademic topics, noneffortful error is measured. This is measured overall across the survey along with being measured by grade. Factors measuring social and emotional learning were analyzed with the effect of removing the noneffortful responses.

### **Investigating the Relationships Between Test-taking Effort and Performance: Self-Report and Time-Based Measures**

*Xiaoxiao Liu, University of Alberta; Ying Cui, University of Alberta*

Process data from students' item-solving process could be used to measure examinees' cognitive effort in testing. Utilizing Programme for International Student Assessment (PISA) 2018 data, structural equation modeling analyses found that self-reported effort mediated the relationship between learning attitudes, resilience and time-based effort as well as performance.

### **Identifying Careless Responses in Surveys with Missing Data**

*Yuan Ge, The College Board; Stefanie A. Wind, University of Alabama; Eli Andrew Jones, The University of Memphis; Chia-Lin Tsai, University of Northern Colorado*

This research will evaluate the sensitivity of IRT indicators in identifying careless responders with incomplete data, which can help with the detection of patterns of missingness in realistic data collection settings and address the need of accurate response pattern indicators. Our preliminary findings suggested to include person-fit statistics in CR investigations.

Discussant:

*Edison Choe, Renaissance*

## **129. Innovations and Practices of AI and Language Models in Large-Scale Assessments**

### **Coordinated Paper Session**

**9:35 to 11:05 am**

**Convention Center: Level 1 - 121A**

Although not new, artificial intelligence (AI) gained extreme popularity with the introduction of ChatGPT in Fall of 2022. It resulted in great excitement and anxiety, particularly for educators. Advances in AI, natural language processing (NLP), and machine learning are now central topics in educational assessment. A proliferation of research and practices have explored the use of these techniques throughout the assessment process. While the potential for automatic item generation (AIG) and automated scoring were promising, concerns arose about privacy and the exclusion of human roles and professional expertise. Many fear the use of automated techniques and decision-making will result in a lack of interpretability of assessment scores and black boxes of accountability. Challenges have also arisen in ensuring the psychometric properties and the fairness of AI-based scores. The focus of this coordinated session is to investigate the advances of AI and NLP in large-scale assessment settings from research to practice. Over four papers, we explore several issues in evaluating the item quality within AIG, the use of language models in flagging learners in danger, and the use of AI in providing assessment feedback in practice. Attendees will improve their understanding of AI as it becomes increasingly prevalent in education.

Session Organizer:

*Shenghai Dai, Washington State University*

Participants:

### **Using Large Language Models for Evaluating Item Quality in Large-Scale Assessments**

*Guher Gorgun, University of Alberta; Okan Bulut, University of Alberta*

### **The Importance of Emotion Recognition for Identifying Learners in Danger in Large-Scale Assessments**

*Joshua Southerland, Pearson; Luis Alejandro Andrade, Pearson VUE; Lee Becker, Pearson; Kyle Habermehl, Pearson*

### **Providing Standards-Based AI-Generated Feedback to Students on Their Writing**

*Susan Lottridge, Cambium Assessment, Inc; Sherri Woolf, Consultant; Chris Ormerod, Cambium Assessment, Inc.; Michael Flynn, Cambium Assessment, Inc.; Julie Benson, Cambium Assessment, Inc.; Kevin Dwyer, Cambium Assessment*

### **Organizational Adoption of Generative AI Assessment and Feedback Tools – Early Successes and Challenges**

*Rachel Forman, McKinsey & Company*

# FULL SCHEDULE

**SUNDAY, APRIL 14**

Discussant:  
*Leslie Rutkowski, Indiana University*

## **130. NCME Mission Fund Selected Projects Showcase**

**Coordinated Paper Session**

**9:35 to 11:05 am**

**Convention Center: Level 1 - 121B**

As the NCME Mission Fund Committee, we are delighted to organize this session, dedicated to the exceptional individuals who have been recipients of our mission fund and honorable mention awards in the last two years. This session serves as a platform for these esteemed awardees to showcase and share their remarkable work, which has not only significantly contributed to the field of education but also exemplified our commitment to excellence and innovation. In this session, we will have the privilege of witnessing the transformative impact of their research and initiatives, highlighting the influence of the NCME Mission Fund on educational measurement.

Session Organizer:  
*Ren Liu, University of California, MERC*

Chair:  
*Stella Kim, UNC Charlotte*

Participants:  
**U**Developing and Evaluating Score Reports Adapted for Parents Whose Primary Language is Not English  
*Samuel Dale Ihlenfeldt, University of Minnesota*

**A Guiding Framework for Effective, Equitable Classroom Assessment for Teacher Practice and Student Success**  
*Dustin Van Orman, Western Washington University; Erin Riley-Lepo, The College of New Jersey*

**Validation & Implementation of a School-Based Gender Identity Screener for Children**  
*Matthew Diemer, University of Michigan*

**Identify and Address Complex, Intersectional Forms of Bias in Large-scale Assessments**  
*James Soland, University of Virginia*

## **131. Equity-Centered Innovations in Early Learning Assessment**

**Coordinated Paper Session**

**9:35 to 11:05 am**

**Convention Center: Level 1 - 121C**

Young children increasingly attend early childhood education settings before they enter kindergarten. Yet practitioners in these diverse settings often lack reliable, unbiased data on children's skills that could inform instruction and other programmatic decisions. This coordinated paper session introduces four novel assessment approaches for early learning settings intended to provide actionable insights to ultimately drive more equitable early learning outcomes. Several of these innovations rely on recent technological advancements, including generative AI, to capture the abilities of preliterate children who often have short attention spans and limited experience with formal learning situations. The first paper describes an approach that builds on traditional observation-based assessment by leveraging large language models to support more efficient and unbiased observations and incorporating tablet-based games. The second describes an integrated, informal assessment approach relying on activities embedded within a widely used learning application. The third explores whether playful tablet-based assessments can capture children's skills across multiple domains simultaneously. The final paper describes a literacy and language assessment developed with an anti-bias anti-racism (ABAR) lens. The session will conclude with a discussion on future directions in early learning assessment to center equity and ensure tools are aligned with the needs of end-users.

Session Organizer:  
*Emily Hanno, MDRC*

Chair:

*Ximena Portilla, MDRC*

Participants:

**An Integrated Observation and Game-Based Early Learning Assessment Approach**

*Ben Thier, Cognitive ToyBox*

**Embedded Assessment for Early Childhood**

*Kristen DiCerbo, Khan Academy; Caroline Hu Flexer, Khan Academy*

**Multidimensional Assessment Approach for Young Learners**

*Sara Bakken, BrainPop; Yigal Rosen, BrainPop*

**Applying an Anti-Bias, Anti-Racist Lens to Early Learning Assessment Development**

*Alisha Wackerle-Hollman, University of Minnesota; Lillian Duran, University of Oregon*

Discussant:

*Emily Hanno, MDRC*

### **132. Co-Design/Participatory Design Approaches to Assessment Development: Learnings from Four Implementation Studies**

**Coordinated Paper Session**

**9:35 to 11:05 am**

**Convention Center: Level 1 - 122A**

In recent years, the education and assessment communities have seen a renewed focus on culturally relevant, justice-oriented, socioculturally responsive approaches. An increasing number of frameworks and theoretical approaches have been proposed in the last years. A common theme across these equity-centered approaches is a call for including diverse stakeholders, including students, in the assessment development process. This approach is sometimes referred to as co-design or participatory design. While promising, the effective implementation of co-design/participatory design practices within the assessment development process remains an open question. During this session, we will present four research projects that implemented co-design/participatory design practices in their assessment development efforts. Each project uses a different approach to the implementation of these practices. After a summary of each project, presenters will discuss what value was gained from the implementation, what challenges were presented, and what was learned about co-design/participatory design practices in assessment development. The four presentations will be followed by a discussion session led by Dr. Solano-Flores, who will first offer his perspective on those practices and then open the floor for a collegiate discussion about co-design/participatory design practices.

Session Organizer:

*Cristina Anguiano-Carrasco, ACT*

Chair:

*Cristina Anguiano-Carrasco, ACT*

Participants:

**Cultural Content in Assessment Items: What Students Notice, Choose, and Prefer**

*Molly Faulkner-Bond, WestEd; Priya Kannan, WestEd; Jaylin N Nesbitt, James Madison University; Marianne Perie, WestEd*

**The Adult Skills Assessment Program (ASAP): From Socioculturally Responsive Principles to Actions**

*Javier Suarez-Alvarez, University of Massachusetts Amherst; April Zenisky, University of Massachusetts Amherst; Stephen G Sireci, University of Massachusetts, Amherst; Maria-Elena Oliveri, University of Nebraska*

**Toward More Equitable Assessments of Social and Emotional Learning (and Beyond)**

*Cristina Anguiano-Carrasco, ACT; Nola Daley, ACT; Jeremy Burrus, ACT; Nancy Lewin, ACT; Temple Lovelace, Assessment for Good; Lauren D. Kendall Brooks*

### **The Promise of Participatory Construct Mapping** *Fernanda Gandara, Room to Read*

Discussant:  
*Guillermo Solano-Flores, Stanford University*

### **133. Educational Measurement, Fifth Edition: A New Approach (Joint AERA/NCME Session)** **Coordinated Paper Session** **9:35 to 11:05 am** **Convention Center: Level 1 - 122B**

The first edition of Educational Measurement was edited by E.F. Lindquist and published in 1951, followed by three more editions. Those editions documented information about theory and techniques, updated it, and compared it to that in previous editions. The primary audience has traditionally been scientists with advanced degrees in measurement. The fifth edition will expand the focus of the audience to scholars in related fields and sciences. Consequently, the chapters have been crafted to be maximally accessible and useful to as broad an audience as possible. In this session, the editors provide a preview of the fifth edition, and discuss their goals and the changes introduced. Then, four chapter authors describe the goals and content of their chapters.

Session Organizer:  
*Mary Pitoniak, ETS*

Chairs:  
*Linda Cook, ETS*  
*Mary Pitoniak, ETS*

Participants:

#### **The History of Educational Measurement**

*Brian Clauser, National Board of Medical Examiners; Jerome Clauser, American Board of Internal Medicine (ABIM); Amanda Clauser, National Board of Medical Examiners*

#### **Validity in Educational Measurement**

*Suzanne Lane, University of Pittsburgh; Scott Marion, National Center for the Improvement of Educational Assessment*

#### **Assessment of Social-Emotional, Character, Behavioral, and Intrapersonal and Interpersonal Skills**

*Patrick Charles Kyllonen, ETS; Jiyun Zu, ETS*

#### **Assessment in International Contexts**

*Henry Braun, Boston College; Irwin Kirsch, ETS*

### **134. Deciphering US Students' Post-COVID Performance Results from National Assessment of Educational Progress** **Coordinated Paper Session** **9:35 to 11:05 am** **Convention Center: Level 1 - 123**

As the largest nationally representative, continuing evaluation of the condition of education in USA, the National Assessment of Educational Progress (NAEP) provides valid measures of the educational achievement and progress of the nation's students at established grades and ages. The NAEP 2022 results, the first post-pandemic NAEP results, showed declines in mathematics and reading at both grades 4 and 8, compared to 2019. The results were expected, as the COVID-19 pandemic brought widespread disruptions to education systems leading to learning losses or lags. However, how to

interpret the NAEP 2022 results remains a challenge as many factors could relate to such large declines. This session tries to decipher US students' post-COVID performance results in NAEP. The first study examines whether demographic compositional changes are associated with changes in grade 8 mathematics performance, applying propensity weights. The second study investigates the role of absenteeism in score declines between 2019 and 2022, using multiple regression analyses. The third study examines the absentee rate change by achievement level and contribution of the change to score declines using Blinder-Oaxaca decomposition. The fourth study investigates achievement gaps by SES before and after the pandemic and how changes differed by student group.

Session Organizer:

*Young Yee Kim, American Institutes for Research*

Chair:

*Xiaying Zheng, American Institutes for Research*

Participants:

### **Public-School Composition Change And Post-Covid Performance In NAEP: A Propensity Weights Approach**

*Martin Hooper, American Institutes for Research; David Bamat, American Institutes for Research; Rebecca Shipan, American Institutes for Research; George William Bohrnstedt, American Institutes for Research*

### **Potential Effect of Absenteeism on Declines in Students' Performance in NAEP 2022**

*Sinan Yavuz, American Institutes for Research; Young Yee Kim, American Institutes for Research*

### **Deeper Dive Into The Relationship Between Absenteeism and NAEP Performance**

*Young Yee Kim, American Institutes for Research; Sinan Yavuz, American Institutes for Research; Xiaying Zheng, American Institutes for Research*

### **Changes In Ses Achievement Gap Pre- And Post-Pandemic: Evidence From NAEP**

*Yifan Bai, American Institutes for Research; Markus Broer, American Institutes for Research*

Discussant:

*David Thissen, University of North Carolina*

## **135. Machine Learning Aspects of Computational Psychometrics**

### **Coordinated Paper Session**

**9:35 to 11:05 am**

**Convention Center: Level 1 - 124**

Computational psychometrics blends theory-based psychometrics with data-driven approaches from machine learning, data science, and artificial intelligence. Focusing on machine learning aspects, this coordinated paper session highlights five recent developments that harness computational psychometrics to confront educational measurement challenges spanning open item scoring, item calibration, and item feature selection. The first paper uses deep learning to automatically score open-response items on several criteria regarding style and content, and illustrates how such analyses can be implemented in R. Papers two to four aim to drastically minimize sample size requirements for piloting and calibration studies. The second paper uses automated machine learning to construct IRT models based on item-level features to speed up the piloting process and improve model fit. The third paper informs Bayesian estimation of the three parameter logistic model (3PL) with item parameters predicted by neural networks trained on items' linguistic features. The fourth paper presents a likelihood-free, neural-network-based estimation approach for the 3PL that draws on a given item pool to minimize sample size requirements. The fifth paper investigates feature selection procedures in the presence of missing data, comparing the performance of commonly used machine learning techniques (LASSO, Elastic Net) and recently developed approaches based on metaheuristics (Genetic Algorithm).

Session Organizers:

*Esther Ulitzsch, University of Oslo*

*Dmitry I. Belov, Law School Admission Council*

Participants:

### **Application of Deep Learning for the Evaluation of Text Responses**

*Rudolf Debelak, University of Zurich; Benjamin Wolf, University of Zurich*

### **AutoIRT: Improving the Fit of IRT Models Using Automated Machine Learning**

*James Sharpnack, Duolingo; Phoebe Mulcaire, Duolingo; Geoff LaFlair, Duolingo; Kevin Yancey, Duolingo*

### **Calibration Sample Reduction by Enriching IRT Estimation with NLP-Informed Parameter Predictions**

*Esther Ulitzsch, University of Oslo; Dmitry I. Belov, Law School Admission Council; Alexander Robitzsch, Leibniz Institute for Science and Mathematics Education; Oliver Lüdtke, Leibniz Institute for Science and Mathematics Education*

### **Neural Networks Approach to Estimate the 3PL in Small Samples**

*Dmitry I. Belov, Law School Admission Council; Esther Ulitzsch, University of Oslo; Oliver Lüdtke, Leibniz Institute for Science and Mathematics Education*

### **Machine Learning Approaches to Variable Selection with Ignorable Missing Data**

*Catherine Bain, University of Oklahoma; Dingjing Shi, University of Oklahoma*

Discussant:

*Hong Jiao, University of Maryland*

## **136. Alignment and Standard Setting: Interrogating Methodologies in Support of Instructional Utility**

### **Coordinated Paper Session**

**9:35 to 11:05 am**

**Convention Center: Level 1 - 125**

Quenemoen & Thurlow (2019) wrote, “PLDs [performance level descriptors] are definitions of grade-to-grade expectations, and as such are essential resources in determining whether an individual student has received access—through specialized instruction and supports—to the same content and expectations as other students, whether based on general or alternate achievement expectations (p. 26).” At the heart of their argument is the notion that PLDs support equal access to the rigor of learning expectations because they illuminate the state standards. This notion centers in using principled assessment design (PAD) and an evidenced-based approach for developing PLDs. It requires item writing and alignment to PLDs, and perhaps standard setting methods that more fully rely on PLDs being accurate. For assessments to be instructionally useful they may need to clearly describe validated theories of learning. This suggests evidence must be collected, interrogated, and if the data is not optimal, PLDs or items may need to be refined over time. When differences exist between Range PLDs and item difficulty, assessment developers may need to determine a framework under which Range PLDs or items can or should be revised.

Session Organizer:

*Christina Schneider, Cambium Assessment*

Participants:

### **Without Achievement Level Descriptors There Cannot Be Alignment**

*Ellen Forte, edCount, LLC*

### **Examining Test Score Interpretations on a Computer Adaptive Assessment for Alternate Assessment**

*Christina Schneider, Cambium Assessment; Jocelynn Pittman, Cambium Assessment; Matthew Gill, South Dakota Department of Education*

### **Refining Range ALDs: Bridging Predicted Learning Progressions with Empirical Insights**

*Karla Egan, EdMetric LLC; Melina Franklin, EdMetric, LLC*

### **Embedded Standard Setting Efficacy Under Different Item Alignment and Empirical Difficulty Conditions**

*Jing Chen, Cambium Assessment; Sangdon Lim, Cambium Assessment; Christina Schneider, Cambium Assessment*

Discussant:

*Susan Davis-Becker, ACS Ventures, LLC*

## **137. Is Your Test Instructionally Useful? How Do You Know?**

### **Coordinated Paper Session**

**9:35 to 11:05 am**

**Convention Center: Level 4 - Terrace Ballroom III**

Many assessments are promoted as instructionally useful. Unfortunately, this term and the companion term, “instructional utility,” are often used without clearly describing the characteristics of the assessment or evidence to support such claims. We define instructional utility “as it pertains to assessment processes and results, as the insights derived from the assessment, broadly speaking, that bear directly on the interactions among the teacher, student, and the content about student learning and instruction relative to specific learning targets that are the focus of current, immediately past, or near future instruction” (Evans and Marion, in press). In addition to clearly conceptualizing the terminology associated with the ways in which assessments might be used to support instruction and learning for the students taking the assessment, this session offers two presentations highlighting empirical analyses from two different types of testing programs to evaluate the degree to which these assessments are meaningfully serving instructional aims. More pragmatically, we aim to provide a framework for thinking about creating instructionally useful assessments to help educators and test developers better support student learning in their schools and districts.

Session Organizer:

*Scott Marion, National Center for the Improvement of Educational Assessment*

Participants:

### **Instructional Utility of Assessments: A Conceptual Framework**

*Scott Marion, National Center for the Improvement of Educational Assessment; Carla M. Evans, National Center for the Improvement of Educational Assessment*

### **Assessment Features that Impact Instructional Utility**

*Carla M. Evans, National Center for the Improvement of Educational Assessment; Scott Marion, National Center for the Improvement of Educational Assessment*

### **Investigating Teacher Use of Benchmark Modules to Improve Instruction in Utah**

*Kyla McClure, University of Colorado Boulder*

### **An Instructional Utility Case Study: Considering the Lousanna Innovative Assessment Pilot**

*Nathan Dadey, Center for Assessment*

Discussant:

*Lorrie Shepard, University of Colorado Boulder*

## **138. Hot Topics in Measurement - Roundtable Discussion II**

### **Organized Discussion**

**9:35 to 11:05 am**

**Convention Center: Level 4 - Terrace Ballroom IV**

Moderator:

*Jonathan Weeks, ETS*

**139. Classification****Paper Session****11:25 am to 12:25 pm****Convention Center: Level 1 - 120C**

Chair:

*Eunbee Kim, Georgia Institute of Technology*

Participants:

**A Unified True-Score Framework for Classification Indices with Application to Curriculum-Based Measures***Edison Choe, Renaissance; Adam E Wyse, Renaissance*

We introduce a unified treatment of computing classification accuracy and consistency indices based on true-score theory using single-administration data. New and existing models for true scores and errors are assembled, and some model fit criteria are suggested. The proposed framework is applied to data from several K-3 curriculum-based measures (CBMs).

**Classification Consistency and Accuracy with Conjunctive Decision Rule***Seohee Park, American Board of Internal Medicine*

This study considers the estimation of classification consistency and accuracy when handling a conjunctive decision. We propose four approaches using  $\alpha$ -cut scores: Integration with MVN, Integration with UVN, Monte-Carlo sampling with MVN, and Monte-Carlo sampling with UVN. To illustrate these methods, this study employs data from a large-scale certification organization.

**Exploration Of Multi-Year Pooling To Bolster The Reliability Of School Accountability Decisions***Brendan Longe, University of Massachusetts Amherst*

Small schools must be included in accountability systems despite the challenge of less reliable decisions from small n-sizes. One solution is pooling data across years. A multi-year accountability system was simulated to investigate the impact of sample pooling. Results help accountability designers evaluate pooling alternatives, a common but understudied practice.

Discussant:

*Won-Chan Lee, University of Iowa***140. Model Considerations****Paper Session****11:25 am to 12:25 pm****Convention Center: Level 1 - 121A**

Chair:

*Ami Lin, Aledev Consulting*

Participants:

**A Comparison of Item Parameter Estimates between Three- and Five-option MC Items***Jiayi Deng, University of Minnesota, Twin Cities; Yun Leng Wong, University of Minnesota; Joseph DeWeese, University of Minnesota; Seulbi Bailey Lee, Graduate Student; Patrick Kennedy, University of Oregon; Gina Biancarosa, University Of Oregon; David Weiss, University of Minnesota; Mark Davison, University of Minnesota*

This study aims to compare the estimates of item difficulty, item discrimination, and information functions between multiple-choice items comprising either three or five options, using data collected from a reading comprehension assessment. The findings suggested significant influence of option count on difficulty, discrimination, and information.

**Examining the Mode Effect on Students' Civic Perception and Political Participation***Yuan-Ling Liaw, IEA Hamburg; Rolf Strietholt, IEA Hamburg*

This paper explores the mode effect using data from IEA's International Civic and Citizenship Education Study (ICCS), which transitioned to digital assessment in 2022 via computers or tablets. Computer-based responses generally indicate higher importance of citizenship behaviors and greater expectancy of participation.

### **The Effects of Estimation Methods and Method Factors in Factor Analytic Techniques**

*Hyeri Hong, California State University, Fresno*

We investigated the effects of estimation methods and controls of negatively and positively phrased items for Neuroticism domain scale in the IPIP-NEO-300. The results showed that Bayesian-IP models best represented correlated factor, bifactor, and hierarchical frameworks of the structure of personality and controlling for method factors improved the model fits.

Discussant:

*Tiago A Caliço, Organization for Economic Cooperation and Development*

### **141. Application Essays and Characters in Higher Education Admissions**

#### **Coordinated Paper Session**

**11:25 am to 12:25 pm**

**Convention Center: Level 1 - 121B**

The recent rejection of race-based affirmative action from the Supreme Court presented new challenges to U.S. higher education, besides challenges related to the use of quantitative (e.g. high school grades) and qualitative (e.g., application essays and recommendation letters) information. These challenges have a complex impact on the diversity and success rate of college students. In this session, we will explore possible answers to these challenges from multiple angles. The first paper examines the application essays' content in its relationship with demographic and socio-economic background variables, admissions status, and post-admissions success indicators (such as first semester GPA), based on a large pool of applications for a public university. The second paper investigates AI's role in application essay preparation and its evaluation in admissions, exploring possible biases associated with the use of large language models in both processes. The third paper shares findings in terms of using behavioral skills assessment to assist with college and graduate admissions, and its implications for diversity, equity, and post-admissions success. Together, these three papers would provide new ideas and information to further the discussion and exploration of a valid and fair admissions process for U.S. higher education.

Session Organizer:

*Guangming Ling, ETS*

Chair:

*Ou Lydia Liu, ETS*

Participants:

#### **Evaluating Application Essays' Content in Shaping College Admissions Decisions and College Success**

*Guangming Ling, ETS; Michael Flor, ETS; Sugene Cho-Baker, ETS*

#### **The Evolution of AI Bias: Large Language Models and Holistic Review**

*Jinsook Lee, Cornell University; Aaron Jordan Alvero; Thorsten Joachims, Cornell University; Rene Kizilecec, Cornell University*

#### **Using A Behavioral Assessment To Assist Graduate Admissions: Some Initial Evidence**

*Patrick Charles Kyllonen, ETS; Daniel Fishtein, ETS*

Discussant:

*Li Cai, UCLA*

### 142. Cognitive Diagnostic Models Applications

#### Paper Session

11:25 am to 12:25 pm

Convention Center: Level 1 - 121C

Chair:

*Youn Seon Lim, Univeristy of Cincinnati*

Participants:

#### **Applications of Cognitive Diagnosis Modeling: A Systematic Review**

*Xiyu Wang, Purdue University; Yukiko Maeda, Purdue University; Hua-Hua Chang, Purdue University*

We conducted a systematic review on Cognitive Diagnosis Models (CDM) applied in educational settings. Although CDM has made significant methodological advancements in the past decade, it is still underutilized in practice. The study identifies key trends, challenges, and offers insights for enhancing the practical applications of CDM in education.

#### **Student Growth Percentiles in a Diagnostic Model Framework**

*Madeline Schellman, Pearson; Matthew James Madison, University of Georgia; Shiyu Wang, University of Georgia*

Modeling growth is important in large-scale assessment. For diagnostic classification models (DCMs) to realize their potential, researchers must examine how current growth modeling techniques can be employed in a DCM framework. This paper introduces the diagnostic growth percentile (DGP) and evaluates the reliability of DGPs under various assessment conditions.

#### **The Impact of Large-Scale Diagnostic Classroom Assessment on Student Achievement**

*Laine Bradshaw, Pearson; Ruchi Sachdeva, Pearson; Madeline Schellman, Pearson; Jordan Wheeler, University of Nebraska-Lincoln*

Diagnostic measurement-based assessment systems hold promise for improving equity in learning by providing detailed, just-in-time feedback for informing personalized learning in the classroom. We examine the impact of a large-scale, diagnostic measurement-based classroom assessment system for learning on student achievement across subjects, grades, and over time.

Discussant:

*Scott Monroe, UMASS Amherst*

### 143. Models for International Assessments

#### Paper Session

11:25 am to 12:25 pm

Convention Center: Level 1 - 122A

Participants:

#### **Applying LASSO Regression to Understand House Effect on PIRLS 2021 Turkey Results**

*Serife Okay, Hacettepe University; Sinan Yavuz, American Institutes for Research; Selahattin Gelbal, Hacettepe University*

This research explores the application of LASSO regression to the PIRLS 2021 Turkey dataset. Given the observed multicollinearity among the selected variables, LASSO, a regularization technique, is employed to derive more precise insights. The analysis also uses the five plausible values (PVs) provided in the dataset to ensure robust conclusions.

#### **Improving Graphical Response Scoring Accuracy in TIMSS 2023 with AI-Human Expert Integration**

*Lillian Tyack, Boston College; Matthias von Davier, Boston College*

Students sometimes respond to graphical response items in unusual ways where it may be unclear what score the response should receive, leading to inconsistent scoring between human raters. This presentation proposes an AI-based validation method to identify and select these “borderline” responses for expert validation and scoring in TIMSS.

#### **Underlying Generic Problem-Solving Competence in PISA 2012 Constructs: Skill Audit Findings**

*Lan Anh Nguyen Khoa, University of Melbourne*

The study investigates the existence of a Generic problem-solving concept in PISA 2012 Computer-

based assessment constructs. It argues for its existence in PISA 2012 Mathematics, Digital Reading, and Creative PS assessments, supported by skill audits of 38 released items. Findings suggest an underlying, generic PS competence across these domains.

Discussant:

*Leslie Rutkowski, Indiana University*

**144. Text & Measurement: Critical Approaches to Measuring Text & Measuring Our Field**  
**Coordinated Paper Session**  
**11:25 am to 12:25 pm**  
**Convention Center: Level 1 - 122B**

Amidst fast-moving change, the field of measurement has much to offer in text analysis, as well as much to learn from it. In this session, we share insights from the intersection of natural language processing (NLP) and measurement, with a focus on measuring biases in language. The first paper will seek to measure the biases of ChatGPT-generated text using NLP methods, in the context of a practice-based concern for high school counselors: ChatGPT-drafted letters of recommendation. The second presentation draws on NLP methods to analyze the use of deficit-based language within the measurement field itself in a quant-crit based study of measurement journals. The final presentation will set forth a framework for combining NLP methods with measurement methods to construct scales that measure constructs of interests across large bodies of text, with an application to syllabi data. Assessing and reducing biases in the production and review of text has remained a difficult problem to measure and address. Through our panel, we grapple with what it means to measure constructs, especially those related to equity, within text mediums that have not historically been examined at scale.

Session Organizer:

*Brein Mosely, Harvard University*

Chair:

*Sebastian Muñoz-Najar Galvez, Harvard University*

Participants:

**Deficit-based Language in Measurement Journals: A Quant-Crit Informed Textual Analysis**

*Brein Mosely, Harvard University; Zach Himmelsbach, Harvard University; Sebastian Munoz-Najar Galvex, Harvard University*

**Measuring Text: A Framework for Developing and Validating Scales for Text Analysis**

*Emma M. Klugman, Harvard Graduate School of Education*

**ChatGPT as a Writing Assistant: Merit and Adversity Framing in Recommendation Letters**

*Amy Desiderio, Harvard; Sheridan Stewart, Harvard; Sebastian Muñoz-Najar Galvez, Harvard; Julius DiLorenzo, Harvard; Jen Ha, Harvard; Matthew Nicola, Harvard; Tara P. Nicola, Harvard Graduate School of Education; Megan Richardson, Harvard; Mandy Savitz-Romer, Harvard*

Discussant:

*Luke Weisman Miratrix, Harvard Graduate School of Education*

## FULL SCHEDULE

SUNDAY, APRIL 14

### 145. Instructing and Assessing Competencies for Future Success: Research From International Baccalaureate

**Coordinated Paper Session**

**11:25 am to 12:25 pm**

**Convention Center: Level 1 - 123**

According to employers globally, competencies such as creative and critical thinking, analytical thinking, communication, and ethical thinking are critical for future workforce success (World Economic Forum, 2023). The International Baccalaureate (IB) recently proposed an ambitious research and development project focused on examining how instruction and assessment can accelerate the development of these competencies in schools and classrooms. In this coordinated session, the IB Global Director of Research and Design will discuss the IB's vision for the IB Learner. She will describe how IB is advancing the operationalization of "competencies of the future" through a focused research agenda. Following, independent researchers will present the results of two literature reviews on two success competencies: creative thinking and ethical thinking. Each presentation will offer a synthesis definition of each competency (creative thinking and ethical thinking), summarize evidence to support their instructional malleability, and summarize evidence informing effective pedagogy and assessment. Presentations will conclude with implications for future assessment and measurement research and practice, which apply to a broad range of success competencies.

Session Organizer:

*Willard Christopher Brandt, National Center for the Improvement of Educational Assessment*

Participants:

**IB International's Role in Supporting the Development of Competencies of the Future**

*Jennifer Merriman, International Baccalaureate*

**A Review of Creative Thinking and Implications for Assessment Design and Use**

*Chris Brandt, Center for Assessment*

**Defining and Assessing Ethical Thinking in Cross-National Education Contexts.**

*William A. Lorie, Center for Assessment*

Discussant:

*Thanos Patelis, Johns Hopkins University & University of Kansas*

### 146. Creating a Cadre of Culturally Responsive Measurement Professionals - GSIC and Digital Promise Joint Session

**Organized Discussion**

**11:25 am to 12:25 pm**

**Convention Center: 126A**

Presenters:

*Janine Jackson, Morgan State University*

*Darius D. Taylor, UMASS Amherst*

*Thao Vo, Washington State University*

*Sara Finney, James Madison University*

*Brian French, Washington State University*

*Drew Gitomer, Rutgers, The State University of New Jersey*

*Brian C Leventhal, James Madison University*

*Jennifer Randall, University of Michigan*

*Mary Roduta Roberts, University of Alberta*

*Michael Russell, Boston College*

*Stephen G Sireci, University of Massachusetts, Amherst*

*David Sul, Sul & Associates International*

*Catherine Taylor, University of Washington*

*Fiona Hinds, Independent Consultant*

*Emi Iwatani, Digital Promise*

*Susan Lyons, Lyons Assessment Consulting*

*Mya Poe, Northeastern University*

# FULL SCHEDULE

SUNDAY, APRIL 14

*Dustin Van Orman, Western Washington University*  
*Amanda Gorham, Center for Measurement Justice*  
*Eduardo Javier Crespo Cruz, UMASS Amherst*  
*Damon Bryant*  
*Maria Hamdani, Center for Measurement Justice*  
*Jenin Sabbah, Center for Measurement Justice*  
*Jade Caines Lee, University of Kansas*

## 147. Surveys of/for the Measurement Community

**Paper Session**

**11:25 am to 12:25 pm**

**Convention Center: Level 4 - Terrace Ballroom III**

Chair:

*Sanford R Student, University of Delaware*

Participants:

### **Analyzing Open-Responses In A Post-Exam Survey Using Natural Language Processing Methods**

*Yunyi Long, National Board of Osteopathic Medical Examiners; Xia Mao, NBOME*

This study compares two Natural Language Processing (NLP) methods and human coding for analyzing open-response survey data in a high-stakes medical examination. Initial findings revealed that NLP categorized open-responses in a similar way as human coding. The possibility and future directions of utilizing NLP to grasp candidates' feedback are discussed.

### **MxML Phase 2: Survey Of The Measurement Community**

*Yi Zheng, Arizona State University; Sijia Huang, Indiana University Bloomington; Susu Zhang, University of Illinois at Urbana-Champaign; Steven Nydick, Duolingo*

Studies incorporating ML into measurement have grown exponentially in the last 10 years. The MxML project explores the impact and consequences of adopting ML in measurement. The current Phase 2 of this project will conduct a survey of the measurement community about their experience, attitude, and thoughts toward ML.

### **The Item Response Warehouse (IRW): A Large Open Repository of Response Data**

*Ben Domingue, Stanford University; Klint Kanopka, New York University; Lucy Caffrey-Maffei, Stanford University; Radhika Kapoor, Stanford University; Susu Zhang, University of Illinois at Urbana-Champaign; Yiqing Liu, University of Illinois*

In contrast with other disciplines, psychometrics has limited public data. The IRW is designed to change that through the collection and standardization of a large volume of item response datasets. We describe key elements of the data standardization process and the nearly 1 billion item responses already in the IRW.

Discussant:

*Brian Gong, Center for Assessment*

## 148. The Digital SAT: Concordance Studies and Post-Admin Analyses

**Coordinated Paper Session**

**1:15 to 2:45 pm**

**Convention Center: Level 1 - 120B**

The SAT suite assessment has been taken digitally by international students since March 2023 and will be taken digitally starting in Fall 2023 for students taking PSAT/NMSQT and 2024 for U.S.-based students taking SAT. The new digital SAT, administered as a multi-stage adaptive test, continues to measure the knowledge and skills that students learn in high school and matter most for college and career readiness. The digital SAT suite assessments are scored on a 1600 scale and scores can be used

# FULL SCHEDULE

**SUNDAY, APRIL 14**

to track growth across the SAT Suite of Assessments over time. Concordance studies and post-admin analyses have been conducted to ensure that the digital SAT assessments maintain the scale score metric and score integrity. In this coordinated session, five presentations will focus on three aspects of maintaining the score metric and integrity of the score for digital SAT assessments: concordance studies, post-admin psychometric and forensic analyses. Analyses results and key findings on these topics will be shared to address the questions and issues brought by the transition to the digital SAT, a multi-stage adaptive test.

Session Organizer:

*Weiwei Cui, College Board*

Participants:

**Straight-line Concordance Studies**

*Thomas Proctor, College Board*

**Concordance Studies: Link Paper SAT to Digital SAT**

*Weiwei Cui, College Board*

**PSAT/NMSQT Concordance Studies**

*Oliver Zhang, College Board*

**Post-Admin Monitoring Studies**

*Jonathan Beard, College Board*

**Post-Admin Forensic Analyses**

*Hui Deng, College Board*

Discussant:

*Richard Melvin Luecht, University of North Carolina at Greensboro*

## **149. Revisiting Common Assumptions and Practices in Vertical Scaling**

**Coordinated Paper Session**

**1:15 to 2:45 pm**

**Convention Center: Level 1 - 121A**

Increasingly, vertical scales are being used in practice to better understand student progress in core academic subjects over time. Yet, the construction of vertical scales often relies on practices and assumptions supported by relatively little evidence, especially given the consequences associated with uses of those scales. In this coordinate paper session, we revisit some of those assumptions and practices. The first study (Briggs) examines an assumption central to use of vertical scales—namely that they possess equal-interval properties—and provides options for psychometricians to test that assumption that are both axiomatic and pragmatic. The second (Soland & Edwards) and third (Liao & Bolt) revisit rules of thumb about how many linking items are needed in adjacent grades, including when common forms of measurement model misspecification occur. Meanwhile, the fourth paper (Student) uses moderated nonlinear factor analysis to probe assumptions underlying an incredibly common practice, namely using different grades from cross-sectional data as a proxy for how students develop longitudinally, which assumes differences across grades are not affecting item parameters. In total, this panel not only raises questions about one of the most important tools in psychometrics, but also tries to provide solutions that can be used by measurement practitioners.

Session Organizer:

*James Soland, University of Virginia*

Participants:

**On the Justification and Evaluation of an Interval Scale Derek**

*Christian Briggs, University of Colorado Boulder*

**Revisiting the 20% Rule of Thumb for Linking Items in Vertical Scales**

*James Soland, University of Virginia; Kelly Edwards, American Board of Pediatrics*

# FULL SCHEDULE

**SUNDAY, APRIL 14**

## **Implications of IRT Model Misspecification on Vertical Scale Linking**

*Xiangyi Liao, University of Wisconsin, Madison; Daniel Bolt, University of Wisconsin, Madison*

## **Validating, Calibrating and Linking Vertical Scales with Moderated Item Response Theory**

*Sanford R. Student, University of Delaware*

Discussant:

*Scott Monroe, UMASS Amherst*

## **150. Incorporating Multiple Data Source to Explore Group Strategy Variations**

### **Coordinated Paper Session**

**1:15 to 2:45 pm**

**Convention Center: Level 1 - 121B**

This coordinated session focuses on incorporating multiple data source to explore the group strategy variations beyond merely response data to address potential measurement fairness issues. Our session covers a broad range of process data exploration and showcases both advanced psychometric modeling and data-driven approach to identify the variations in problem-solving strategies and provide potential solution to leverage these variations in the measurement. The first paper presents a response time-based finite-mixture item response theory approach for flexible modeling examinee-and-item-specific item-solving strategies. The second paper introduces a mixture hidden Markov model to leverage action sequences with elapsed time to highlight test takers' strategy variation in reading pause states. The third paper demonstrates how to combine the Seldonian algorithm and adversarial neural networks to estimate the degree that students' online or remote learning preparedness levels impact feature extraction and predictive model interpretation, focusing on the "time-on-task" features. The fourth paper presents the strategy variations in missing responses and behavioral pattern extraction in interactive computational thinking tasks across countries via a sequence mining approach. These studies lead a pressing direction of integrating multiple data source in large-scale assessments to provide a new angle to assure the equitable measurement across groups with various backgrounds.

Session Organizer:

*Qiwei He, Georgetown University*

Participants:

### **A Response Time-based Finite-mixture Item Response Theory Model for Item-Solving Strategy**

*Sijia Huang, Indiana University Bloomington; Jinwen Luo, UCLA; Minjeong Jeon, UCLA*

### **Identifying Group Difference by Latent Sequence Modeling on Process Data with Elapsed Time**

*Ling Chen, Columbia University; Qiwei He, Georgetown University*

### **Towards Fair Learning Analytics: An Application of the Seldonian Algorithm**

*Okan Bulut, University of Alberta; Jinnie Shin, University of Florida; Wallace Nascimento Pinto, University of Florida*

### **Profiling Behavioral Patterns in Interactive Computational Thinking Tasks Across Countries**

*Qiwei He, Georgetown University; Eugenio Gonzalez, ETS*

Discussant:

*Steven Culpepper, University of Illinois Urbana-Champaign*

## **151. Cognitive Diagnostic Models II**

### **Paper Session**

**1:15 to 2:45 pm**

**Convention Center: Level 1 - 121C**

Chair:

*Yesim Ozer Ozkan, Gaziantep University*

Participants:

### **A More Accurate Evaluation of the Potential Advantages of the eMC-DINA Model**

*Zechu Feng, The University of Hong Kong; Jimmy de la Torre, University of Hong Kong*

The eMC-DINA model is a flexible CDM for multiple-choice data. Although its advantages over the MC-DINA and DINA models have been demonstrated, the results may not be accurate given how the comparison models were treated. This paper presents a more accurate comparison, and shows the advantages of the eMC-DINA remain.

### **Combining Regularization and Logistic Regression Model to Validate the Q-Matrix for CDM**

*Xiaojuan Sun; Tao Xin, Beijing Normal University; Naiqing Song, Southwest University*

In this study, we combine the L1 regularization and multiple logistic regression (MLR) model to validate the Q-matrix. Results show that the regularized MLR method: (a) produces better Q-matrix recovery rate, true positive rate, and true negative rate than either the MLR-B or the Hull method for most conditions.

### **On Posterior Probability of Latent Classifications in Diagnostic Models**

*Jimmy de la Torre, University of Hong Kong; Rodrigo Schames Kreitchmann, Universidad Nacional de Educacion a Distancia*

Parametric estimation of person parameters in cognitive diagnosis models hinges on posterior probabilities that are sample-dependent. This study investigates the impact of sample dependency on model calibration and scoring, proposing a multiple imputation procedure to correct it. Simulation results highlight the superiority of the new method with small samples.

### **Small Sample Cognitive Diagnosis Using Bayesian Catalytic Priors**

*David Arthur, Purdue University Statistics; Arman Sabbaghi, Purdue University Department of Statistics; Hua-Hua Chang, Purdue University*

Cognitive Diagnosis Models (CDMs) are powerful tools but can be underutilized in classroom settings due to large sample size requirements. Through simulation studies and using Tatsuoka's fraction subtraction dataset, we demonstrate how Bayesian catalytic priors can lead to improved estimation and classification accuracy of CDMs in small sample sizes.

Discussant:

*Anne Corinne Huggins-Manley, University of Florida*

## **152. Process Data**

### **Paper Session**

**1:15 to 2:45 pm**

**Convention Center: Level 1 - 122A**

Chair:

*Magdalen Beiting-Parrish, CUNY Graduate Center*

Participants:

### **A Scanpath Trajectory Reading Eye-Tracking Spatio-Temporal Similarity (STRESS) Measure**

*Matthew David Naveiras, Riverside Insights; Sun-Joo Cho, Peabody College of Vanderbilt; Amanda Goodwin, Vanderbilt University; Jorge Salas, Vanderbilt University; Eduardo Davalos, Vanderbilt University*

Eye-tracking data can be compared between people to investigate how readers process text differently. This study presents a similarity measure (STRESS) designed to capture the dynamic spatio-temporal characteristics of eye-tracking data that are unique to reading data. The STRESS measure is validated through a qualitative analysis and a simulation study.

### **Test-Taking Experience: How Does Automated Feedback Impact Achievement Emotions Over Time?**

*Ute Mertens, Leibniz Institute for Science and Mathematics Education; Livia Kuklick, IPN Kiel, Germany; Marlit Annalena Lindner, IWM - Leibniz Institut für Wissensmedien, University of Tübingen*

We investigated the impact of two automated feedback approaches on the emotional responses of 335 undergraduate participants during a fact-based science test. Analysis of process data revealed that providing feedback significantly enhanced student emotions compared to the absence of feedback, particularly towards the end of the test.

### **Using Process Data to Assess Item Readability Impact on Accessibility Feature Usage**

*Adam Hearn, American Institutes for Research; Burhan Ogut, American Institutes for Research; Ruhan Circi, American Institutes for Research*

Using process data from the NAEP G8 Math assessment, this study explored the relationship between item readability and accessibility feature usage. Using a random-effects approach, we found that more complex text was associated with varying usage of certain accessibility features, with significant differential effects for students with disabilities.

Discussant:

*Klint Kanopka, New York University*

### **153. A Debate on the Merits of Embedded Standard Setting**

#### **Organized Discussion**

**1:15 to 2:45 pm**

**Convention Center: Level 1 - 122B**

Embedded standard setting (ESS) has emerged as an alternative to the traditional multi-round in-person or online standard setting workshop. Daniel Lewis will argue that when item writers create test items purposefully aligned to carefully crafted performance level descriptors (PLDs), and when empirical data support the alignment of items to PLDs, then cut scores can be estimated directly (as described by Lewis & Cook, 2020 and Lewis, Lee, & Choi; 2022). Marianne Perie has collected data from at least one state that, she will argue, challenge some of the assertions made by Dr. Lewis regarding ESS. She will share those data as well as caveats about the method. In the course of the debate, Drs. Lewis and Perie will have an opportunity to present their respective cases and challenge each other's arguments. Gregory Cizek will provide commentary on both presentations and ask questions to elicit clarification of points made by each speaker. Following the debate, audience members will have an opportunity to ask questions, comment, or prompt further debate.

Session Organizer:

*Michael Brannen Bunch, Measurement Incorporated*

Moderator:

*Michael Brannen Bunch, Measurement Incorporated*

Presenters:

*Daniel Lewis, Creative Measurement Solutions LLC*

*Marianne Perie, WestEd*

Discussant:

*Gregory Cizek, University of North Carolina Chapel Hill*

### **154. Equity and Fairness: Building Relationships that Bridge Gaps**

#### **Organized Discussion**

**1:15 to 2:45 pm**

**Convention Center: Level 1 - 123**

As our nation continues to recover from the pandemic's impact, our marginalized students continue to recover from the years of racism and inequities that preceded the pandemic. While progress is noted in both areas, it is slow. Much work lies ahead of us, particularly in equity. The work towards educational equity and social justice has been exhausting but advocates nationwide continue to advance the work by challenging long-standing beliefs and promoting new pathways. Despite some progress, passionate advocates continue to push for much needed change to our current education, assessment, and

accountability systems. Change will require us to build relationships and come together as champions. This session does just that. Panelists will utilize their diverse experiences to explore the benefits of bridging gaps within and between existing advocacy organizations; between research and practice; between federal, state, and local leaders. The challenge? No gap will or can be bridged without building relationships first. The goal? To highlight the efforts of some of our most respected equity in assessment advocates as they discuss the importance of building relationships that improve our educational instruction and assessments especially for those most historically marginalized.

Session Organizers:

*Elda Garcia, Imagine Learning*

*Christina Laster, Parent and Family Advocate*

Presenters:

*Maria Armstrong, Association of Latino Administrators and Superintendents*

*Molly Faulkner-Bond, WestEd*

*Jade Caines Lee, University of Kansas*

**155. Culturally Responsive Assessment in Classrooms and Large-Scale Contexts: Theory, Research, and Practice**  
**Organized Discussion**  
**1:15 to 2:45 pm**  
**Convention Center: Level 1 - 124**

This invited session describes how a culturally responsive lens shapes assessment purpose, design, implementation, use, and validation. The panelists will introduce an NCME volume that explores the conceptual foundations and implications of culturally responsive assessment in K-12 large-scale and classroom contexts. Editors will provide an overview of the book's purpose, organization, and contributors. Scholars, who serve as discussants within the book, will provide an overview of each section of the book and their perspectives about the themes and ideas in the given section. Questions relevant to the discussion include: • What is culturally responsive assessment, why is it important and needed? • What foundations of current assessment practice and shape contemporary assessment design, implementation, use, and validation practices? • What practices and conceptual foundations will need to change (or be adjusted) to yield assessment tools and processes that are culturally responsive? • How will culturally responsive assessment differ in K-12 large-scale versus classroom assessment contexts? The goals for the session are to provide current perspectives on culturally responsive assessment and to prompt discussion among the panelists and the audience about how measurement and assessment must change to better align with a pluralistic and multicultural society.

Session Organizer:

*Carla M. Evans, National Center for the Improvement of Educational Assessment*

Moderator:

*Kadriye Ercikan, ETS*

Presenters:

*Catherine Taylor, University of Washington*

*Carla M. Evans, National Center for the Improvement of Educational Assessment*

*Guillermo Solano-Flores, Stanford University*

*Elena Diaz-Bilello, University of Colorado Boulder, Center for Assessment, Design, Research and Evaluation (CADRE)*

*Robert Joseph Mislevy, retired*

### 156. Medical and Credentialing Test Applications

#### Paper Session

1:15 to 2:45 pm

Convention Center: Level 1 - 125

Chair:

Leslie Keng, National Board of Medical Examiners

Participants:

#### Comparability between In-Person versus Virtual Modality of Performance Examinations

Karen Fung, The Pharmacy Examining Board of Canada (PEBC); Salma Satchu, The Pharmacy Examining Board of Canada; Mahmoud Suleiman, The Pharmacy Examining Board of Canada; Yuen Chu, The Pharmacy Examining Board of Canada; John Pugsley, Pharmacy Examining Board of CA

A pilot study was conducted by a Canadian certification body to compare the traditional in-person format of a performance exam (OSCE) to the virtual modality. Comparability from the perspectives of reliability, validity, performance, and participant experiences were examined. Adaptations made, analyses performed, and results will be discussed in this presentation.

#### Evaluating Chatbot Performance on Board Continuous Certification Examinations: Methodological Considerations and Results

Caroline Prendergast, American Board of Surgery; Andrew Jones, American Board of Surgery; Carol L Barry, American Board of Surgery; Chandler Scott-Smith, American Board of Surgery

The recent expansion of access to LLM-powered chatbots has raised questions about chatbot performance on certification and licensure assessments. This study evaluates the performance of multiple chatbot programs on a medical board continuous certification assessment and outlines the implications of various research design considerations when evaluating chatbot performance.

#### Investigating Predictors of Response Time on a Medical Knowledge Test

Marc Anthony Johnson, National Board of Medical Examiners; Leslie Keng, National Board of Medical Examiners; Linette P. Ross, NBME; Jada Willse, University of North Carolina

The evaluation of item response time has become a popular investigative analysis with the rise of computer-based testing. This study investigates how examinee characteristics, performance, and item type interact as predictors of item response time. This study will aid test developers in using item response time data in constructing exams.

Discussant:

Chad W. Buckendahl, ACS Ventures, LLC

### 157. Personalization in the Service of Equity: What, Why, How, and for Whom?

#### Organized Discussion

1:15 to 2:45 pm

Convention Center: Level 4 - Terrace Ballroom III

The growing recognition of the need to address longstanding issues of social justice is causing some education departments to revise content standards and some assessment organizations, and many members of our field, to rethink traditional definitions of equity and fairness. Evolving definitions include the need to account for differences in sociocultural backgrounds, funds of knowledge, interests, values, and practices that individuals from diverse cultures bring to learning and assessment. In keeping with this need, a variety of approaches to assessment have been proposed. Among those approaches are culturally responsive assessment, socioculturally responsive assessment, antiracist assessment, culturally sustaining assessment, justice-oriented assessment, and universal design for assessment. These approaches put into focus various student groups, some targeted (e.g., African-American students, English learners, students with disabilities) and others more broadly constituted (e.g., minoritized learners). The goal of this organized discussion is to interrogate the idea that rather than conceptualizing approaches to assessment around such groupings, it might be more productive substantively, logistically, and politically to design for the individual—i.e., to personalize. Among the questions the panel will address are what does such design entail, how might we know if it was successful, and what conceptual and technical issues might it raise.

Session Organizer:  
*Randy Bennett, ETS*

Presenters:  
*Lauren Kendall Brooks, Assessment for Good*  
*Blair Lehman, ETS*  
*Aneesha Badrinarayan, Learning Policy Institute*  
*Sandip Sinharay, ETS*

### **158. Electronic Board Session IV** **Electronic Board Session** **1:15 to 2:45 pm** **Convention Center: Level 4 - Terrace Ballroom IV**

Participants:

#### **1. A Fit Propensity Effect Size Approach to Improve Model Selection**

*Holmes Finch, Ball State University; Brian French, Washington State University; Jason Immekus, University of Louisville*

Despite their limitations, latent variable model selection relies on fit indices. Fit propensity (FP) is advanced to overcome limitations, where good model fit and low FP compared to alternative models are preferred. An effect size approach to characterizing FP is investigated using a simulation study. Results demonstrate improved model selection.

#### **2. An Experimental Investigation of Preknowledge Effects on Speed and Accuracy Using IRTrees**

*Justin L. Kern, University of Illinois at Urbana-Champaign; Hahyeong Kim, University of Illinois at Urbana-Champaign*

IRTtree models are IRT models that assume a tree structure for response processes. In the current project, we develop a two-level IRTtree model to investigate the effects of preknowledge on speed and accuracy, accounting for speed effects on accuracy, in an experimentally obtained data set.

#### **3. A Two-Step Q-Matrix Estimation Method**

*Hyunjoo Kim, University of Illinois at Urbana-Champaign; Hans Friedrich Koehn, University of Illinois at Urbana-Champaign; Chia-Yi Chiu, University of Minnesota*

The validity of cognitive diagnosis (CD) depends crucially on the correct specification of the Q-matrix. Data-driven Q-matrix estimation methods have been developed that purportedly are more accurate than curricular experts. An efficient two-step Q-matrix estimation is proposed for use with any CD model.

#### **4. Ensuring Knowledge in Certification Using Natural Language Processing**

*Heath Kincaid, American Board of Obstetrics and Gynecology; Anthony Sparks, American Board of OBGYN; Pooja Shivraj, American Board of OBGYN; George Wendel, American Board of Obstetrics and Gynecology*

The American Board of Obstetrics and Gynecology has created a novel procedure using Natural Language Processing combined with Subject Matter Experts to identify overlapping content on an examination. This novel procedure ensures displaying a breadth and depth of knowledge are mandatory to become certified, and helps ensure certification decisions defensibility.

#### **5. Evaluation of Vertical Scaling for the Alternate English Language Proficiency Assessment**

*Yage Leah Guo, Center for Applied Linguistics; Shu Jing Yen, Center for Applied Linguistics*

Alternate assessments have historically not been developed with the same level of psychometric rigor as the regular assessments developed for general education population. Therefore, schools cannot measure growth across grades or over time. The purpose of this study is to evaluate the new vertical scale by comparing two calibration procedures.

**6. Examining Long-Term Dynamics in Multistage Adaptive Testing**

*Hacer Karamese, WIDA at the University of Wisconsin-Madison; Won-Chan Lee, University of Iowa; Kyoungwon Lee Bishop, WIDA at University of Wisconsin Madison*

This study explores the impact of accumulated linking error, particularly on the final scores obtained through a long-term linking chain established during field-testing. A series of simulations were conducted to investigate the long-term impact of the linkage plan, calibration method, field-test sample distribution, and population shifts on the MST scores.

**7. Examining the dimensionality for the OSCE using Exploratory Graph Analysis**

*Nai-En Tang; Igor Himelfarb, NBCE; Andrew Gow, National Board of Chiropractic Examiners (NBCE); Chia-Lin Tsai, University of Northern Colorado*

This study examined the dimensionality of the chiropractic objective structured clinical examinations (OSCE) license exam using exploratory graph analysis (EGA). Varying numbers of dimensions and low item stability from the replicate bootstrap samples were found. This may suggest the existence of complex structure and within-item multidimensional properties of OSCE exam.

**8. Investigating Measurement Invariance in the Content-Language Integrated Education for Multilingual Learners Scale**

*Jung Yeon Park, George Mason University; Sujin Kim, George Mason University; xiaowen chen, George Mason University*

This study validates a scale assessing teachers' critical awareness for MLs in content classrooms based upon 458 U.S. teacher responses. Using categorical multiple-group CFA approach, we evaluate measurement invariance across roles and grade-levels. We conduct a simulation study to examine the utility of alternative model fit statistics and cut-off values.

**9. Optimizing Online Calibrations: Strategies for Polytomous Item Sets in Computer Adaptive Tests**

*William J Muntean, National Council of State Boards of Nursing; Joe Betts, NCSBN; Zhuoran Wang, National Council of State Boards of Nursing (NCSBN)*

Computerized Adaptive Testing (CAT) efficiently estimates candidate abilities through strategic item administration. Despite its merits, adaptive principles are not widely used for pretesting polytomous items. This study explores the advantages of online calibrations for polytomous items and item sets using novel selection algorithms and termination criteria.

**10. Predicting SJT Item Difficulty in Military Promotion Testing**

*Avery Nicole Garcia, United States Air Force; Isabelle Ebert, United States Air Force; Mark Rose*

This study investigated prediction of SJT item difficulties ( $n_s = 35,139 - 42,561$  for two tests) by SME ratings of SJT response effectiveness. About 30 subject matter experts rated forty SJT items. Findings indicated that SME ratings predicted difficulty, and certain characteristics (rater job level) affected these predictions.

**11. Item Parameter Drift of Social and Emotional Learning Measures Over Time**

*Alejandra Miranda, University of Minnesota - Twin Cities; Pablo Vivas Corrales, University of Minnesota; Mireya Carmen-Martinez Smith, Amplify Education; Michael C. Rodriguez, University of Minnesota*

Item parameter drift is important, may jeopardize validity and fairness of score interpretation, and is rarely examined for noncognitive measures. A large dataset was used to examine IPD over nine years in multiple ways for three noncognitive measures. No serious IPD was found, suggesting good stability of item parameter estimates.

**12. Detection of Differential Item Functioning in Multidimensional Testing Using IRT-Residual DIF Framework**

*Bowen Wang, University of Florida; Hwanggyu Lim, GMAC; Kyung (Chris) T. Han, GMAC*

We explore the IRT-residual based DIF (RDIF) detection method in multidimensional testing contexts. Given its promising performance in the preliminary simulation, along with appealing features like computational simplicity and no need for separate item calibration, the RDIF framework can be an effective methodology for assessing DIF within multidimensional testing environments.

**13. Exploring the Feasibility of Micro-Certification for Out-of-School Youth in the Philippines**

*Farhan Azim, The University of Melbourne; Bruce Beswick, The University of Melbourne; Rebekah Luo, University of Melbourne; Marlene Ferido, The University of the Philippines; Lalaine Bagui, The University of the Philippines*

This study explored the potential of micro-certification within the Philippine Alternative Learning System (ALS) to improve educational outcomes and address low enrolment and completion rates for out-of-school youth. Focusing on two life skills – Communication and Teamwork & Cooperation – the study developed and validated assessment frameworks, instruments, and supporting materials.

**14. Impact of Item Characteristics on Item Selection Procedures in Mokken Scale Analysis**

*Benjamin Kweku Lugu, University of Alabama; Stefanie A. Wind, University of Alabama*

We examined the impact of problematic item characteristics on the accuracy of item scale classifications based on automated selection procedures (AISP) in Mokken Scale Analysis (MSA). Although AISP was fairly robust to violation of invariant item ordering, results v monotonicity and scalability violations compromised the accuracy of scale assignment.

**15. Improving Transformer-Based Automated Scoring of K-12 Science Items**

*Alexander Kwako, Cambium Assessment, Inc.; Chris Ormerod, Cambium Assessment, Inc.*

Transformer-based language models, such as ELECTRA, are often effective at scoring short, constructed response (SCR) items. Yet, for some times, ELECTRA underperforms. This paper focuses on six, difficult-to-autoscore K-12 science items. We explore four techniques for improving model performance. Results indicate our novel approaches can improve SCR autoscore with ELECTRA.

**16. Measurement Invariance of Social and Emotional Learning Measures across Four Administrations**

*Mohammed Abulela, University of Minnesota; Kyle Nickodem, University of Minnesota - Twin Cities; Michael C. Rodriguez, University of Minnesota*

We assessed measurement invariance for five social and emotional learning measures across four administrations (2013, 2016, 2019, 2022) of a statewide student survey using RMSEAD for the nested model comparisons with ordinal items. We also estimated standardized mean differences, which indicated sharp post-COVID decline.

**17. Using Process Data and Neuroimaging to Enhance the Validity of Neuropsychological Assessment**

*Nixi Wang, University of Washington; Shuping Lim, National University of Singapore; Aveline A. Vasu, National University of Singapore; Suan Peen Foo, National University of Singapore; Evelyn Law, National University of Singapore*

This study examined the utility of process data and neuroimaging evidences in advancing neuropsychological assessment research. Specifically, it showed that the integration of multiple data sources allows for a more precise characterization of the predictive validity of neuropsychological assessment of executive function (EF).

**159. Culturally Responsive and Sustaining Approaches to Constructed Response Scoring in Assessment**

**Coordinated Paper Session**

**3:05 to 4:35 pm**

**Convention Center: Level 1 - 120C**

Purposeful action to reduce majority bias is required throughout a scoring system to ensure valid and fair score interpretations for all students. The four papers in this session weave together recommendations for scoring essays in large-scale assessment from a panel of diverse voices and perspectives. Authors will provide a path that leads to a culturally responsive scoring system: focusing on item and rubric development; prioritizing the representation of diverse groups in the scoring sample; and providing hand- and automated-scoring recommendations designed to be practical to implement and posed to offer an immediate impact. Authors will discuss assumptions that underlie scoring and how bias may encroach into the most well-intentioned scoring system without an active antiracist stance; identify approaches that can lead to more culturally responsive and sustainable scoring processes to promote equitable measurement practices; provide evidence to support their claims; and, encourage large-scale assessment providers to be intentional in applying culturally responsive practices in the development of human and automated scoring systems.

Session Organizer:  
*Sarah Quesen, WestEd*

Participants:  
**Culturally Responsive and Sustaining Approaches to Item and Rubric Alignment**  
*Heather Roeters-Solano, Pearson*

**Culturally Responsive and Sustaining Approaches to Sample Representation and Metrics for Evaluation**  
*Jaylin Nesbitt, WestEd*

**Culturally Responsive and Sustaining Approaches to Hand-scoring Processes**  
*Susan Lottridge, Cambium Assessment, Inc*

**Culturally Responsive and Sustaining Approaches to Automated-Scoring Processes**  
*Sarah Quesen, WestEd; Karen Lochbaum, Pearson*

Discussant:  
*Lauren White, Pearson*

**160. A Principled Design Approach to Developing an Alternate English Language Proficiency Assessment**  
**Coordinated Paper Session**  
**3:05 to 4:35 pm**  
**Convention Center: Level 1 - 121A**

The valid assessment of a complex construct like English language development targeted for the small and diverse population of students with the most significant cognitive disabilities presents unique challenges and requires a systematic test development process. This session's presenters will detail the design and development of the Alt ELPA—an alternate English language development test developed under a principled assessment design (PAD) framework. Specifically, the presenters will (a) provide a context for the challenges due to the complexity of the construct and diversity of the small student population, (b) recount how states and researchers applied principled assessment design to understand, articulate expectations for, and measure the receptive and productive language development of students while creating a fair and accessible assessment with meaningful scores, (c) describe calibrating and scoring the Alt ELPA, given the challenges of federal reporting requirements, small sample sizes and high degrees of missingness and incompleteness of data, and (d) describe the Embedded Standard Setting (ESS) process for the Alt ELPA and in particular, how ESS places specific, but modest requirements throughout the PAD test development lifecycle to support a coherent assessment system. A nationally recognized expert in special student populations will discuss and contextualize the various presentations.

Session Organizer:  
*Daniel Lewis, Creative Measurement Solutions LLC*

Chair:  
*Cathryn Still, ELPA21*

Participants:  
**Developing a Research-Based Alt ELPA Theory of Action and Accessibility Policy**  
*Kristin Liu, University of Minnesota*

**Expanding Expectations of Communication for English Learners to Increase Fairness in Assessment**  
*Cathryn Still, ELPA21*

**Calibration and Scoring of the Alt ELPA**  
*Yun-Kyung Kim, UCLA; Li Cai, UCLA*

**Embedded Standard Setting for the Alt ELPA**  
*Daniel Lewis, Creative Measurement Solutions LLC*

Discussant:  
*Martha Thurlow, National Center on Educational Outcomes*

**161. Large Language Models in Assessment: Guidelines, Characteristics and Biases**  
**Coordinated Paper Session**  
**3:05 to 4:35 pm**  
**Convention Center: Level 1 - 121B**

In the rapidly evolving landscape of education, testing and assessment, AI-powered large language models (LLMs), such as ChatGPT, are becoming increasingly prevalent. This coordinated session aims to explore various aspects of LLMs that help to better understand their properties, strengths, and weaknesses. While the focus is on writing assessments, we conjecture that many of the presented findings will generalize to other areas as well. The session features four talks: The first two talks explore the linguistic characteristics and overall quality of AI-generated essays. The first talk studies the difference between several state-of-the-art LLMs, both proprietary and open source, while the second talk focuses on the effect of the sampling temperature – a simple yet consequential parameter. The remaining two talks focus on detection of AI-generated essays. The third talk provides guidelines and discusses best practices for the use of such detectors, while the fourth talk investigates the issues of potential biases detectors might show against or in favor of non-native speakers.

Session Organizer:  
*Michael Fauss, ETS*

Participants:  
**Comparison of Essays Generated by Different LLMs**  
*Yang Zhong, University of Pittsburgh; Jiangang Hao, ETS; Michael Fauss, ETS*

**Effects of Sampling Temperature on Writing Style and Quality**  
*Michael Fauss, ETS; Jiangang Hao, ETS; Chen Li, ETS*

**Practical Considerations of Using Detectors of AI-Generated Essays: Dos and Don'ts**  
*Jiangang Hao, ETS; Michael Fauss, ETS*

**Towards Investigating the Fairness of Detecting LLM-Generated Essays**  
*Yang Jiang, ETS; Jiangang Hao, ETS; Michael Fauss, ETS; Chen Li, ETS*

**162. Addressing Fairness Concerns in Digital-First, AI-Powered Testing Environments**  
**Coordinated Paper Session**  
**3:05 to 4:35 pm**  
**Convention Center: Level 1 - 121C**

The testing landscape has dramatically changed in recent years due to the digitization of test content and the use of AI in test development and scoring. Although these changes have led to lower costs, shorter assessments, and a greater focus on delightful, personalized experiences for test takers, they also raise new fairness concerns that threaten the credibility of assessments among stakeholders. Some of these concerns include a general distrust in “black box” AI decision-making, a belief that assessments produce unfair outcomes (e.g., in university admissions), and that professional practices

# FULL SCHEDULE

**SUNDAY, APRIL 14**

in the assessment industry are insensitive to complex forms of bias (e.g., bias due to intersectionality). In this coordinated session, the presenters seek to address some of these fairness concerns via the development of responsible AI standards, the justification of delightful test content, the analysis of measurement bias across multiple background variables, and the personalization of test content under well-defined constraints. The discussant will prepare questions of each presenter and facilitate discussion among the attendees.

Session Organizer:  
*Will Belzak, Duolingo*

Participants:

### **Responsible AI for Assessment as Professional Responsibility**

*Jill Burstein, Duolingo; Geoff LaFlair, Duolingo; Kevin Yancey, Duolingo; Alina A. von Davier, Duolingo*

### **Digital-First Content Development for Test-Taker Delight and Fairness**

*Yena Park, Duolingo*

### **Evaluating DIF Across Multiple Background Variables Simultaneously**

*Will Belzak, Duolingo*

### **Constraining for Fairness**

*Stephen G Sireci, UMASS Amherst; Duy N. Pham, UMASS Amherst*

Discussant:  
*Victoria Yaneva, NBME*

## **163. Process Data and Data Quality**

### **Paper Session**

**3:05 to 4:35 pm**

**Convention Center: Level 1 - 122A**

Chair:  
*Mo Zhang, ETS*

Participants:

### **Detecting Pre-Knowledge Via Classification Tree-Based Modeling Of Responses, Response Times, and Visual Fixations**

*Kaiwen Man, University of Alabama; Qipeng Chen, University of Alabama; Peida Zhan, Zhejiang Normal University*

Technology-enhanced learning system (TELS) has drawn much attention in educational assessment recently. To gain behavioral inferences by properly modeling multi-source data collected from TELS, this study proposes an innovative structural classification tree-based model for jointly modeling responses, response times, and visual fixations to detect aberrant test-taking behavior like preknowledge responding.

### **Susceptibility Of Data Quality Indicators To Prevalence Of Low Quality Responses**

*Nivedita Bhaktha, GESIS*

In this study, we examined the performance of the three most commonly used data quality indicators - maximum long string, intra-individual response variability, and Mahalanobis distance, in identifying low-quality responses (LQR) in the sample. We present the results of a simulation study that considered different types and prevalence of LQR.

### **Treating Rapid Responses As Incorrect Improves Estimates Of Test Performance**

*Daniel B. Wright, UNLV; Sarah Wolff, UNLV*

Rapid responding can be a sign of non-thoughtful responding. Several approaches have been put forward that can have value in different situations. Using grade 9-12 personal learning assessments, NAEP, and ACT data, we show that treating rapid response as incorrect improves the reliability for all three of these assessments.

### **Upsetting the Apple CART: Classification Trees for Process Data from Performance-Based Assessments**

*Nate Smith, UMASS Amherst; Rich Feinberg, National Board of Medical Examiners; Chunyan Liu, National Board of Medical Examiners; Lisa Keller, UMASS Amherst*

This research analyzes process data from a virtual performance-based assessment (VPBA), where examinees manage a simulated patient scenario. Through the application of classification trees to sequenced actions, this study illuminates an interpretable, data-driven methodology for highlighting salient patterns in VPBAs, enhancing understanding of examinee actions within complex action spaces.

Discussant:

*Okan Bulut, University of Alberta*

## **164. Moving from Accountability to Improvement: A Framework and Lessons from the Field**

### **Coordinated Paper Session**

**3:05 to 4:35 pm**

**Convention Center: Level 1 - 122B**

Recent policy debates about the future of accountability in education have emphasized the need to re-imagine accountability systems in ways that will promote equitable, high-quality learning opportunities for all students. Educators and researchers have offered numerous policy recommendations including an expansion of systems to measure a broader set of outcomes, an increased emphasis on monitoring opportunity to learn (OTL), and, perhaps most significantly, a prioritization of supports for school improvement that will benefit all learners, especially those who have been denied access to well-resourced, effective, and supportive learning environments. This session explores efforts to re-imagine accountability systems as accountability and improvement systems, drawing on recent research on prior efforts to use large-scale data to promote school improvement. The first presentation discusses five strategies for improving these systems, making them more inclusive, equitable, and informative. The remaining four presenters will describe concrete applications of these strategies, exploring their implications in more detail. The session will include a discussant who will synthesize the work presented and moderate a discussion with the audience.

Session Organizer:

*Laura Hamilton, American Institutes for Research*

Participants:

### **A Framework to Improve Accountability, and Examples from the Field**

*Chris Brandt, Center for Assessment; Juan Manuel D'Brot, National Center for the Improvement of Educational Assessment*

### **Filling in the (Data) Gap for Equitable Decision Making**

*Samantha Neiman, American Institutes for Research; Kyosin Kang, American Institutes for Research*

### **Measuring and Monitoring Comprehensive, Whole-Child Supports**

*Jonathan Schweig, RAND Corporation*

### **Incorporating Teachers' Perspectives into Accountability and Improvement System Design**

*Sam Rikoon, American Institutes for Research; Margarita Olivera Aguilar, ETS GLOBAL; Laura Hamilton, American Institutes for Research*

### **Student Experience and Measurement in Education: A DisCrit QuantCrit Turn**

*David Hernandez-Saca, University of Northern Iowa*

Discussant:

*Lillian Pace, KnowledgeWorks*

### 165. Performance Levels and Standard Setting

#### Paper Session

3:05 to 4:35 pm

Convention Center: Level 1 - 124

Chair:

*Xinhui Maggie Xiong, ETS*

Participants:

#### **A Task-based Approach to External Validation in Standard Setting Research**

*Kathryn Hille, ETS; Jonathan Schmidgall, ETS*

Overlapping Task Type Performance is a novel source of external validity evidence for cut scores on newly (re)designed tests that share some task types with an existing test that already has operational data and established cut scores. Methodology, a successful implementation, and potential applications for personalized assessments will be shared.

#### **Incorporating Cluster Analysis Results into Standard Setting**

*Ji Zeng, Michigan Department of Education; Shiqi Hao, Michigan Department of Education; Jessalyn Smith, DRC*

The primary goal of this study is to employ statistical methodology to help determine plausible cut score ranges that can narrow panelist's focus at standard setting. We conducted a K-modes cluster analysis on two summative NGSS science assessments to examine the method's effectiveness.

#### **The Stability of Cut Score Estimates from the Beuk Standard Setting Method**

*Joe Grochowalski, College Board; Lauren Molin, College Board; Lei Wan, College Board; Amy H. Hendrickson, College Board*

We examined the stability of cut scores from the Beuk method. We studied the effects of panelist agreement and score distributions, and we introduce a bootstrapping procedure to estimate standard errors. We simulated varying panel disagreement, sizes, and number of bootstrap replications. The results provide guidance about achieving optimal stability.

Discussant:

*Michael Brannen Bunch, Measurement Incorporated*

### 166. Differential Item Functioning II

#### Paper Session

3:05 to 4:35 pm

Convention Center: Level 1 - 125

Chair:

*Montserrat B Valdivia Medinaceli, UC DAVIS*

Participants:

#### **A Framework for Evaluating Intersectional Fairness in Algorithmic Decision Making**

*Youmi Suk, Teachers College Columbia University; Kyung (Chris) T. Han, GMAC*

This study proposes an extension of a recently proposed framework for algorithmic fairness, known as "differential algorithmic functioning" (DAF), to include the concept of intersectionality. The proposed intersectional DAF framework provides theoretical and technical definitions, as well as practical tools and guidelines for assessing intersectional fairness in modern algorithms.

#### **Detecting and Investigating Polytomous Items with Nonuniform-Bidirectional Drift**

*Hao Jia, National Council of State Boards of Nursing NCSBN; William J Muntean, National Council of State Boards of Nursing; Joe Betts, NCSBN; Zhuoran Wang, National Council of State Boards of Nursing (NCSBN)*

The Monte Carlo simulation was employed to model two testing administrations, encompassing 81 scenarios, where polytomous items demonstrated nonuniform-bidirectional drift. The findings revealed notable impacts of the drift on distinct ability groups and substantiated the efficacy of the Weighted Root Mean Square Difference in identifying nonuniform-bidirectional drift over conventional methods.

### **Detecting DIF in the Multi-Unidimensional Pairwise Preference (MUPP) IRT Model** Lavanya

*Shravan Kumar, University of South Florida; Naidan Tu, University of South Florida; Sean Joo, University of Kansas; Stephen Stark, University of South Florida*

Multidimensional forced-choice (MFC) measures are gaining prominence in noncognitive assessment. Yet there has been little research on detecting differential item functioning (DIF) with models for forced-choice measures. This research examines Lord's chi-square and item parameter replication (IPR) methods with MFC tests based on the Multi-Unidimensional Pairwise Preference (MUPP) model.

Discussant:

*Dubravka Svetina Valdivia, Indiana University*

## **167. Validating Assessment in a Multicultural, Pluralistic Context**

### **Coordinated Paper Session**

**3:05 to 4:35 pm**

**Convention Center: Level 4 - Terrace Ballroom III**

This coordinated paper session brings together four papers, each of which introduces a new framework for validating assessments in a culturally, linguistically, and socially diverse society. The papers in this session each offer a different approach, but all recognize the central roles that language and culture play in learning, and therefore in the assessment of learning. Additionally, each paper grapples with the sociopolitical power of assessment to perpetuate or disrupt patterns of systemic marginalization. This session is intended to create conversation among leading measurement professionals who are redefining what it means to validate an assessment in a pluralistic context, where diversity is explicitly valued. After a brief overview of the frameworks from each of the presenting authors, our discussant, Dr. Solano-Flores, will share insights into where our frameworks overlap and where they present opposing viewpoints. We will close with a facilitated panel conversation among the presenting authors that aims to build consensus and identify the most promising paths forward for advancing the theoretical and conceptual foundations of validity within a multicultural, pluralistic setting.

Session Organizer:

*Susan Lyons, Lyons Assessment Consulting*

Participants:

### **Assessment Validation Frameworks For Culturally Responsive Assessment**

*David Slomp, University of Lethbridge; Mya Poe, Northeastern University; Maria Elena Oliveri, Buros Center for Testing-UNL*

### **Validity Considerations For Culturally Responsive Assessments**

*Scott Marion, National Center for the Improvement of Educational Assessment; Suzanne Lane, University of Pittsburgh*

### **A Framework For Enacting Equity Aims In Assessment Use: A Justice-Oriented Approach**

*Susan Lyons, Lyons Assessment Consulting; Maria Elena Oliveri, Buros Center for Testing-UNL; Mya Poe, Northeastern University*

### **Prioritizing Just Social Outcomes For Diverse Readers And Writers: A Conceptual Framework**

*Elena Forzani, Boston University; Julie Corrigan, Concordia University; David Slomp, University of Lethbridge*

Discussant:

*Guillermo Solano-Flores, Stanford University*

# PARTICIPANT INDEX

(Last Name, First Name, Session Number)

- Abeyasinghe, Bhashithe, 127  
Abulela, Mohammed, 128, 158  
Ackerman, Terry, 008  
Adam, Hugh, 058  
Adams, Elizabeth L., 114  
Adams, Grant, 093  
Adams, Raymond J, 060  
Addis, Kerbie, 127  
Adkins, Deborah, 099  
A Keller, Lisa A., 105  
Alahmadi, Sarah, 045  
Alarcon, Lorena, 106  
Albano, Tony, 041, 081, 098  
Aldrup, Karen, 045  
Ali, Usama, 078  
Almond, Russell G, 033  
Alvero, Aaron Jordan, 141  
Alves, Cecilia, 115  
Andersen, Nico, 120  
Andrade-Lotero, Luis Alejandro, 117, 129  
Angehr, Emily, 099  
Anguiano-Carrasco, Cristina, 132  
Ankomah, Francis, 081  
Antal, Judit, 074  
Araneda, Sergio, 059  
Armoush, Nour, 081, 098  
Armstrong, Maria, 154  
Arslan, Burcu, 066  
Arthur, Ann, 109  
Arthur, David, 151  
Atitsogbey, Desmond, 069  
Avvisati, Francesco, 060, 084  
Awwal, Nafisa, 121  
Ayers-Wright, Elizabeth, 031, 069, 103  
Azim, Farhan, 158
- Basaraba, Deni, 024  
Bates, Shaun, 091  
Beach, Sarah, 091  
Bean, Jerry, 069  
Beard, Jonathan, 148  
Beaty, Roger E., 084  
Becker, Kirk, 069, 096  
Becker, Lee, 117, 129  
Bediwy, Ahmed, 045, 081  
Beigman Klebanov, Beata, 081  
Beiting-Parrish, Magdalen, 032, 152  
Belov, Dmitry I., 135  
Belur, Vinetha, 046  
Belzak, Will, 162  
Benedek, Mathias, 084  
Bennett, Randy, 112, 157  
Benson, Julie, 129  
Berezner, Alla, 060  
Bergner, Yoav, 100  
Berry, Yufeng, 069  
Bertling, Jonas, 078  
Beswick, Bruce, 158  
Betebenner, Damian, 048, 080  
Betts, Joe, 158, 166  
Betzberger, Justin, 038  
Bhaduri, Indrani, 078  
Bhaktha, Nivedita, 163  
Bhandari, Shreya, 098  
Biancarosa, Gina, 140  
Bilir, Mustafa Kuzey, 049  
Binici, Salih, 099  
Bishop, Kyoungwon Lee, 057, 103, 158  
Bissel, Edie, 127  
Blanco, Maria Paz, 077  
Blankenship, Charles, 122  
Block, Jared, 081  
Bohrnstedt, George William, 134  
Bolender, Brad, 053  
Bolt, Daniel, 022, 041, 110, 119, 149  
Bonifay, Wes, 015, 063, 072  
Boothe-Corley, Rebekah, 069  
Bouy, Alicia, 117  
Boykin, Allison, 094  
Bradshaw, Laine, 067, 142  
Braithwaite, Reina, 119  
Brandt, Chris, 145, 164  
Brantley, Wyman, 043  
Braum-Bharti, Nina James, 105  
Braun, Henry, 044, 122, 133  
Brennan, Robert, 059  
Brice, Amanda, 118  
Briggs, Derek Christian, 044, 048, 080, 092, 100, 116, 149  
Briscoe, Michael, 097  
Broer, Markus, 111, 126, 134
- Brookhart, Susan, 092  
Brown, Emily Anne, 069  
Brown, Richard, 069  
Brunetti, Matthew, 085  
Bryant, Damon, 108, 146  
Bryk, Anthony, 090  
Buckendahl, Chad W., 045, 101, 156  
Buczek, Matthew, 119  
Bulut, Hatice Cigdem, 069  
Bulut, Okan, 027, 045, 057, 066, 081, 086, 098, 129, 150, 163  
Bunch, Michael Brannen, 153, 165  
Burkhardt, Amy, 055  
Burrus, Jeremy, 132  
Burstein, Jill, 105, 162  
Buzick, Heather, 075
- Caffrey-Maffei, Lucy, 147  
Cai, Li, 141, 160  
Cai, Qing, 076  
Caliço, Tiago A, 084, 122, 140  
Camara, Wayne J., 038, 059  
Cancado, Luciana, 051  
Cao, Yichong, 045  
Cardwell, Ramsey, 105  
Carling, Dona, 051  
Carney, Michele, 100  
Casabianca, Jodi M, 034, 054  
Castaneda, Ruben, 039, 057  
Catts, Hugh, 061  
Chakraborty, Roti, 091  
Chalhoub-Deville, Micheline, 108  
Cham, Heining, 045  
Chang, Hua-Hua, 105, 116, 142, 151  
Chang, Wanchen, 049  
Chao, Hsiu-Yi, 087, 105  
Chatzidimitriou, Despoina, 115  
Chatzimichail, Myria, 115  
Chawala, Kamal, 078  
Chen, Becky, 045  
Chen, Dandan, 077, 098, 121  
Chen, Guanyu, 087  
Chen, Hong, 057, 081  
Chen, Jianshen, 033, 087, 111  
Chen, Jihang, 029, 122  
Chen, Jing, 136  
Chen, Jinsong, 069  
Chen, Jyun-Hong, 087, 105  
Chen, Ling, 150  
Chen, Ping, 083, 152  
Chen, Qipeng, 163  
Chen, Tingting, 035, 073  
Chen, Xiaowen, 158  
Chen, Ying, 083  
Cheng, Ming-Min, 111  
Cheng, Yiling, 022

# PARTICIPANT INDEX

(Last Name, First Name, Session Number)

Cheng, Ying, 087, 118  
Cheruvu, Abhinav, 127  
Chiu, Chia-Yi, 067, 105, 158  
Cho, Sun-Joo, 152  
Cho, Youngmi, 057, 105  
Cho, YoungWoo, 109  
Cho-Baker, sugene, 141  
Choe, Edison, 128, 139  
Choi, Hanna, 105  
Choi, Hye-Jeong, 069  
Choi, Ikkyu, 054, 089, 098  
Choi, Jaehwa, 050  
Choi, Jeongwon, 081, 125  
Choi, Jinah, 069, 124  
Choi, Jinnie, 057  
Choi, Seung W., 016  
Choi, Youn-Jeng, 050, 105  
Chu, Yuen, 156  
Chung, Gregory, 001, 105  
Cipriano, Christina, 102  
Circi, Ruhan, 093, 127, 152  
Cizek, Gregory, 153  
Clark, Amy, 099  
Clauser, Amanda, 025, 133  
Clauser, Brian, 035, 133  
Clauser, Jerome, 133  
Claxton, Patricia, 119  
Clelland, Ashley, 027  
Cloney, Daniel, 060  
Cohen, Allan, 027, 117  
Collins, Chris, 090  
Connolly, Helena, 123  
Cook, Linda, 133  
Corrigan, Julie, 167  
Cotterell, Michael, 093  
Cox, Olivia, 048  
Crawford, Kristen, 097  
Crespo Cruz, Eduardo Javier, 146  
Cui, Mengyao, 110  
Cui, Weiwei, 089, 148  
Cui, Wenju, 047  
Cui, Ying, 045, 128  
Culpepper, Steven, 116, 150  
  
Dadey, Nathan, 085, 137  
Dai, Shenghai, 129  
Dailey, David, 069  
Daisher, Ted, 118  
Daley, Nola, 132  
Dallas, Andrew, 037, 105  
Dartnell, Pablo, 077  
Davalos, Eduardo, 152  
Davidson, Anne H, 095  
Davis, Jennifer, 093  
Davis, Laurie, 048, 080, 087  
Davis-Becker, Susan, 025, 136  
  
Davison, Mark, 140  
D'Brot, Juan Manuel, 009, 164  
Deane, Paul, 032, 114  
Debelak, Rudolf, 135  
DeCarlo, Lawrence T., 110  
de la Torre, Jimmy, 004, 067, 151  
De Lisle, Jerome, 119  
DeMars, Christine, 045, 077  
Demir, Cihan, 081  
Demirkaya, Onur, 057, 069  
Deng, Hui, 148  
Deng, Jiayi, 033, 069, 140  
Deng, Xinyue, 045  
DePascale, Charles, 080  
Deshmukh Towery, Ila, 102  
Desiderio, Amy, 144  
DeWeese, Joseph, 140  
Diao, Hongyu, 049  
Diaz, Maria, 058  
Diaz-Bilello, Elena, 155  
DiCerbo, Kristen, 032, 131  
Diemer, Matthew, 130  
DiLorenzo, Julius, 144  
Dixon-Roman, Ezekiel, 116  
Dobria, Lidia, 027  
Dogan Lacin, Betul Gokcen, 069  
Dolan, Robert, 095  
Domaleski, Chris, 065  
Domingue, Ben, 041, 061, 072, 147  
Dray, Amy, 048  
Drossel, Kerstin, 029  
Drost, Bryan, 100  
Du, Kuixi, 045  
Dumas, Denis, 084  
Dunbar, Stephen, 081  
Dunn, Jennifer, 010  
Duran, Lillian, 061, 131  
Dworak, Elizabeth, 058  
Dwyer, Kevin, 129  
  
Ebert, Isabelle, 158  
Edeh, Ejike, 063, 075  
Edwards, Kelly, 149  
Egan, Karla, 040, 136  
Eickelmann, Birgit, 029  
Embretson, Susan, 041  
Engelhard, George, 114  
Englert, Kerry, 057, 099, 106  
Ercikan, Kadriye, 033, 155  
Evans, Carla M., 137, 155  
Everson, Howard, 035, 107  
  
Fan, Fen, 035  
Fan, Meng, 069  
Fang, Luyang, 086  
Faulkner-Bond, Molly, 068, 095, 132, 154  
  
Fauss, Michael, 054, 161  
Federiakin, Denis, 027, 051  
Feinberg, Rich, 057, 065, 163  
Feng, Tianying, 001, 105  
Feng, Wanyong, 086  
Feng, Xiaoying, 126  
Feng, Zechu, 151  
Ferido, Marlene, 158  
Ferland, Alane, 062  
Fernandez, Maria-Paz, 038  
Ferrara, Steve, 028, 051, 062, 091  
Fina, Anthony D., 045, 081  
Finch, Holmes, 072, 158  
Fincher, Melissa, 028  
Fink, Aron, 120  
Finnegan, Robert, 085  
Finney, Sara, 057, 127, 146  
Firoozi, Tahereh, 086, 117  
Fishtein, Daniel, 141  
Fleckenstein, Johanna, 081  
Fleetwood, April, 125  
Flor, Michael, 054, 141  
Flynn, Michael, 129  
Fnu, Chandramani, 086  
Foo, Suan Peen, 158  
Forman, Rachel, 129  
Forsyth, Carol McGregor, 033, 038, 090  
Forte, Ellen, 028, 036, 107, 136  
Forzani, Elena, 167  
Foster, David, 059  
Franklin, Melina, 136  
French, Brian, 088, 146, 158  
Frey, Andreas, 120  
Frey, Sharon, 069  
Friedman, Tim, 074  
Fu, Jianbin, 105  
Fu, Yanyan, 030, 042, 093, 096  
Fulmer, Gavin, 123  
Fung, Karen, 156  
Furgol Castellano, Katherine, 048, 085, 124  
  
Gales, Mark, 055  
Gandara, Fernanda, 132  
Gane, Brian, 099  
Gao, Shilin, 055  
Gao, Yizhu, 045  
Garcia, Avery Nicole, 158  
Garcia, Elda, 154  
Ge, Yuan, 026, 043, 128  
Gelbal, Selahattin, 143  
Gershon, Richard, 058  
Gezer, Tuba, 119  
Gholson, Melissa L., 095

# PARTICIPANT INDEX

(Last Name, First Name, Session Number)

- Gianopulos, Garron, 124  
Gierl, Mark J., 042, 081, 098, 117  
Gilbar, Charlotte, 068  
Gill, Matthew, 136  
Gitomer, Drew, 146  
Gochyyev, Perman, 094, 110  
Goldhammer, Frank, 066, 120  
Gombert, Sebastian, 120  
Gong, Brian, 097, 147  
Gonzalez, Eugenio, 150  
Gonzalez, Mario, 031  
Gonzalez-Wegener, Xaviera, 082  
Gooch, Reginald M, 046  
Goodman, Joshua, 037, 101  
Goodman, Madeline, 085  
Goodwin, Amanda, 152  
Gorgun, Guher, 057, 086, 098, 129  
Gorham, Amanda, 146  
Gorham, Jerry L., 038  
Gorney, Kylie, 033, 041, 066, 087, 115  
Gorno Tempini, Marilu, 061  
Gotzmann, Andrea Julie, 115  
Gow, Andrew, 158  
Grabovsky, Irina, 025, 037  
Graf, Edith Aurora, 005  
Grant, Rosalie, 106  
Griger, Cassandra, 081  
Griph, Gerald, 069  
Grochowalski, Joseph, 115, 165  
Gu, Yuqi, 083  
Guez, Ava, 084  
Guo, Hongwen, 033, 066, 120  
Guo, Qi, 115  
Guo, Yage Leah, 158  
Guven, Ahmet, 121
- Ha, Jen, 144  
Haab, Sergio, 045, 093  
Haberman, Shelby Joel, 065  
Habermehl, Kyle, 117, 129  
Ham, Eun Hye, 120  
Hamdani, Maria, 082, 146  
Hamilton, Laura, 102, 164  
Hamilton, Lauren, 069  
Han, Catherine, 058  
Han, Kyung (Chris) T., 017, 030, 049, 096, 158, 166  
Han, Suhwa, 049, 094  
Han, Youngjin, 093  
Hanno, Emily, 131  
Hanrahan, Rachel, 058  
Hao, Jiangang, 030, 047, 104, 161  
Hao, Shiqi, 165  
Hardy, Michael, 114  
Harris, Deborah J., 019, 088, 104, 109  
Haryanto, Bakir, 060
- Hatcher, Wilson, 069  
Hauenstein, Clifford Erhardt, 063  
Hauerwas, Laura Boynton, 114  
Haviland, Sara, 046  
Hayes, Anne, 095  
Haynie, Kathleen C, 038  
He, Qiwei, 150  
He, Yi, 039, 069  
Hearn, Adam, 152  
Hellman, Scott, 117  
Hendrickson, Amy H., 115, 165  
Hernandez-Saca, David, 164  
Hester, Olivia, 069  
Hicks, Juanita, 057, 093  
Hille, Kathryn, 165  
Hilz, Anna, 045  
Himelfarb, Igor, 057, 073, 158  
Himmelsbach, Zach, 144  
Hinds, Fiona, 082, 108, 146  
Hipfner-Boucher, Kathleen, 045  
Ho, Andrew, 092, 116  
Ho, TsungHan, 057  
Hoang, Andrew, 090  
Hofman, Abe, 045  
Höft, Lars, 081  
Hong, Hyeri, 035, 063, 128, 140  
Hooks-Singletary, Brianna, 108  
Hooper, Martin, 134  
Horbach, Andrea, 081  
Horowitz, Karen, 048  
Houriet, Andrew, 093  
Howard, Tyrone, 068  
Hsieh, Ching-Ni, 117  
Hu, An, 069  
Hu, Ann, 062  
Hua, Cheng, 073  
Huang, Chi-Yu, 069  
Huang, Fuchun, 060  
Huang, Qi, 110  
Huang, Sijia, 147, 150  
Huang, Yingshi, 083  
Huang, Yue, 055  
Huff, Kristen, 052, 092, 097, 107  
Hu Flexer, Caroline, 131  
Huggins-Manley, Anne Corinne, 063, 088, 151  
Huo, Huade, 122  
Hutchinson, Angelica, 127
- Ihlenfeldt, Samuel Dale, 026, 130  
Ikoma, Sakiko, 126  
Ilagan, Michael, 081  
Immekus, Jason, 158  
Ingrisone, James, 115  
Ingrisone, Soo, 115  
Isaacs, Jazmin, 066
- Ishii, Hidetoki, 105  
Iwatani, Emi, 146
- Jackson, Janine, 091, 108, 146  
Jaeger, Garrett, 047  
Jansen, Brenda, 045  
Jansen, Thorben, 081  
Jarr, Karoline, 085  
Jasso, Javier, 061  
Jeon, Eunjeong, 105  
Jeon, Minjeong, 074, 093, 105, 150  
Ji, Jeongmin, 081  
Jia, Hao, 166  
Jiang, Tao, 103  
Jiang, Yang, 161  
Jiang, Yanlin, 105  
Jiao, Hong, 086, 135  
Jin, Kuan Yu, 128  
Jin, Yi, 069  
Jing, Zeyuan, 043  
Joachims, Thorsten, 141  
Johnson, Evelyn, 119  
Johnson, Janice Lee, 062  
Johnson, Marc Anthony, 156  
Johnson, Matthew S, 033, 054, 066, 116  
Johnstone, Christopher, 095  
Jolivette, Kristine, 069  
Jones, Andrew, 025, 156  
Jones, Christopher, 055  
Jones, Edmund, 055  
Jones, Eli Andrew, 128  
Joo, Sean, 105, 166  
Julian, Marc W, 077  
Jung, Ae Kyong, 063  
Jung, Ji Yoon, 086  
Junker, Brian, 116  
Jurich, Daniel, 127
- Kaat, Aaron, 058  
Kadir, Latif, 069  
Kamata, Akihito, 045, 076, 125  
Kamei, Toshiko, 121  
Kang, Hyeon-Ah, 076  
Kang, Kyosin, 164  
Kang, Yulim, 069  
Kannan, Priya, 085, 132  
Kanopka, Klint, 063, 072, 147, 152  
Kaplan, David, 033, 060  
Kapoor, Radhika, 057, 147  
Kapoor, Shalini, 109  
Kara, Yusuf, 024, 045, 076  
Karamese, Hacer, 057, 103, 158  
Karimi, Honeiah, 055  
Karunakaran, Shiv, 123  
Karvonen, Meagan, 095

# PARTICIPANT INDEX

(Last Name, First Name, Session Number)

- Katz, Daniel, 123  
Kaye, Elaine, 127  
Kean, Sam, 021  
Kelberlau, Darin, 036  
Keller, Lisa, 045, 163  
Keller, Stefan Daniel, 114  
Kelley, Rebecca, 097  
Kelly, Justin, 106  
Kendall Brooks, Lauren D., 132  
Keng, Leslie, 031, 156  
Kennedy, Patrick, 140  
Kern, Justin L., 067, 088, 158  
Kerzabi, Emily, 033, 038, 047  
Ketan, 045, 084  
Ketterlin Geller, Leanne, 095  
Kevelson, Marisol, 046  
Kim, Eunbee, 063, 139  
Kim, Hahyeong, 057, 158  
Kim, Han Yi, 103  
Kim, Hyung Jin, 125  
Kim, Hyunjoo, 158  
Kim, Kyung Yong, 031, 089  
Kim, Seongeun, 031, 037  
Kim, Sohee, 069  
Kim, Stella, 031, 059, 105  
Kim, Sujin, 158  
Kim, Yeonwho, 031  
Kim, Yewon, 045, 050  
Kim, YoungKoung, 035  
Kim, Young Yee, 126, 134  
Kim, Yun-Kyung, 160  
Kincaid, Heath, 158  
King, Emma, 126  
Kingston, Neal, 027, 064  
Kinsey, Devon, 043  
Kirsch, Irwin, 133  
Kiser, Keith, 090  
Kislyonkova, Alexandra Levin, 034  
Kizilecec, Rene, 141  
Klugman, Emma M., 144  
Koehn, Hans Friedrich, 158  
Kolen, Michael, 057  
Kong, Lingchen, 043  
Koprowicz, Nick, 053  
Kowalski, Susan, 123  
Kozikowski, Andrzej, 037  
Kreitchmann, Rodrigo Schames, 067, 151  
Krist, Andrew, 051  
Krost, Kevin, 075, 143  
Kuhfeld, Megan, 066  
Kukea Shultz, Pohai, 057, 068, 092, 099, 106  
Kuklick, Livia, 057, 152  
Kumar, Lavanya Shravan, 166  
Kwako, Alexander, 055, 158  
Kwon, Muwon, 093  
Kwon, Tae Yeon, 063  
Kyllonen, Patrick Charles, 043, 047, 133, 141  
Kyung, Jungwon, 059  
LaFlair, Geoff, 135, 162  
Laforte, Erica, 069  
Lai, Hollis, 098, 125  
Lai, Viet, 023  
Laitusis, Cara C, 095  
Lakin, Joni, 051  
Lallmamode, Sheila, 082  
Lamm, Rik, 128  
Lan, Andrew, 032, 086  
Lan, Min, 069  
Landl, Erika, 097  
Lane, Suzanne, 133, 167  
Langi, Meredith, 123  
Lansing-Stoeffler, Kristin, 047  
Laster, Christina, 154  
La Torre, Deborah, 038  
Laufer, Alice, 051  
Law, Evelyn, 158  
Law, Nancy, 069  
Lawler, Mark, 062  
Lazarus, Sheryl, 075  
Lazenby, Katherine, 123  
Lazendic, Goran, 060  
Le, Anh Thu, 024  
Lee, Chansoon, 037, 087  
Lee, Dayeon, 050  
Lee, Dukjae, 045, 121  
Lee, Eunji, 050  
Lee, Guemin, 105  
Lee, Gyeong-Geon, 086  
Lee, Haeju, 022  
Lee, Hyeryung, 035, 073  
Lee, Jade Caines, 100, 112, 146, 154  
Lee, Jake, 086  
Lee, Ja-yoon, 045  
Lee, Jinsook, 141  
Lee, Meng-Hsun, 045  
Lee, Mina, 049  
Lee, Minho, 074  
Lee, Rachel, 045  
Lee, Selene Sunmin, 026  
Lee, Seulbi Bailey, 140  
Lee, Seunghyun, 083  
Lee, Soo, 126  
Lee, Sung-Hyuck, 096  
Lee, Sunhyoung, 050  
Lee, Won-Chan, 031, 059, 081, 125, 139, 158  
Lee, Youngjun, 037  
Lehman, Blair, 026, 066, 090, 157  
Lekwa, Adam, 024  
LeMahieu, Paul, 090  
Leng, Dihao, 081  
Leonard, Jana, 090  
Leventhal, Brian C, 127, 146  
Lewin, Nancy, 132  
Lewis, Crystal, 003  
Lewis, Daniel, 058, 107, 153, 160  
Lewis, Regina A., 036  
Li, Chen, 054, 161  
Li, Dongmei, 109  
Li, Grace, 106  
Li, Isaac, 079  
Li, Jie, 069  
Li, Jiehan, 063  
Li, Jingyang, 045  
Li, Xin (Grace), 057, 103  
Li, Xueming, 062  
Li, Zhushan Mandy, 029  
Li, Zikun, 105  
Liang, Min, 081  
Liang, Qianru, 069  
Liang, Xinya, 063  
Liao, Xiangyi, 149  
Liao, Yuqi, 033, 093  
Liaw, Yuan-Ling, 029, 087, 140  
Lim, Hwanggyu, 031, 096, 121, 158  
Lim, Sangdon, 016, 136  
Lim, Shuping, 158  
Lim, Youn Seon, 067, 142  
Lin, Ami, 140  
Lindner, Marlit Annalena, 057, 066, 152  
Lindsay, Constance, 105  
Ling, Guangming, 033, 090, 141  
Lions, Séverin, 077  
Lipien, Lodi, 125  
Lipscomb, Thomas J, 045  
Liu, Chen-Wei, 128  
Liu, Chunyan, 035, 057, 163  
Liu, Guangyun, 105  
Liu, Huan, 081  
Liu, Jinghua, 121  
Liu, Kristin, 160  
Liu, Lei, 090  
Liu, Liu, 042, 057  
Liu, Na, 023, 045  
Liu, Ou Lydia, 046, 090, 141  
Liu, Ren, 130  
Liu, Tuo, 120  
Liu, Xiang, 054  
Liu, Xiaoxiao, 045, 128  
Liu, Yan, 087  
Liu, Yang, 072, 093  
Liu, Yiqing, 147  
Liu, Yunting, 098

# PARTICIPANT INDEX

(Last Name, First Name, Session Number)

- Lizano, Cheryl, 005  
Lochbaum, Karen, 159  
Lohmann, Julian, 114  
Long, Yunyi, 147  
Longe, Brendan, 139  
Lorie, William A., 009, 145  
Lottridge, Susan, 055, 086, 104, 117, 129, 159  
Love, Quintin, 057  
Lovelace, Temple, 132, 157  
Lu, Xiangyu, 083, 152  
Lu, Yang, 049  
Lu, Yikai, 087  
Lu, Ying, 074  
Lu, Zhenqiu Laura, 045  
Lüdtke, Oliver, 135  
Luecht, Richard Melvin, 062, 089, 148  
Lugu, Benjamin Kweku, 022, 083, 158  
Luo, Hao, 105  
Luo, Jinwen, 150  
Luo, Rebekah, 158  
Luo, Yachen, 099  
Lyons, Susan, 097, 146, 167  
Lyu, Weicong, 033, 041
- Ma, Jing, 045, 081  
Ma, Mingjia, 069  
Ma, Wenchao, 004, 069, 083  
Ma, Ye, 006, 069  
Maas, Tricia, 048  
MacIntosh, Alexander, 043  
Madison, Matthew James, 014, 045, 093, 142  
Maeda, Hotaka, 042  
Maeda, Yukiko, 142  
Majerus, Mary, 091  
Makinde, Henry, 108  
Malatesta, Jaime, 049, 096, 121  
Man, Kaiwen, 007, 163  
Mang, Julia, 120  
Mansolf, Maxwell, 058  
Mao, Xia, 025, 035, 147  
Marder, Andrew, 059  
Mardones, Constanza, 027  
Marini, Jessica, 099  
Marion, Scott, 056, 064, 097, 133, 137, 167  
Mark-Worrell, Nicola, 119  
Martineau, Joseph A., 029, 085  
Martinez, Jose Felipe, 038, 112  
Matewere, Roshi, 069  
Matta, Tyler, 123  
Mazzullo, Elisabetta, 081, 098
- McCaffrey, Daniel, 054, 078  
McClarty, Katie, 064  
McClure, Kyla, 137  
McCormick, Carina, 077  
McMillan, James, 100  
McNichols, William, 086  
Mee, Janet, 049, 093, 127  
Melanlioglu, Deniz, 030  
Meng, Huijuan, 006  
Merkle, Edgar, 013  
Merriman, Jennifer, 145  
Mertens, Ute, 057, 081, 152  
Meyer, Jennifer, 081  
Meyer, Patrick, 062, 113  
Michaelides, Michalis, 115  
Michaels, Hillary, 040  
Michel, Rochelle, 112  
Middlestead, Andrew J., 036  
Middleton, Kyndra, 091  
Mikeska, Jamie, 081  
Mills, Christine, 053  
Min, Shangchao, 057  
Minchen, Nathan, 048, 118  
Minshew, Karen, 090  
Miranda, Alejandra, 158  
Miratrix, Luke Weisman, 018, 042, 144  
Mireles, Nicolas, 083  
Mislevy, Robert Joseph, 155  
Miyashiro, David, 090  
Mojoyinola, Mubarak O, 045, 081  
Molin, Lauren, 165  
Möller, Jens, 114  
Monroe, Scott, 111, 142, 149  
Montanano, Amber, 127  
Moore, John, 127  
Morell, Linda, 023  
Morris, Armani Michia, 057  
Morris, Sarah Ruth, 081  
Morrison, Carol, 025  
Morrison, Kristin M., 103  
Mosely, Brein, 144  
Moses, Tim, 020, 031, 035  
Moteane, Malitsitso, 108  
Mulcaire, Phoebe, 135  
Mulvihill, Megan M, 110  
Mundhenk, Kimberly, 085  
Muñoz-Najar Galvez, Sebastian, 144  
Muntean, William J, 158, 166  
Murphy, Daniel, 057, 094  
Myers, Aaron, 124  
Myers, Matthew, 118
- Nagasawa, Mirai, 034  
Nash, Brooke, 091, 095
- Nason, Kelsey, 045  
Nava, Imelda, 038  
Naveiras, Matthew David, 152  
Neiman, Samantha, 164  
Nemeth, Yvette M, 069  
Nesbitt, Jaylin N., 085, 108, 132, 159  
Nese, Joseph F. T., 045, 076  
Newstadt, Michelle, 047  
Nguyen, Cuc Thi Kim, 034  
Nguyen, Tram-Anh Tran, 045  
Nguyen Khoa, Lan Anh, 121, 143  
Nickodem, Kyle, 026, 128, 158  
Nicola, Matthew, 144  
Nicola, Tara P., 144  
Noh, Jaehee, 045  
Novack, Miriam, 058  
Nowinski, Cindy, 058  
Nugba, Regina Mawusi, 081  
Nydick, Steven, 105, 147
- Ober, Teresa, 026, 039, 043, 046  
Odom, Kimberly, 069  
O'Donnell, Francis, 025, 057, 156  
Ogut, Burhan, 152  
Oh, Hyeon-Joo, 020, 057, 069, 105  
Oh, Hyunjee, 067  
Okay, Serife, 143  
Olivarez-Durden, Sandra, 069  
Olivera Aguilar, Margarita, 164  
Oliveri, Maria-Elena, 132, 167  
Olson, Jacinta, 045  
Oluwalana, Olasumbo, 046  
O'Neil, Timothy, 094  
Ong, Thai Quang, 093  
Oppenzato, Colleen, 123  
Oppong, Frank Agyemang, 081  
Oquendo, Jodine, 006  
Organisciak, Peter, 084  
Ormerod, Chris, 032, 117, 129, 158  
Ortega, Gabriel, 077  
Osses-Vargas, Alejandra, 060  
Ouyang, Wenli, 093  
Ozer Ozkan, Yesim, 151
- Pace, Lillian, 164  
Page, Jane, 121  
Pan, Qianqian, 069  
Pan, Tianshu, 069  
Pan, Yanfang, 083  
Pan, Yiqin, 007  
Paraboni, Ivandre, 084  
Pardos, Zachary A., 098  
Paris, Joseph, 046  
Park, John, 045  
Park, Jung Yeon, 105, 158

# PARTICIPANT INDEX

(Last Name, First Name, Session Number)

- Park, Seohee, 139  
Park, Yena, 162  
Park, Yooyoung, 025  
Parker, Emily Hope, 119  
Parks, Charles, 105  
Parra-Martinez, Fabio Andres, 081  
Partchev, Ivailo, 012  
Patel, Milan, 117  
Patel, Nirmal, 093  
Patelis, Thanos, 051, 064, 145  
Patterson, Jim, 051  
Patterson, John D., 084  
Patton, Elizabeth Adele, 077  
Patz, Richard, 010, 058, 073  
Pearson, Jami-Jon, 040  
Peasley, Donald, 036  
Pei, Francesca, 061  
Pellegrino, James, 056, 094, 107, 123  
Peng, Fang, 069  
Peng, Luyao, 105  
Perez, Leticia, 038  
Perie, Marianne, 094, 119, 132, 153  
Petscher, Yaacov, 061  
Pham, Duy N., 023, 162  
Phillip, kristy, 119  
Piacentini, Mario, 084  
Pila, Sarah, 058  
Pinsonneault, Laura, 009  
Pinto, Wallace Nascimento, 125, 150  
Pitoniak, Mary, 133  
Pittman, Jocelynn, 069, 136  
Poe, Mya, 146, 167  
Por, Han-Hui, 078  
Portilla, Ximena, 131  
Potgieter, Cornelis, 076  
Powers, Sonya, 030, 039, 048, 069, 124  
Prasad, Seema, 074  
Pratt, Amy, 061  
Prest, June, 069  
Prendergast, Caroline, 156  
Prewitt, Nicole, 069  
Primi, Ricardo, 084  
Proctor, Thomas, 089, 148  
Pugsley, John, 156
- Qiao, Xin, 024, 057, 076, 079  
Qin, Qi, 103  
Qiu, Chen, 109  
Quan, Yale, 079  
Quanbeck, Mari, 075  
Quesen, Sarah, 094, 159  
Quranda, Milton, 093
- Rabinowitz, Stanley N., 026, 040  
Rabkin, Sasha, 102
- Rahimi, Ahmad, 047  
Raimondi, Matt, 113  
Rajeb, Mehdi, 051  
Ramazan, Onur, 025  
Ramnarine, Nirmala, 119  
Ramsawak-Jodha, Nalini, 119  
Randall, Jennifer, 056, 146  
Raymond, Krystina, 045  
Razavi, Pooya, 030  
Reckase, Mark, 002, 022, 126  
Redman, Elizabeth, 001  
Reise, Steve, 081  
Re-McGlynn, Leta S., 127  
Ren, He, 057  
Ren, Shixiu, 067  
Resanovich, Mary, 062  
Reshetar, Rosemary, 101  
Rhodes, Sally, 069  
Richardson, Julie, 062  
Richardson, Megan, 144  
Rijmen, Frank, 094  
Rikoon, Sam, 164  
Rios, Joseph A., 081  
Rios, Oscar, 081  
Roberts, James S., 045  
Roberts, Jeremy, 001  
Robin, Frederic, 026  
Robitaille, Jennifer, 119  
Robitzsch, Alexander, 135  
Rocha, Horacio A., 081  
Rodriguez, Michael C., 010, 026, 052, 068, 099, 128, 158  
Roduta Roberts, Mary, 146  
Roeters-Solano, Heather, 159  
Rome, Logan, 077, 118  
Rosales de Veliz, Leslie Vanessa, 082  
Rose, Mark, 158  
Rosen, Yigal, 047, 131  
Ross, Linette P., 127, 156  
Roussos, Louis, 103  
Rozman, Mojca, 029  
Rozunick, Chris, 124  
Ruan, Chunyi, 054  
Rubel, Jennifer, 069  
Rubright, Jonathan, 130  
Runge, Andrew, 032  
Runyon, Christopher, 093  
Rupp, Andre, 009  
Rushkin, Ilia, 047  
Russell, Michael, 146  
Rutkowski, Leslie, 084, 122, 129, 143  
Rutstein, Daisy Wise, 107  
Ruzek, Erik, 123  
Ryoo, Ji Hoon, 047
- Sabatini, John, 024  
Sabbaghi, Arman, 151  
Sabbah, Jenin, 146  
Sachdeva, Ruchi, 142  
Saklani, Dinesh Prasad, 078  
Salas, Jorge, 152  
Saldivia, Luis, 033, 066, 113  
Saleh, Yustina, 090  
Samson, Sally, 106  
Sanchez, Edgar, 074  
Sanders, Elizabeth A., 057  
Sanders, Sara, 069  
Sarac, Merve, 103, 120  
Satchu, Salma, 156  
Satkus, Paulius, 030, 049, 096, 124  
Sato, Edynn, 057, 075, 103, 106  
Savitz-Romer, Mandy, 144  
Sayin, Ayfer, 030, 042  
Scarlato, Alex, 086  
Schaller, Nils-Jonathan, 081  
Schellman, Madeline, 142  
Schemel, Lynn, 085  
Schiller, Ronja, 081  
Schmidgall, Jonathan, 165  
Schmidt, Susanne, 027  
Schneider, Christina, 028, 069, 103, 136  
Schneider, Wei, 049  
Schnittka Hoskins, Jessica Elizabeth, 119  
Schultz, Matthew, 101  
Schwantner, Ursula, 060  
Schwappach, Bethany, 090  
Schweig, Jonathan, 164  
Scott-Smith, Chandler, 156  
Setzer, Carl, 057  
Shan, Xintong, 067  
Sharairi, Sid, 069  
Sharpnack, James, 135  
Shaw, Emily, 099  
Shaw, Timothy, 085  
Shelley, Kyna J., 045  
Shelton, Stephanie, 069  
Shen, Yawei, 039, 049  
Shenavai, Kevin, 027  
Shepard, Lorrie, 056, 137  
Sherer, David, 090  
Shermis, Mark D., 055, 117  
Shi, Dingjing, 135  
Shi, Qingzhou, 051  
Shih, Ching-Lin, 087  
Shimizu, Yuki, 105  
Shin, David, 079  
Shin, Hyesung, 105  
Shin, Hyo Jeong, 047  
Shin, Jinnie, 043, 125, 150

# PARTICIPANT INDEX

(Last Name, First Name, Session Number)

- Shipan, Rebecca, 134  
Shivraj, Pooja, 158  
Shotts, Bruce L., 057  
Siebert, Julian Maximilian, 061  
Siegel, Danielle, 111  
Siefert, Paula, 058  
Sikali, Emmanuel, 033, 093  
Sinclair, Andrea, 028  
Sinharay, Sandip, 044, 065, 105, 157  
Sireci, Stephen G, 110, 132, 146, 162  
Sirganci, Gozde, 125  
Skoblow, Hanamori, 015, 072  
Slomp, David, 167  
Smith, Charity, 040  
Smith, Digory, 086  
Smith, Jessalyn, 165  
Smith, Larissa, 049  
Smith, Mireya Carmen-Martinez, 158  
Smith, Nate, 163  
Soland, James, 066, 130, 149  
Solano-Flores, Guillermo, 056, 112, 132, 155, 167  
Someshwar, Shonai, 050  
Somsong, Sarunya, 125  
Song, Dan, 125  
Song, Naiqing, 151  
Soni, Parth Tusharbai, 078  
Sotelo, Jose de Jesus, 046  
Soto, Consuelo Andrea, 077  
Southerland, Joshua, 129  
Soyoye, Olushola Olufemi, 081  
Sparks, Anthony, 158  
Sparks, Jesse R., 026, 039  
Spikings, Desiree, 062  
Stark, Stephen, 166  
Steed, Teneka, 108  
Steedle, Jeffrey, 047, 109  
Stethen, Kelley, 025  
Stewart, Sheridan, 144  
Stickney, Eric, 080  
Still, Cathryn, 160  
Stopek, Joshua, 101  
Strietholt, Rolf, 140  
Student, Sanford R., 048, 080, 118, 147, 149  
Su, YuLan, 024, 057, 081  
Suarez-Alvarez, Javier, 132  
Subedi, Dipendra, 069  
Subhiyah, Raja G., 057  
Sudheesh, Athul, 039  
Suh, Hongwook, 031, 103  
Suh, Yon Soo, 083, 123  
Suk, Youmi, 166  
Sul, David, 146  
Suleiman, Mahmoud, 156  
Sun, Huaping, 037  
Sun, Ting, 057  
Sun, Xiaojian, 151  
Sung, Youjin, 072, 093  
Supovitz, Jonathan, 105  
Sussman, Joshua, 023, 091  
Svetina Valdivia, Dubravka, 166  
Swain, Matthew, 081  
Sweeney, Sandra, 103  
Swygert, Kimberly, 127  
Talreja, Vinita, 006, 093  
Tan, Bin, 081, 098  
Tan, Cheng Yong, 069  
Tan, Ling, 060  
Tan, Yanyan, 105  
Tang, Alexander, 125  
Tang, Nai-En, 057, 158  
Tang, Xiuxiu, 105  
Tao, Sisi, 069  
Taylor, Catherine, 091, 146, 155  
Taylor, CJ, 006  
Taylor, Darius D., 146  
Templin, Jonathan, 045, 063, 081  
Tenison, Caitlin, 039  
Thacker, Arthur, 028, 113  
Thier, Ben, 131  
Thissen, David, 065, 134  
Thompson, W. Jake, 011, 083  
Thurlow, Martha, 160  
Tian, Wei, 069  
Todd, Jessica Andrews, 033, 038  
Tong, Ye, 010, 112, 127  
Toton, Sarah Linnea, 007  
Toutkoushian, Emily Karen, 037  
Traynor, Anne, 099, 111  
Trüb, Ruth, 114  
Tsai, Chia-Lin, 128, 158  
Tsey, Enoch, 081  
Tu, Naidan, 166  
Tyack, Lillian, 086, 143  
Ulitzsch, Esther, 072, 135  
Valdivia Medinaceli, Montserrat B., 068, 122, 166  
Van der Ark, Andries, 047  
Vanlwaarden, Adam, 048  
Van Orman, Dustin, 100, 130, 146  
van Rijn, Peter, 005, 078  
van Schijndel, Tessa, 047  
Vasu, Aveline A., 158  
Vele, Veronica, 098  
Ventura, Claudia J, 081  
Verkuilen, Jay, 081  
Vispoel, Walter, 035, 073  
Vivas Corrales, Pablo, 158  
Vo, Thao, 146  
Vo, Yen, 081  
Von Davier, Alina A., 017, 047, 092, 104, 116, 162  
Von Davier, Matthias, 081, 086, 104, 111, 122, 143  
Wackerle-Hollman, Alisha, 131  
Wainer, Howard, 065  
Walker, Marcus, 042, 105  
Walker, Michael E., 046, 068, 070  
Wall, Nathan, 019  
Wan, Lei, 165  
Wan, Siyu, 049  
Wang, Aijun, 057  
Wang, Bowen, 158  
Wang, Changjiang, 069  
Wang, Chuang, 057  
Wang, Chun, 057, 079  
Wang, Chunxin, 069  
Wang, Dongwei, 057, 079  
Wang, Hongling, 109  
Wang, Jue, 114  
Wang, Kuo, 045  
Wang, Min, 089  
Wang, Nixi, 158  
Wang, Shiyu, 027, 039, 083, 105, 117, 142  
Wang, Shuang, 069  
Wang, Songtao, 045  
Wang, Xi, 045  
Wang, Xiyu, 142  
Wang, Xuan, 069  
Wang, Yang Caroline, 119  
Wang, Yu, 105  
Wang, Yuan, 090  
Wang, Zhen, 057  
Wang, Zhuoran, 158, 166  
Warner, Zachary, 036  
Waters, Charlotte, 060  
Watts, Ashley, 072  
Webb, Blue, 033, 093, 122  
Weeks, Jonathan, 071, 103, 138  
Wei, Hsin-Ro, 057, 105  
Weiner, Drew, 101  
Weiner, John, 047  
Weiss, David, 140  
Welch, Catherine, 081  
Wells, Craig, 059, 079  
Wendel, George, 158  
Westrick, Paul, 099  
Wheeler, Jordan M., 114, 117, 142  
White, Lauren, 057, 159  
Whitfield, Erik, 048  
Whitmer, John, 032  
Whitney, Sarah, 062

# PARTICIPANT INDEX

(Last Name, First Name, Session Number)

Williams, Kevin, 043  
Williams, Natasha Jayne, 038  
Williamson, Shelby, 025  
Willse, Jada, 156  
Wilmurth, Gina, 062  
Wilson, Joshua, 055, 117  
Wilson, Mark, 023, 073, 094, 110  
Wilson, Nicole, 127  
Wind, Stefanie A., 022, 034, 051,  
073, 128, 158  
Winter, Sonja D, 015, 072  
Wise, Steven, 066  
Withycombe, Adam, 062  
Wolf, Benjamin, 135  
Wolfe, Edward W, 045, 120  
Wolff, Sarah, 163  
Wollack, James, 041, 103, 115  
Wong, Yun Leng, 140  
Wongvorachan, Tarid, 057, 081  
Woo, Ada, 130  
Woo, Yejin, 050  
Woodhead, Simon, 086  
Woods, Scott, 109  
Woolf, Sherri, 129  
Workman, Trent, 036  
Worthington, Michelle, 033, 066  
Wright, Daniel B., 163  
Wu, Hao, 081  
Wu, Lizzy, 072  
Wu, Tong, 069  
Wu, Yi-Fang, 019, 039  
Wu, Zebing, 049  
Wuyts, Celine, 084  
Wylie, E. Caroline, 085  
Wyse, Adam E, 076, 139  
  
Xia, Yan, 105  
Xiao, Xingyao, 030, 034, 073  
Xie, Qing, 008  
Xin, Tao, 069, 151  
Xiong, Jiawei, 027, 049  
Xiong, Xinhui Maggie, 165  
Xiong, Yimei, 111  
Xu, Xin, 067  
Xu, Yangmeng, 073  
Xue, Mingfeng, 023, 076  
  
Yan, Duanli, 017  
Yancey, Kevin, 135, 162  
Yaneva, Victoria, 093, 104, 162  
Yang, Huidi, 046  
Yang, Ji Seung, 093  
Yang, Xiangdong, 105  
Yang, Xiaotong, 057  
Yang, Yanyun, 121  
Yao, Lihua, 058, 074  
  
Yavuz, Sinan, 033, 093, 111, 134, 143  
Ye, Daisy, 077  
Ye, Shaojie, 033  
Yen, Shu Jing, 158  
Yi, Karen, 093  
Yildirim-Erbasli, Seyma N., 066  
Yin, Yue, 083  
Yoo, Hanwook, 020  
Yoon, Hyejin, 105  
Yu, Haoyang, 081  
Yuan, Ye, 114  
  
Zapata-Rivera, Diego, 026, 048, 090  
Zavitkovsky, Paul, 124  
Zechner, Klaus, 117  
Zegers, Mónica, 061  
Zehner, Fabian, 120  
Zeng, Biao, 105  
Zeng, Ji, 029, 165  
Zenisky, April, 132  
Zhai, Xiaoming, 027, 086  
Zhan, Peida, 163  
Zhang, Anna Yinqi, 045  
Zhang, Bo, 105  
Zhang, Danhui, 067  
Zhang, Jin, 069  
Zhang, Jinming, 121  
Zhang, Lijin, 072  
Zhang, Litong, 077  
Zhang, Mengyao, 069, 101  
Zhang, Mingqin, 008, 069  
Zhang, Mo, 054, 163  
Zhang, Oliver, 148  
Zhang, Susu, 083, 147  
Zhang, Susu, 147  
Zhang, Ting, 033, 093, 111, 122  
Zhang, Xiao, 077  
Zhang, Yichi, 045  
Zhang, Yu, 057  
Zhang, Yuan, 126  
Zhang, Yuxiao, 069, 105  
Zhang, Zhonghua, 034  
Zheng, Xiaying, 126, 134  
Zheng, Yi, 147  
Zhong, Xiaoting, 081  
Zhong, Yang, 161  
Zhou, Xinchang, 057, 105  
Zhou, Xuechun, 105  
Zhu, Danqi, 103, 115  
Zlatkin-Troitschanskaia, Olga, 027,  
051  
Zopluoglu, Cengiz, 061  
Zou, Chunrui, 045  
Zou, Tongtong, 124  
Zou, Yuqing, 045  
Zoucha, James, 057

## PARTICIPANT EMAIL

(Last Name, First Name, Affiliation, Email)

Abeyasinghe, Bhashithe, American Institutes for Research, babeysinghe@air.org  
Abulela, Mohammed, University of Minnesota, mhady001@umn.edu  
Ackerman, Terry, University of Iowa, terryackerman2@gmail.com  
Adam, Hugh, Northwestern University, hubert.adam@northwestern.edu  
Adams, Elizabeth L., American Institutes For Research, eadams@air.org  
Adams, Grant, American Institutes for Research, gadams@air.org  
Adams, Raymond J., Australian Council for Educational Research, ray.adams@acer.org  
Addis, Kerbie, NBME, kaddis@nbme.org  
Adkins, Deborah, Accessible Teaching, Learning, and Assessment Systems ATLAS, dadkins386@gmail.com  
Alahmadi, Sarah, James Madison University, alahmasi@jmu.edu  
Alarcon, Lorena, University of Illinois at Urbana-Champaign, lorenaa2@illinois.edu  
Albano, Tony, University of California, Davis, adalbano@ucdavis.edu  
Aldrup, Karen, IPN - Leibniz Institute for Science and Mathematics Education, aldrup@leibniz-ipn.de  
Ali, Usama, ETS, uali@ets.org  
Almond, Russell G., Florida State University, ralmond@fsu.edu  
Alvero, Aaron Jordan, aaronjalvero@gmail.com  
Alves, Cecilia, Medical Council of Canada, calves@mcc.ca  
Andersen, Nico, DIPF | Leibniz Institute for Research and Information in Education, Andersen.Nico@dipf.de  
Andrade-Lotero, Luis Alejandro, PEARSON, alejandro.andradelotero@pearson.com  
Angehr, Emily, College Board, eangehr@collegeboard.org  
Anguiano-Carrasco, Cristina, ACT, cristina.anguiano-carrasco@act.org  
Ankomah, Francis, Ohio University, fa013522@ohio.edu  
Antal, Judit, College Board, jantal@collegeboard.org  
Araneda, Sergio, Caveon, sergio.araneda@caveon.com  
Armoush, Nour, University of Alberta, armoush@ualberta.ca  
Armstrong, Maria, Association of Latino Administrators and Superintendents, maria@alasedu.org  
Arslan, Burcu, ETS Global B.V., barslan@ets.org  
Arthur, Ann, ACT, ann.arthur@act.org  
Arthur, David, Purdue University Statistics, arthur22@purdue.edu  
Atitsogbey, Desmond, The Ohio State University, atitsogbey.1@buckeyemail.osu.edu  
Avvisati, Francesco, OECD, francesco.avvisati@gmail.com  
Awwal, Nafisa, University of Melbourne, n.awwal@unimelb.edu.au  
Ayers-Wright, Elizabeth, Cambium Assessment, Inc., eaa18sb@gmail.com  
Azim, Farhan, The University of Melbourne, farhan.azim@unimelb.edu.au  
Badrinarayan, Aneasha, Learning Policy Institute, abadrinarayan@learningpolicyinstitute.org  
Bagui, Lalaine, The University of the Philippines, l.bagui@actrc.org  
Bahr, Jan Luca, bahr@leibniz-ipn.de  
Bai, Yifan, American Institutes for Research, ybai@air.org  
Bailey, Alison, UCLA, abailey@gseis.ucla.edu  
Bailey, Paul, American Institutes for Research, pbailey@air.org  
Bain, Catherine, University of Oklahoma, cbain1@ou.edu  
Baker, Doris Luft, The University of Texas at Austin, doris.baker@austin.utexas.edu  
Bakken, Sara, BrainPop, sarab@brainpop.com  
Balderama, Earvin, California State University, Fresno, earvin@csufresno.edu  
Baldwin, Miles Somay, Harriton, miles.baldwin07@gmail.com  
Baldwin, Peter, National Board of Medical Examiners, pbaldwin@nbme.org  
Bamat, David, American Institutes for Research, dbamat@air.org  
Banjanovic, Erin, Curriculum Associates, erin.banjanovic@gmail.com  
Bao, Yu, James Madison University, bao2yx@jmu.edu  
Barber, Justin O., Pearson, justin.barber@pearson.com  
Baron, Helen, Saville & Holdsworth LTD, helen.baron@shlgroup.com  
Barry, Carol L., American Board of Surgery, cbarry@absurgery.org  
Basaraba, Deni, Amplify Education, dbasaraba@amplify.com  
Bates, Shaun, Missouri Department of Elementary and Secondary Education, shaun.bates@dese.mo.gov  
Beach, Sarah, University of Virginia, seb3mr@virginia.edu  
Bean, Jerry, The Ohio State University, bean.8@osu.edu

## PARTICIPANT EMAIL

(Last Name, First Name, Affiliation, Email)

Beard, Jonathan, College Board, jonathan.j.beard@gmail.com  
Beaty, Roger E., Penn State University, rebeaty@psu.edu  
Becker, Kirk, PEARSON, kirk.becker@pearson.com  
Becker, Lee, Pearson, lee.becker@pearson.com  
Bediwy, Ahmed, The University of Iowa, ahmed-bediwy@uiowa.edu  
Beigman Klebanov, Beata, ETS, bbeigmanklebanov@ets.org  
Beiting-Parrish, Magdalen, Federation of American Scientists, magdalen.beiting@gmail.com  
Belov, Dmitry I., Law School Admission Council, dbelov@lsac.org  
Belur, Vinetha, vbelur@ets.org, vbelur@ets.org  
Belzak, Will, Duolingo, wbelzak@duolingo.com  
Benedek, Mathias, University of Graz, mathias.benedek@uni-graz.at  
Bennett, Randy, ETS, rbennett@ets.org  
Benson, Julie, Cambium Assessment, Inc., julie.benson@cambiumassessment.com  
Berezner, Alla, Australian Council for Educational Research, Alla.Berezner@acer.org  
Bergner, Yoav, New York University, yoav.bergner@nyu.edu  
Berry, Yufeng, Minnesota Department of Education, yufeng.berry@state.mn.us  
Bertling, Jonas, ETS, jbertling@ets.org  
Beswick, Bruce, The University of Melbourne, bruceab@unimelb.edu.au  
Betebenner, Damian, National Center for the Improvement of Educational Assessment, dbetebenner@nciea.org  
Betts, Joe, NCSBN, jbetts@ncsbn.org  
Betzberger, Justin, UCLA, betzberger@g.ucla.edu  
Bhaduri, Indrani, National Council of Educational Research and Training, India, indranibhaduri@gmail.com  
Bhaktha, Nivedita, GESIS, nivedita.bhaktha@gesis.org  
Bhandari, Shreya, University of California, Berkeley, shreya.bhandari@berkeley.edu  
Biancarosa, Gina, University Of Oregon, Ginab@uoregon.edu  
Bilir, Mustafa Kuzey, PEARSON, kuzey.bilir@pearson.com  
Binici, Salih, Florida Department of Education, salih.binici@fldoe.org  
Bishop, Kyoungwon Lee, WIDA at University of Wisconsin Madison, kei.bishop2@gmail.com  
Bissel, Edie, NBME, ebissel@nbme.org  
Blanco, María Paz, Center for Advanced Research in Education (FB0003), Institute of Education, Universidad de Chile, Santiago, Chile, mariapaz.blanco@ciae.uchile.cl  
Blankenship, Charles, American Institutes for Research, cblankenship@air.org  
Block, Jared, UCLA, jblock@psych.ucla.edu  
Bohrnstedt, George William, AIR, GBohrnstedt@air.org  
Bolender, Brad, Finetune, brad.bolender@gmail.com  
Bolt, Daniel, University of Wisconsin, Madison, dmbolt@wisc.edu  
Bonifay, Wes, University of Missouri, bonifayw@missouri.edu  
Boothe-Corley, Rebekah, Edmentum, Inc., rebekah.boothe-corley@edmentum.com  
Bouy, Alicia, Pearson, alicia.bouy@pearson.com  
Boykin, Allison, Cambium Learning Group, allison.boykin@cambiumassessment.com  
Bradshaw, Laine, Pearson, lainebradshaw@gmail.com  
Braithwaite, Reina, The University of the West Indies, reinabrathwaite@my.uwi.edu  
Brandt, Chris, National Center for the Improvement of Educational Assessment, cbrandt@nciea.org  
Brantley, Wyman, ETS, wbrantley@ets.org  
Braun-Bharti, Nina James, Germantown Friends School, nbraumbharti25@germantownfriends.org  
Braun, Henry, Boston College, braunh@bc.edu  
Brennan, Robert, University of Iowa, robert-brennan@uiowa.edu  
Brice, Amanda, Curriculum Associates, abrace@cainc.com  
Briggs, Derek Christian, University of Colorado Boulder, derek.briggs@colorado.edu  
Briscoe, Michael, Ed Reports, michael.briscoe@gmail.com  
Broer, Markus, American Institutes for Research, markus.broer@gmail.com  
Brookhart, Susan, Duquesne University, suebrookhart@gmail.com  
Brown, Emily Anne, emilybrown8@my.unt.edu  
Brown, Richard, rich@westcoastanalytics.com  
Brunetti, Matthew, WestEd, mbrunet@wested.org  
Bryant, Damon, dbryant@aastest.com

## PARTICIPANT EMAIL

(Last Name, First Name, Affiliation, Email)

Bryk, Anthony, Carnegie Foundation, Abryk@carnegiefoundation.org  
Buckendahl, Chad W., ACS Ventures, LLC, cbuckendahl@acsventures.com  
Buczek, Matthew, Aperture Education, mbuczek@apertureed.com  
Bulut, Hatice Cigdem, Northern Alberta Institute of Technology, hcyavuz@gmail.com  
Bulut, Okan, University of Alberta, bulut@ualberta.ca  
Bunch, Michael Brannen, Measurement Incorporated, mbunch@measinc.com  
Burkhardt, Amy, Cambium Assessment, Inc., burkhardt.amy@gmail.com  
Burrus, Jeremy, ACT, jeremy.burrus@act.org  
Burstein, Jill, Duolingo, jill@duolingo.com  
Buzick, Heather, ETS, hbuzick@ets.org  
Caffrey-Maffei, Lucy, Stanford University, l.caffreymaffei@stanford.edu  
Cai, Li, UCLA, lcai@ucla.edu  
Cai, Qing, University of California, Berkeley, qingcai@berkeley.edu  
Caliço, Tiago A., Organization for Economic Cooperation and Development, tcalico@proton.me  
Camara, Wayne J., LSAC, waynecamara@gmail.com  
Cancado, Luciana, Curriculum Associates, lcancado@cainc.com  
Cao, Yichong, University of Iowa, yichong-cao@uiowa.edu  
Cardwell, Ramsey, Duolingo, ramsey@duolingo.com  
Carling, Dona, College Board, dcarling@collegeboard.org  
Carney, Michele, michelecarney@boisestate.edu  
Casabianca, Jodi M., ETS, jcasabianca@ets.org  
Castaneda, Ruben, College Board, z5rcastaneda@gmail.com  
Catts, Hugh, Florida State University, hugh.catts@cci.fsu.edu  
Chakraborty, Roti, Georgia State University, rchakraborty3@student.gsu.edu  
Chalhoub-Deville, Micheline, University of North Carolina, GRE, chalhoub-deville@uncg.edu  
Cham, Heining, Fordham University, hcham@fordham.edu  
Chang, Hua-Hua, Purdue University, chang606@purdue.edu  
Chang, Wan Chen, Cambium Assessment, claire.chang@cambiumassessment.com  
Chao, Hsiu-Yi, Soochow University, psyhyc@scu.edu.tw  
Chatzidimitriou, Despoina, University of Cyprus, chatzidimitriou.despoina@ucy.ac.cy  
Chatzimichail, Myria, University of Cyprus, chatzimichael.myria@ucy.ac.cy  
Chawala, Kamal, University of Delaware, kamalc@udel.edu  
Chen, Becky, University of Toronto, xi.chen.bumgardner@utoronto.ca  
Chen, Dandan, Pearson, dandan.c.chen@gmail.com  
Chen, Guanyu, The University of British Columbia, chenguanyu.ubc@gmail.com  
Chen, Hong, The University of Iowa, hchen102@uiowa.edu  
Chen, Jianshen, College Board, cachem@collegeboard.org  
Chen, Jihang, Boston College, jihang@bc.edu  
Chen, Jing, Cambium Assessment, jingchen2022@gmail.com  
Chen, Jinsong, The University of Hong Kong, jinsong.chen@live.com  
Chen, Jun-Hong, National Cheng Kung University, psyjhc@gs.ncku.edu.tw  
Chen, Ling, Columbia University, lc3521@columbia.edu  
Chen, Ping, Beijing Normal University, pchen@bnu.edu.cn  
Chen, Qipeng, University of Alabama, qipengchen\_psy@outlook.com  
Chen, Tingting, University of Iowa, tingting-chen@uiowa.edu  
Chen, Xiaowen, George Mason University, xchen29@gmu.edu  
Chen, Ying, ychen406@uic.edu  
Cheng, Ming-Min, Purdue University, cheng637@purdue.edu  
Cheng, Yiling, Kaohsiung Medical University, yilingcheng@kmu.edu.tw  
Cheng, Ying, University of Notre Dame, ycheng4@nd.edu  
Cheruvu, Abhinav, Texas A&M University, acheruvu@air.org  
Chiu, Chia-Yi, University of Minnesota, cchiu@umn.edu  
Cho, Sun-Joo, Peabody College of Vanderbilt, sj.cho@vanderbilt.edu  
Cho, Youngmi, Riverside Insights, youngmi.cho@riversideinsights.com  
Cho, YoungWoo, ACT, youngwoo.cho@act.org  
Cho-Baker, sugene, ETS, SCHOBAKER@ets.org

## PARTICIPANT EMAIL

(Last Name, First Name, Affiliation, Email)

Choe, Edison, Renaissance, edison.choe@renaissance.com  
Choi, Hanna, EWAH Womans University, chje1834@ewhain.net  
Choi, Hye-Jeong, Human Resources Research Organization, hchoi@humro.org  
Choi, Ikkyu, ETS, ichoi001@ets.org  
Choi, Jaehwa, George Washington University, jaechoi@gwu.edu  
Choi, Jeongwon, Vanderbilt University, jeongwon.choi@vanderbilt.edu  
Choi, Jinah, Edmentum, Inc., Jinah.Choi@edmentum.com  
Choi, Jinnie, Savvas Learning Company, jinnie.choi@gmail.com  
Choi, Seung W., University of Texas at Austin, schoi@austin.utexas.edu  
Choi, Youn-Jeng, EWAH Womans University, younjengchoi@ewha.ac.kr  
Chu, Yuen, The Pharmacy Examining Board of Canada, yuenchu@pebc.ca  
Chung, Gregory, UCLA CRESST, chung@cresst.org  
Cipriano, Christina, Education Collaboratory at Yale, christina.cipriano@yale.edu  
Circi, Ruhan, American Institutes for Research, rcirci@air.org  
Cizek, Gregory, University of North Carolina-C, cizek@unc.edu  
Clark, Amy, ATLAS: University of Kansas, akclark@ku.edu  
Clauser, Amanda, National Board of Medical Examiners, aclouser@nbme.org  
Clauser, Brian, National Board of Medical Examiners, bclouser@nbme.org  
Clauser, Jerome, American Board of Internal Medicine (ABIM), jclouser@abim.org  
Claxton, Patricia, patriciaclaxton27@yahoo.com  
Clelland, Ashley, University of Alberta, adc@ualberta.ca  
Cloney, Daniel, Australian Council for Educational Research, Dan.Cloney@acer.org  
Cohen, Allan, University of Georgia, acohen@uga.edu  
Collins, Chris, Cajon Valley Union School District, collinsch@cajonvalley.net  
Connolly, Helena, helena.lunts@gmail.com  
Cook, Linda, ETS, lcook@ets.org  
Corrigan, Julie, Concordia University, julie.corrigan@concordia.ca  
Cotterell, Michael, University of Georgia, mepcott@uga.edu  
Cox, Olivia, University of Colorado Boulder, olivia.cox@colorado.edu  
Crawford, Kristen, Kristen.S.Crawford@doe.nh.gov  
Crespo Cruz, Eduardo Javier, UMASS Amherst, ecrespocruz@umass.edu  
Cui, Mengyao, Cambium Assessment, Inc., mengyao.cui@cambiumassessment.com  
Cui, Weiwei, College Board, wcui@collegeboard.org  
Cui, Wenju, ETS, wcui@ets.org  
Cui, Ying, University of Alberta, yc@ualberta.ca  
Culpepper, Steven, University of Illinois at Urbana-Champaign, sculpepp@illinois.edu  
Dadey, Nathan, Center for Assessment, ndadey@nciae.org  
Dai, Shenghai, Washington State University, s.dai@wsu.edu  
Dailey, David, Dailey Data Group, david@daileydatagroup.com  
Daisher, Ted, Curriculum Associates, t3daisher@gmail.com  
Daley, Nola, ACT, Nola.Daley@act.org  
Dallas, Andrew, National Commission on Certification of Physician Assistants (NCCPA), drewd@nccpa.net  
Dartnell, Pablo, Center for Advanced Research in Education (FB0003), Institute of Education; Center for Mathematical Modeling (ACE210010 and FB210005); and Department of Mathematical Engineering; all from Universidad de Chile, Santiago, Chile, dartnell@ciae.uchile.cl  
Davalos, Eduardo, Vanderbilt University, eduardo.davalos.anaya@vanderbilt.edu  
Davidson, Anne H, EdMetric LLC, anne.davidson@edmetric.com  
Davis, Jennifer, Amazon Web Services (AWS), jedavisr@amazon.com  
Davis, Laurie, Curriculum Associates, laurie@davistx.com  
Davis-Becker, Susan, ACS Ventures, LLC, sdavisbecker@acsventures.com  
Davison, Mark, University of Minnesota, mld@umn.edu  
D'Brot, Juan Manuel, National Center for the Improvement of Educational Assessment, jdbrot@nciae.org  
Deane, Paul, ETS, pdeane@ets.org  
Debelak, Rudolf, University of Zurich, rudolf.debelak@psychologie.uzh.ch  
DeCarlo, Lawrence T., Teachers College, Columbia University, decarlo@tc.edu  
de la Torre, Jimmy, University of Hong Kong, j.delatorre@hku.hk

## PARTICIPANT EMAIL

(Last Name, First Name, Affiliation, Email)

De Lisle, Jerome, University of the West Indies,, jeromedelisle@yahoo.com  
DeMars, Christine, James Madison University, demarsce@jmu.edu  
Demir, Cihan, Washington State University, cihan.demir@wsu.edu  
Demirkaya, Onur, Riverside Insights, onurdmrkaya@gmail.com  
Deng, Hui, College Board, hdeng0466@yahoo.com  
Deng, Jiayi, University of Minnesota, deng0194@umn.edu  
Deng, Xinyue, Fordham University, xdeng27@fordham.edu  
DePascale, Charles, Psychometric Confections, LLC, cdepascale@nciea.org  
Deshmukh Towery, Ila, Education First, ideshmukhtowery@education-first.com  
Desiderio, Amy, Harvard, adesiderio@gse.harvard.edu  
DeWeese, Joseph, University of Minnesota, gree2903@umn.edu  
Diao, Hongyu, ETS, hdiao@ets.org  
Diaz, Maria, Northwestern University, Mvarela@northwestern.edu  
Diaz-Bilello, Elena, University of Colorado Boulder, Center for Assessment, Design, Research and Evaluation (CADRE), Elena.diazb@colorado.edu  
DiCerbo, Kristen, Khan Academy, kdicerbo@cox.net  
Diemer, Matthew, University of Michigan, diemerm@umich.edu  
DiLorenzo, Julius, Harvard, jdilorenzo@gse.harvard.edu  
Dixon-Roman, Ezekiel, University of Pennsylvania, ezekield@sp2.upenn.edu  
Dobria, Lidia, Wilbur Wright College, LDobria@ccc.edu  
Dogan Lacin, Betul Gokcen, Mustafa Kemal University, betulgokcen.doganlacin@mku.edu.tr  
Dolan, Robert, Diverse Learners Consulting, bob@diverselearnersconsulting.com  
Domaleski, Chris, Center for Assessment, cdomaleski@nciea.org  
Domingue, Ben, Stanford University, ben.domingue@gmail.com  
Dray, Amy, Spencer Foundation, ajdray@gmail.com  
Drossel, Kerstin, Universität Paderborn, kerstin.drossel@upb.de  
Drost, Bryan, Rocky River City School District, drost.bryan@rrcs.org  
Du, Kuixi, University of Southern Mississippi, kuixi.du@usm.edu  
Dumas, Denis, University of Georgia, Denis.Dumas@uga.edu  
Dunbar, Stephen, University of Iowa, steve-dunbar@uiowa.edu  
Dunn, Jennifer, Pearson Assessments, Jenn.Dunn@pearson.com  
Duran, Lillian, University of Oregon, lduran@uoregon.edu  
Dworak, Elizabeth, Northwestern University, elizabeth.dworak@northwestern.edu  
Dwyer, Kevin, Cambium Assessment, kevin.dwyer@cambiumassessment.com  
Ebert, Isabelle, U.S. Air Force, isabelle.ebert@us.af.mil  
Edeh, Ejike, University of Arkansas, ejedeh@uark.edu  
Edwards, Kelly, American Board of Pediatrics, kedwards@abpeds.org  
Egan, Karla, EdMetric LLC, karla.egan@edmetric.com  
Eickelmann, Birgit, Universität Paderborn, birgit.eickelmann@upb.de  
Embretson, Susan, Georgia Institute of Technology, susan.embretson@psych.gatech.edu  
Engelhard, George, UGA, engelgh@uga.edu  
Englert, Kerry, Seneca Consulting, LLC, kenglert@senecaconsulting.org  
Ercikan, Kadriye, ETS, kercikan@ets.org  
Evans, Carla M., National Center for the Improvement of Educational Assessment, cevans@nciea.org  
Everson, Howard, Graduate Center, City University of New York, howard.everson@gmail.com  
Fan, Fen, NBME, FFan@nbme.org  
Fan, Meng, Human Resources Research Organization, mfan@humrro.org  
Fang, Luyang, University of Georgia, Luyang.Fang@uga.edu  
Faulkner-Bond, Molly, WestEd, mfaulkn@wested.org  
Fauss, Michael, ETS, mfauss@ets.org  
Federiakin, Denis, Johannes Gutenberg University Mainz, denis.federiakin@gmail.com  
Feinberg, Rich, National Board of Medical Examiners, rfeinberg@nbme.org  
Feng, Tianying, UCLA CRESST, tfeng@cresst.org  
Feng, Wanyong, UMass Amherst, wanyongfeng@umass.edu  
Feng, Xiaoying, American Institutes for Research, ctr\_xfeng@air.org  
Feng, Zechu, The University of Hong Kong, jason521@connect.hku.hk

## PARTICIPANT EMAIL

(Last Name, First Name, Affiliation, Email)

Ferido, Marlene, The University of the Philippines, m.ferido@actrc.org  
Ferland, Alane, NWEA, alane.ferland@nwea.org  
Fernandez, Maria-Paz, UCLA, mpfernandez@g.ucla.edu  
Ferrara, Steve, HumRRO, sferrara1951@gmail.com  
Fina, Anthony D., University of Iowa, anthony-fina@uiowa.edu  
Finch, Holmes, Ball State University, whfinch@bsu.edu  
Fincher, Melissa, edCount, LLC, mfincher@edcount.com  
Fink, Aron, Goethe University Frankfurt, a.fink@psych.uni-frankfurt.de  
Finnegan, Robert, ETS, RFinnegan@ets.org  
Finney, Sara, James Madison University, finneysj@jmu.edu  
Firoozi, Tahereh, University of Alberta, tahereh@ualberta.ca  
Fishtein, Daniel, ETS, DFISSTEIN@ets.org  
Fleckenstein, Johanna, IPN KIEL, fleckenstein@ipn.uni-kiel.de  
Fleetwood, April, Florida Virtual School, afleetwood@flvs.net  
Flor, Michael, ETS, mflor@ets.org  
Flynn, Michael, Cambium Assessment, Inc., michael.flynn@cambiumassessment.com  
Fnu, Chandramani, University of Maryland, hjiao1@umd.edu  
Foo, Suan Peen, National University of Singapore, foosp@nus.edu.sg  
Forman, Rachel, McKinsey & Company, rachel\_forman@mckinsey.com  
Forsyth, Carol McGregor, CForsyth@ets.org  
Forte, Ellen, edCount, LLC, eforte@edcount.com  
Forzani, Elena, Boston University, eforzani@bu.edu  
Foster, David, Caveon Test Security, david.foster@caveon.com  
Franklin, Melina, EdMetric, LLC, melia.franklin@edmetric.com  
French, Brian, Washington State University, frenchb@wsu.edu  
Frey, Andreas, Goethe University Frankfurt, frey@psych.uni-frankfurt.de  
Frey, Sharon, Riverside Insights, sharon.frey@riversideinsights.com  
Friedman, Tim, tim.friedman@acer.org  
Fu, Jianbin, ETS, jfu@ets.org  
Fu, Yanyan, Graduate Management Admission Council, yfu@gmac.com  
Fulmer, Gavin, HMH, Gavin.Fulmer@hnhco.com  
Fung, Karen, The Pharmacy Examining Board of Canada (PEBC), karen.fung@ualberta.ca  
Furgol Castellano, Katherine, ETS, KCastellano@ets.org  
Gales, Mark, Enhanced Speech Technology, mjfg@eng.cam.ac.uk  
Gandara, Fernanda, Room to Read, fernanda.gandara@roomtoread.org  
Gane, Brian, Accessible Teaching, Learning, and Assessment Systems ATLAS University of Kansas, bgane@ku.edu  
Gao, Shilin, Cambridge University Press & Assessment, shilin.gao@cambridge.org  
Gao, Yizhu, Government of Alberta, yizhu@ualberta.ca  
Garcia, Avery Nicole, avery.garcia@us.af.mil  
Garcia, Elda, Imagine Learning, elda.garcia@imaginelearning.com  
Ge, Yuan, The College Board, yge@collegeboard.org  
Gelbal, Selahattin, Hacettepe University, sgelbal@gmail.com  
Gershon, Richard, Northwestern University, gershon@northwestern.edu  
Gezer, Tuba, tgezer@uncc.edu  
Gholson, Melissa L., ATLAS, University of Kansas, melissa.gholson@gmail.com  
Gianopulos, Garron, Cambium Assessment, garron@gianopulos.com  
Gierl, Mark, University of Alberta, mark.gierl@ualberta.ca  
Gilbar, Charlotte, Natrona County School District, charlotte\_gilbar@natronaschools.org  
Gill, Matthew, South Dakota Department of Education, matthew.gill@state.sd.us  
Gitomer, Drew, Rutgers, The State University of New Jersey, drew.gitomer@gse.rutgers.edu  
Gochyyev, Perman, University of California, Berkeley, perman@berkeley.edu  
Goldhammer, Frank, DIPF | Leibniz Institute for Research and Information in Education, ZIB, goldhammer@dipf.de  
Gombert, Sebastian, Leibniz Institute for Research and Information in Education, s.gombert@dipf.de  
Gong, Brian, Center for Assessment, bgong@nceia.org  
Gonzalez, Eugenio, ETS, sreign@netscape.net  
Gonzalez, Mario, Cambium Assessment, mario.gonzalez@cambiumassessment.com

## PARTICIPANT EMAIL

(Last Name, First Name, Affiliation, Email)

Gonzalez-Wegener, Xaviera, UCL Institute of Education, xgonzalezwe@gmail.com  
Gooch, Reginald M, ETS, rmgooch@ets.org  
Goodman, Joshua., National Commission on Certification of Physician Assistants (NCCPA), joshuag@nccpa.net  
Goodman, Madeline, ETS, mgoodman@ets.org  
Goodwin, Amanda, Vanderbilt University, amanda.goodwin@vanderbilt.edu  
Gorgun, Guher, University of Alberta, gorgun@ualberta.ca  
Gorham, Amanda, agorham@umass.edu  
Gorham, Jerry L., Tropical Gulf Measurement, jerry.gorham@gmail.com  
Gorney, Kylie, Michigan State University, kgorney@msu.edu  
Gorno Tempini, Marilu, University of California, San Francisco, marialuisa.gornotempini@ucsf.edu  
Gotzmann, Andrea Julie, Medical Council of Canada, agotzmann@mcc.ca  
Gow, Andrew, National Board of Chiropractic Examiners (NBCE), agow@nbce.org  
Grabovsky, Irina, NBME, igrabovsky@nbme.org  
Graf, Edith Aurora, ETS, agraf@ets.org  
Grant, Rosalie, WIDA at the University of Wisconsin – Madison, rosalie.grant@wisc.edu  
Griger, Cassandra, University of Iowa, cgriger@uiowa.edu  
Griph, Gerald, PEARSON, gerald.griph@pearson.com  
Grochowalski, Joe, College Board, jgrochowalski@collegeboard.org  
Gu, Yuqi, Columbia University, yuqi.gu@columbia.edu  
Guez, Ava, OECD, ava.guez@oecd.org  
Guo, Hongwen, ETS, hguo@ets.org  
Guo, Qi, Medical Council of Canada, qguo@mcc.ca  
Guo, Yage Leah, Center for Applied Linguistics, yageguo@gmail.com  
Guven, Ahmet, Augusta University, aguyen@augusta.edu  
Ha, Jen, Harvard, jenniferha@g.harvard.edu  
Haab, Sergio, shaab@uiowa.edu  
Haberman, Shelby Joel, sjhaberman@gmail.com  
Habermehl, Kyle, Pearson, kyle.habermehl@pearson.com  
Ham, Eun Hye, Kongju National University, thanks02@gmail.com  
Hamdani, Maria, Center for Measurement Justice, mhamdani@measurementjustice.org  
Hamilton, Laura, American Institutes for Research, lhamilton@air.org  
Hamilton, Lauren, Edmentum, Inc., Lauren.Hamilton@edmentum.com  
Han, Catherine, Northwestern University, ych@northwestern.edu  
Han, Suhwa, University of Texas at Austin, suhwa@utexas.edu  
Han, Youngjin, University of Maryland College Park, younghan@umd.edu  
Hanno, Emily, emily.hanno@mdrc.org  
Hanrahan, Rachel, Northwestern University, r-hanrahan@northwestern.edu  
Hao, Jiangang, ETS, jhao@ets.org  
Hao, Shiqi, Michigan Department of Education, haos@michigan.gov  
Hardy, Michael, Stanford University, hardy.mike@gmail.com  
Harris, Deborah, University of Iowa, deborah-harris@uiowa.edu  
Haryanto, Bakir, Australian Council for Educational Research, Bakir.Haryanto@acer.org  
Hatcher, Wilson, UNLV CREA, wilson.hatcher@unlv.edu  
Hauenstein, Clifford Erhardt, Johns Hopkins School of Medicine, chauens2@jh.edu  
Hauerwas, Laura Boynton, Providence College, lhauerwa@providence.edu  
Haviland, Sara, ETS, shaviland@ets.org  
Hayes, Anne, Inclusive Development Partners, anne@inclusivedevpartners.com  
Haynie, Kathleen C, Haynie Research and Evaluation, kchaynie@stanfordalumni.org  
He, Qiwei, Georgetown University, qiwei.he@georgetown.edu  
He, Yi, Edmentum, yi.he@edmentum.com  
Hearn, Adam, American Institutes for Research, ahearn@air.org  
Hellman, Scott, Pearson, scott.hellman@pearson.com  
Hendrickson, Amy H., College Board, ahendrickson@collegeboard.org  
Hernandez-Saca, David, University of Northern Iowa, david.hernandez-saca@uni.edu  
Hester, Olivia, University of Alabama, orhester@ua.edu  
Hicks, Juanita, AIR, jhicks@air.org

## PARTICIPANT EMAIL

(Last Name, First Name, Affiliation, Email)

Hille, Kathryn, ETS, khille@ets.org  
Hilz, Anna, ahilz@leibniz-ipn.de  
Himelfarb, Igor, NBCE, ihmelfarb@nbce.org  
Himmelsbach, Zach, Harvard University, zah972@g.harvard.edu  
Hinds, Fiona, Independent Consultant, fiona@womeninmeasurement.org  
Hipfner-Boucher, Kathleen, Research Officer, khibou59@gmail.com  
Ho, Andrew, Harvard Graduate School of Education, Andrew\_Ho@gse.harvard.edu  
Ho, TsungHan, ETS, tho@ets.org  
Hoang, Andrew, ETS, ahoang@ets.org  
Hofman, Abe, University of Amsterdam, a.d.hofman@uva.nl  
Höft, Lars, Leibniz Institute for Science and Mathematics Education, hoeft@leibniz-ipn.de  
Hong, Hyeri, California State University, Fresno, hyerihong@mail.fresnostate.edu  
Hooks-Singletary, Brianna, UNCG, briannahooks41@gmail.com  
Hooper, Martin, American Institutes for Research, mhooper@air.org  
Horbach, Andrea, Universität Hildesheim, horbach@uni-hildesheim.de  
Horowitz, Karen, Curriculum Associates, KHorowitz@cainc.com  
Houriet, Andrew, NBME, ahouriet@nbme.org  
Howard, Tyrone, UCLA, thoward@gseis.ucla.edu  
Hsieh, Ching-Ni, ETS, chsieh@ets.org  
Hu, An, State Key Laboratory for Artificial Microstructure and Mesoscopic Physics, School of Physics, Peking University, lakecool.x@gmail.com  
Hu, Ann, NWEA, ann.hu@nwea.org  
Hua, Cheng, University of Montevallo, chua@montevallo.edu  
Huang, Chi-Yu, ACT, INC, chiyu.huang@act.org  
Huang, Fuchun, Australian Council for Educational Research, Fuchun.Huang@acer.org  
Huang, Qi, University of Wisconsin - Madison, qhuang85@wisc.edu  
Huang, Sijia, Indiana University Bloomington, sijhuang@iu.edu  
Huang, Yingshi, University of California, Los Angeles, yingshi@ucla.edu  
Huang, Yue, Measurement Incorporated, yueh@udel.edu  
Huff, Kristen, Curriculum Associates, khuff@cainc.com  
Hu Flexer, Caroline, Khan Academy, carolineflexer@khanacademy.org  
Huggins-Manley, Anne Corinne, University of Florida, amanley@coe.ufl.edu  
Huo, Huade, AIR, hhuo@air.org  
Hutchinson, Angelica, NBME, achutchinson@nbme.org  
Ihlenfeldt, Samuel Dale, University of Minnesota, ihlen010@umn.edu  
Ikoma, Sakiko, American Institutes for Research, sikoma@air.org  
Ilagan, Michael, ETS Canada, milagan@etscanada.ca  
Immekus, Jason, University of Louisville, jcimme01@louisville.edu  
Ingrisone, James, Pearson VUE, ingrisone@gmail.com  
Ingrisone, Soo, PEARSON, singrisone@gmail.com  
Isaacs, Jazmin, NWEA, jazmin.isaacs@nwea.org  
Ishii, Hidetoki, Nagoya university, ishii.hidetoki.y1@f.mail.nagoya-u.ac.jp  
Iwatani, Emi, Digital Promise, eiwatani@digitalpromise.org  
Jackson, Janine, Morgan State University, Jajac64@morgan.edu  
Jaeger, Garrett, LEGO Foundation, garrett.jaeger@legofoundation.com  
Jansen, Brenda, University of Amsterdam, b.r.j.jansen@uva.nl  
Jansen, Thorben, Leibniz Institute for Science and Mathematics Education, tjansen@leibniz-ipn.de  
Jarr, Karoline, Curriculum Associates, KJarr@cainc.com  
Jasso, Javier, The Ohio State University, javjasso@gmail.com  
Jeon, Eunjeong, EWAH Womans University, euneun1202@ewhain.net  
Jeon, Minjeong, UCLA, mjjeon@ucla.edu  
Ji, Jeongmin, University of Iowa, jeongmin-ji@uiowa.edu  
Jia, Hao, National Council of State Boards of Nursing NCSBN, hjia@ncsbn.org  
Jiang, Tao, Cambium Assessment, Inc, tao.jiang@cambiumassessment.com  
Jiang, Yang, ETS, yjiang002@ets.org  
Jiang, Yanlin, National Commission on Certification of Physician Assistants, jiangyanlin@hotmail.com

## PARTICIPANT EMAIL

(Last Name, First Name, Affiliation, Email)

Jiao, Hong, University of Maryland, hjiao@umd.edu  
Jin, Kuan Yu, HKEAA, kyjin@hkeaa.edu.hk  
Jin, Yi, kimm@connect.hku.hk  
Jing, Zeyuan, University of Florida, jingzeyuan@ufl.edu  
Joachims, Thorsten, Cornell University, Tjoachims@Cornell.edu  
Johnson, Evelyn, Aperture Education, ejohnson@apertureed.com  
Johnson, Janice Lee, NWEA, janice.johnson@nwea.org  
Johnson, Marc Anthony, National Board of Medical Examiners, majsax01@hotmail.com  
Johnson, Matthew S., ETS, msjohnson@ets.org  
Johnstone, Christopher, University of Minnesota, john4810@umn.edu  
Jolivet, Kristine, University of Alabama, jolivet@ua.edu  
Jones, Andrew, American Board of Surgery, ajones@absurgery.org  
Jones, Christopher, College Board, chjones@collegeboard.org  
Jones, Edmund, Cambridge University Press & Assessment, edmund.jones@cambridge.org  
Jones, Eli Andrew, The University of Memphis, eli.jones@memphis.edu  
Joo, Sean, University of Kansas, sjoo@ku.edu  
Julian, Marc W., DRC, mjulian@datarecognitioncorp.com  
Jung, Ae Kyong, University of Iowa, aekyong-jung@uiowa.edu  
Jung, Ji Yoon, Boston College, jiyoon.jung@bc.edu  
Junker, Brian, Carnegie Mellon University, brian@stat.cmu.edu  
Jurich, Daniel, National Board of Medical Examiners, DJurich@nbme.org  
Kaat, Aaron, Northwestern University, aaron.kaat@northwestern.edu  
Kadir, Latif, The Ohio State University, kadir.11@osu.edu  
Kamata, Akhito, Southern Methodist University, akamata@smu.edu  
Kamei, Toshiko, University of Melbourne, kamei.t@unimelb.edu.au  
Kang, Hyeon-Ah, University of Texas at Austin, hkang@austin.utexas.edu  
Kang, Kyosin, American Institutes for Research, kkang@air.org  
Kang, Yulim, Yonsei University, kangyulim@naver.com  
Kannan, Priya, WestEd, pkannan@wested.org  
Kanopka, Klint, New York University, klint.kanopka@nyu.edu  
Kaplan, David, University of Wisconsin, Madison, david.kaplan@wisc.edu  
Kapoor, Radhika, Stanford University, rkap786@stanford.edu  
Kapoor, Shalini, ACT, INC., shalinikapoor.ia@gmail.com  
Kara, Yusuf, Southern Methodist University, ykara@smu.edu  
Karamese, Hacer, WIDA at University of Wisconsin, haker.karamese@wisc.edu  
Karimi, Honeiah, University of California – Santa Barbara, hkarimi@ucsb.edu  
Karunakaran, Shiv, HMM, Shiv.Karunakaran@hmmco.com  
Karvonen, Meagan, University of Kansas, karvonen@ku.edu  
Katz, Daniel, NWEA, daniel.katz@nwea.org  
Kaye, Elaine, James Madison University, robertef@jmu.edu  
Kelberlau, Darin, Millard Public Schools, dkelberlau@mpsomaha.org  
Keller, Lisa, UMASS Amherst, lkeller@educ.umass.edu  
Keller, Stefan Daniel, Zurich University of Teacher Education, stefandaniel.keller@phzh.ch  
Kelley, Rebecca, University of New Hampshire, becky.bierman@gmail.com  
Kelly, Justin, Center for Applied Linguistics, jkelly@cal.org  
Kendall Brooks, Lauren D., lkendallbrooks@aerdf.org  
Keng, Leslie, National Board of Medical Examiners, lesliekeng@gmail.com  
Kennedy, Patrick, University of Oregon, ppaine@uoregon.edu  
Kern, Justin L., University of Illinois at Urbana-Champaign, kern4@illinois.edu  
Kerzabi, Emily, ETS, ekerzabi@ets.org  
Ketan, University of Massachusetts, ketan@umass.edu  
Ketterlin Geller, Leanne, Southern Methodist University, lkeller@mail.smu.edu  
Kevelson, Marisol, ETS, mkevelson@ets.org  
Kim, Eunbee, Georgia Institute of Technology, eunbee.kim@gatech.edu  
Kim, Hahyeong, University of Illinois at Urbana-Champaign, hk33@illinois.edu  
Kim, Han Yi, Cognia, HanYi.Kim@cognia.org

## PARTICIPANT EMAIL

(Last Name, First Name, Affiliation, Email)

Kim, Hyung Jin, University of Iowa, hyungjin-kim@uiowa.edu  
Kim, Hyunjoo, University of Illinois at Urbana-Champaign, hyunjoo5@illinois.edu  
Kim, Kyung Yong, University of North Carolina Greensboro, k\_kim9@uncg.edu  
Kim, Seongeun, University of North Carolina at Greensboro, s\_kim45@uncg.edu  
Kim, Sohee, University of Alabama, skim@southalabama.edu  
Kim, Stella, University of North Carolina at Charlotte, stella-kim@uncc.edu  
Kim, Sujin, George Mason University, skim222@gmu.edu  
Kim, Yewon, EWHA Womans University, kyw8084@ewhain.net  
Kim, YoungKoung, College Board, ykim@collegeboard.org  
Kim, Young Yee, American Institutes for Research, ykim@air.org  
Kim, Yun-Kyung, UCLA, YUNKIM2729@GMAIL.COM  
Kincaid, Heath, American Board of Obstetrics and Gynecology, hkincaid@abog.org  
King, Emma, American Institutes for Research, eking@air.org  
Kingston, Neal, University of Kansas, nkingston@ku.edu  
Kinsey, Devon, ETS, dkinsey001@ets.org  
Kirsch, Irwin, ETS, ekirsch@ets.org  
Kiser, Keith, ETS, kkiser@ets.org  
Kislyonkova, Alexandra Levin, UC Merced, akislyonkova@ucmerced.edu  
Kizilecec, Rene, Cornell University, rkizilecec@cornell.edu  
Klugman, Emma M., Harvard Graduate School of Education, eklugman@g.harvard.edu  
Koehn, Hans Friedrich, University of Illinois at Urbana-Champaign, hkoehn@illinois.edu  
Kolen, Michael, The University of Iowa, kolenmichael@gmail.com  
Kong, Lingchen, University of Florida, l.kong@ufl.edu  
Koprowicz, Nick, Sr. AI and Machine Learning Scientist at Prometric, Nick.Koprowicz@prometric.com  
Kowalski, Susan, HMH, Susan.Kowalski@hnhco.com  
Kozikowski, Andrzej, National Commission on Certification of Physician Assistants (NCCPA), andrzejk@nccpa.net  
Kreitchmann, Rodrigo Schames, Universidad Nacional de Educacion a Distancia, rschames@psi.uned.es  
Krist, Andrew, The University of Alabama, atkrist@crimson.ua.edu  
Krost, Kevin, Virginia Polytechnic Institute and State University, kevinkrost@vt.edu  
Kuhfeld, Megan, NWEA, megan.kuhfeld@nwea.org  
Kukea Shultz, Pohai, University of Hawaii at Manoa, pohai@hawaii.edu  
Kuklick, Livia, IPN Kiel, Germany, kuklick@leibniz-ipn.de  
Kumar, Lavanya Shraavan, University of South Florida, lkumar1@usf.edu  
Kwako, Alexander, Cambium Assessment, akwako@ucla.edu  
Kwon, Muwon, University of Maryland - College Park, mwkwon21@umd.edu  
Kwon, Tae Yeon, University of Florida, taeyeon.kwon@ufl.edu  
Kyllonen, Patrick Charles, ETS, pkyllonen@ets.org  
Kyung, Jungwon, Umass Amherst, jkyung@umass.edu  
LaFlair, Geoff, Duolingo, geoff@duolingo.com  
Laforte, Erica, Northwestern University, ericalaforte@gmail.com  
Lai, Hollis, University of Alberta, hollis1@ualberta.ca  
Lai, Viet, University of Oregon, vietl@uoregon.edu  
Laitusis, Cara C., Center for Assessment NCIEA, claitusis@nciea.org  
Lakin, Joni, University of Alabama, jlakin@ua.edu  
Lallmamode, Sheila, Independent Assessment Researcher, slallmamode@gmail.com  
Lamm, Rik, Lammx020@umn.edu  
Lan, Andrew, University of Massachusetts at Amherst, andrewlan@cs.umass.edu  
Lan, Min, Zhejiang Normal University, lanmin@zjnu.edu.cn  
Landl, Erika, Center for Assessment, elandl@nciea.org  
Lane, Suzanne, University of Pittsburgh, sl@pitt.edu  
Langi, Meredith, WestEd, mlangi@wested.org  
Lansing-Stoeffler, Kristin, ACT, INC., kristin.stoeffler@act.org  
Laster, Christina, Parent and Family Advocate, christinaclaster@gmail.com  
La Torre, Deborah, UCLA, dlatorre@ucla.edu  
Laufer, Alice, Johannes Gutenberg University, alaufer@students.uni-mainz.de  
Law, Evelyn, National University of Singapore, paelecn@nus.edu.sg

## PARTICIPANT EMAIL

(Last Name, First Name, Affiliation, Email)

Law, Nancy, University of Hong Kong, nlaw@hku.hk  
Lawler, Mark, NWEA, mark.lawler@nwea.org  
Lazarus, Sheryl, NATIONAL CENTER ON EDUCATIONAL, laza0019@umn.edu  
Lazenby, Katherine, HMH, Katherine.Lazenby@hnhco.com  
Lazendic, Goran, ACER, goran.lazendic@glazendic.com  
Le, Anh Thu, ale3@nd.edu  
Lee, Chansoon, American Board of Internal Medicine, dlee@abim.org  
Lee, Dayeon, University of Maryland - College Park, daylee@umd.edu  
Lee, Dukjae, University of Virginia, dlee@umass.edu  
Lee, Eunji, University of Georgia, ejlee@uga.edu  
Lee, Guemin, Yonsei University, guemin@yonsei.ac.kr  
Lee, Gyeong-Geon, University of Georgia, crusaderlee@snu.ac.kr  
Lee, Haeju, University of North Carolina Greenboro, HLEE@uncg.edu  
Lee, Hyeryung, University of Iowa, hyerlee@uiowa.edu  
Lee, Jade Caines, University of Kansas, jade.caines.lee@gmail.com  
Lee, Jake, UMass Amherst, jaewooklee@umass.edu  
Lee, Ja-yoon, Seoul National University, muesli99@snu.ac.kr  
Lee, jinsook, Cornell University, JLee@cornell.edu  
Lee, Meng-Hsun, University of Toronto, menghsun.lee@mail.utoronto.ca  
Lee, Mina, Cambium Assessment, mina.lee@cambiumassessment.com  
Lee, Minh, University of California Los Angeles, lemino72@ucla.edu  
Lee, Rachel, yl3751@tc.columbia.edu  
Lee, Selene Sunmin, slee003@ets.org  
Lee, Seulbi Bailey, Graduate Student, blee3@uoregon.edu  
Lee, Seunghyun, Columbia University, sl4963@columbia.edu  
Lee, Soo, American Institutes for Research, slee@air.org  
Lee, Sung-Hyuck, GMAC, lee9410@hotmail.com  
Lee, Sunhyoung, University of Nebraska-Lincoln, slee82@huskers.unl.edu  
Lee, Won-Chan, University of Iowa, won-chan-lee@uiowa.edu  
Lee, Youngjun, The American Board of Anesthesiology, youngjun.lee@theaba.org  
Lehman, Blair, ETS, blehman@ets.org  
Lekwa, Adam, Rutgers, The State University of New Jersey, adam.lekwa@rutgers.edu  
LeMahieu, Paul, Carnegie Foundation, plem@carnegiefoundation.org  
Leng, Dihao, Boston College, dihao.leng@bc.edu  
Leonard, Jana, Cajon Valley Union School District, leonardj@cajonvalley.net  
Leventhal, Brian C., James Madison University, leventbc@jmu.edu  
Lewin, Nancy, ACT, Nancy.Lewin@act.org  
Lewis, Crystal, Crystal Lewis Consulting, LLC, cgh.lewis@gmail.com  
Lewis, Daniel, Creative Measurement Solutions LLC, dan.lewis@creativemeasurement.com  
Lewis, Regina A., Maine Department of Education, regina.lewis@maine.gov  
Li, Chen, (ETS), cli@ets.org  
Li, Dongmei, ACT, dongmei.li@act.org  
Li, Grace, WIDA at the University of Wisconsin - Madison, grace.li@wisc.edu  
Li, Isaac, American Board of Psychiatry and Neurology, hushbush@hotmail.com  
Li, Jie, lijdbc@gmail.com  
Li, Jiehan, SELECT OR ENTER, jiehanli@ufl.edu  
Li, Jingyang, University of Georgia, jl61533@uga.edu  
Li, Xin (Grace), University of Wisconsin-Madison, gracexinlee@gmail.com  
Li, Xueming, NWEA, sylvia.li@nwea.org  
Li, Zhushan Mandy, Boston College, zhushan.li@bc.edu  
Li, Zikun, George Mason University, zli35@gmu.edu  
Liang, Min, min-liang-1@uiowa.edu  
Liang, Qianru, Jinan University, qr.liang@outlook.com  
Liang, Xinya, Department of Rehabilitation, Human Resources and Communication Disorders, University of Arkansas, xl014@uark.edu  
Liao, Xiangyi, University of Wisconsin, Madison, xliao36@wisc.edu

## PARTICIPANT EMAIL

(Last Name, First Name, Affiliation, Email)

Liao, Yuqi, American Institutes for Research, ylia@air.org  
Liaw, Yuan-Ling, IEA Hamburg, yuan-ling.liaw@iea-hamburg.de  
Lim, Hwanggyu, GMAC, hlim@gmac.com  
Lim, Sangdon, Cambium Assessment, Sangdon.Lim@cambiumassessment.com  
Lim, Shuping, National University of Singapore, shuping.lim@nus.edu  
Lim, Youn Seon, University of Cincinnati, limyo@ucmail.uc.edu  
Lin, Ami, Aledev Consulting, ahoffman@aledev.com  
Lindner, Marlit Annalena, IWM - Leibniz Institut für Wissensmedien, University of Tübingen, m.lindner@iwm-tuebingen.de  
Lindsay, Constance, University of North Carolina at Chapel Hill, clindsay@unc.edu  
Ling, Guangming, ETS, gling@ets.org  
Lions, Séverin, CIAE - Universidad de Chile, severin.lions@ciae.uchile.cl  
Lipien, Lodi, Florida Virtual School, llipien@flvs.net  
Lipscomb, Thomas J, The University of Southern Mississippi, Thomas.Lipscomb@usm.edu  
Liu, Chen-Wei, National Taiwan Normal University, cwliu@ntnu.edu.tw  
Liu, Chunyan, National Board of Medical Examiners, cliu@nbme.org  
Liu, Guangyun, University of Iowa, lguangyun@uiowa.edu  
Liu, Huan, The University of Iowa, huan-liu-1@uiowa.edu  
Liu, Jinghua, National Board of Osteopathic Medical Examiners, jinghuawalkerliu@gmail.com  
Liu, Kristin, University of Minnesota, kline010@umn.edu  
Liu, Lei, ETS, lliu001@ets.org  
Liu, Liu, University of Washington, liuul@uw.edu  
Liu, Na, na.liu@gatech.edu  
Liu, Ou Lydia, ETS, lliu@ets.org  
Liu, Ren, University of California, MERC, rliu45@ucmerced.edu  
Liu, Tuo, Goethe University Frankfurt, Germany, t.liu@psych.uni-frankfurt.de  
Liu, Xiang, ETS, xliu003@ets.org  
Liu, Xiaoxiao, University of Alberta, xiaoxia6@ualberta.ca  
Liu, Yan, Carleton University, YanLiu5@cunet.carleton.ca  
Liu, Yang, University of Maryland, College Park, yliu87@umd.edu  
Liu, Yiqing, University of Illinois, yiqing8@illinois.edu  
Liu, Yunting, UC Berkeley, yunting99@berkeley.edu  
Lizano, Cheryl, Southern Illinois University Edwardsville, ceames@siue.edu  
Lochbaum, Karen, Pearson, Karen.Lochbaum@pearson.com  
Lohmann, Julian, Kiel University, jlohmann@ipl.uni-kiel.de  
Long, Yunyi, National Board of Osteopathic Medical Examiners, ylong@nbome.org  
Longe, Brendan, University of Massachusetts Amherst, blonge@umass.edu  
Lorie, William A., Center for Assessment, william.lorie@gmail.com  
Lottridge, Susan, Cambium Assessment, Inc, susan.lottridge@cambiumassessment.com  
Love, Quintin, WestEd, qllove@wested.org  
Lovelace, Temple, Assessment for Good, tlovelace@aerdf.org  
Lu, Xiangyu, xyliu31415@gmail.com  
Lu, Yang, Pearson, Yang.Lu@pearson.com  
Lu, Yikai, University of Notre Dame, ylu22@nd.edu  
Lu, Ying, College Board, ylu@collegeboard.org  
Lu, Zhenqiu Laura, University of Georgia, readlaura@yahoo.com  
Lüdtke, Oliver, Leibniz Institute for Science and Mathematics Education, oluedtke@leibniz-ipn.de  
Luecht, Richard Melvin, University of North Carolina at Greensboro, rmluecht@uncg.edu  
Lugu, Benjamin Kweku, University of Alabama, benjamin.lugu@aims.ac.rw  
Luo, Hao, Department of Educational Psychology, Faculty of Education, East China Normal University, 549116113@qq.com  
Luo, Jinwen, UCLA, jevan.luo@gmail.com  
Luo, Rebekah, University of Melbourne, rebekah.luo@unimelb.edu.au  
Luo, Yachen, Human Resources Research Organization, yluo@humrro.org  
Lyons, Susan, Lyons Assessment Consulting, susan@lyonsassessment.com  
Lyu, Weicong, University of Washington, Seattle, wlyu4@uw.edu  
Ma, Jing, The University of Iowa, majing@uiowa.edu

## PARTICIPANT EMAIL

(Last Name, First Name, Affiliation, Email)

Ma, Mingjia, Cambium Assessment, Inc., mingjia.ma@cambiumassessment.com  
Ma, Wenchao, University of Alabama, wenchao.ma@ua.edu  
Ma, Ye, AWS, ymcheryl@amazon.com  
Ma, Ye, cherylma Iowa@gmail.com  
Maas, Tricia, Curriculum Associates, TMaas@cainc.com  
MacIntosh, Alexander, Acuity Insights, amacintosh@acuityinsights.com  
Madison, Matthew James, University of Georgia, mjmadison@uga.edu  
Maeda, Hotaka, Smarter Balanced, hotaka.maeda@gmail.com  
Maeda, Yukiko, Purdue University, ymaeda@purdue.edu  
Majerus, Mary, Missouri Department of Elementary and Secondary Education, mary.majerus@dese.mo.gov  
Makinde, Henry, Student, hsmakinde@uncg.edu  
Malatesta, Jaime, GMAC, jmalatesta@gmac.com  
Man, Kaiwen, University of Alabama, kman@ua.edu  
Mang, Julia, TUM School of Education, Centre for International Student Assessment (ZIB), Technical University of Munich (TUM), Julia.Mang@tum.de  
Mansolf, Maxwell, Northwestern University, maxwell.mansolf@northwestern.edu  
Mao, Xia, NBOME, xmao@nbome.org  
Marder, Andrew, Caveon, andrew.marder@caveon.com  
Mardones, Constanza, University of Georgia, cam04214@uga.edu  
Marini, Jessica, College Board, jmarini@collegeboard.org  
Marion, Scott, National Center for the Improvement of Educational Assessment, smarion@nceia.org  
Mark-Worrell, Nicola, silvermaru@gmail.com  
Martineau, Joseph A., ETS, jmartineau@ets.org  
Martinez, Jose Felipe, UCLA - School of Education and Information Studies, jfmtz@ucla.edu  
Matewere, Roshi, Edmentum, Inc., Roshi.Matewere@Edmentum.com  
Matta, Tyler, NWEA, tyler.matta@nwea.org  
Mazzullo, Elisabetta, University of Alberta, mazzullo@ualberta.ca  
McCaffrey, Daniel, ETS, dmccaffrey@ets.org  
McClarty, Katie, McGraw Hill Education, katie.mcclarty@gmail.com  
McClure, Kyla, University of Colorado Boulder, Kyla.McClure@colorado.edu  
McCormick, Carina, Southeast Community College Board of Governors, carinamc@gmail.com  
McMillan, James, Virginia Commonwealth University, jhmcmill@vcu.edu  
McNichols, William, UMass Amherst, wmcnichols@umass.edu  
Mee, Janet, NBME, jmee@nbme.org  
Melanlioglu, Deniz, Istanbul University, deniz.melanlioglu@istanbul.edu.tr  
Meng, Huijuan, Amazon Web Services AWS, huijuam@amazon.com  
Merkle, Edgar, University of Missouri, merklee@missouri.edu  
Merriman, Jennifer, International Baccalaureate, jen.merriman@ibo.org  
Mertens, Ute, Leibniz Institute for Science and Mathematics Education, mertens@leibniz-ipn.de  
Meyer, Jennifer, Leibniz Institute for Science and Mathematics Education, jmeyer@leibniz-ipn.de  
Meyer, Patrick, NWEA, Patrick.Meyer@nwea.org  
Michaelides, Michalis, University of Cyprus, michaelides.michalis@ucy.ac.cy  
Michaels, Hillary, HUMRRO, hillarymichaels@gmail.com  
Michel, Rochelle, Smarter Balanced, rochelle.michel@gmail.com  
Middlestead, Andrew J., Michigan Department of Education, middlestead@michigan.gov  
Middleton, Kyndra, Howard University, kvmiddleton@gmail.com  
Mikeska, Jamie, ETS, Jmikeska@ets.org  
Mills, Christine, Ascend Learning, Christine.Mills@ascendlearning.com  
Min, Shangchao, Zhejiang University, msc@zju.edu.cn  
Minchen, Nathan, Curriculum Associates, LLC, nathanminchen@gmail.com  
Minshew, Karen, Cajon Valley Union School District, minshew@cajonvalley.net  
Miranda, Alejandra, University of Minnesota - Twin Cities, miran143@umn.edu  
Miratrix, Luke Weisman, Harvard Graduate School of Education, lmiratrix@g.harvard.edu  
Mireles, Nicolas, James Madison University, mirelene@dukes.jmu.edu  
Mislevy, Robert Joseph, retired, robert.mislevy@gmail.com  
Miyashiro, David, Cajon Valley Union School District, miyashirod@cajonvalley.net

## PARTICIPANT EMAIL

(Last Name, First Name, Affiliation, Email)

Mojoyinola, Mubarak O., The University of Iowa, mubarak-mojoyinola@uiowa.edu  
Molin, Lauren, College Board, lmolin@collegeboard.org  
Möller, Jens, Kiel University, jmoeller@ipl.uni-kiel.de  
Monroe, Scott, UMASS Amherst, smonroe@educ.umass.edu  
Montanano, Amber, NBME, amontanano@nbme.org  
Moore, John, NBME, jmoore@nbme.org  
Morell, Linda, University of California, BER, lindamorell@berkeley.edu  
Morris, Armani Michia, armanimorris@yahoo.com  
Morris, Sarah Ruth, srm041@uark.edu  
Morrison, Carol, NBME, cmorrison@nbme.org  
Morrison, Kristin M., Curriculum Associates, KMorrison@cainc.com  
Mosely, Brein, Harvard University, breinmosely@g.harvard.edu  
Moses, Tim, College Board, tmoses@collegeboard.org  
Moteane, Malitsitso, University of North Carolina at Greensboro, m\_motean@uncg.edu  
Mulcaire, Phoebe, Duolingo, phoebe@duolingo.com  
Mulvihill, Megan M., University of Kansas, megan.mulvihill@ku.edu  
Mundhenk, Kimberly, California Department of Education, kmundhenk@cde.ca.gov  
Munoz-Najar Galvex, Sebastian, Harvard University, sebastian\_munozng@gse.harvard.edu  
Muntean, William J., National Council of State Boards of Nursing, williamjmuntean@gmail.com  
Murphy, Daniel, WestEd, dmurphy@wested.org  
Myers, Aaron, ABIM, amyers@abim.org  
Myers, Matthew, University of Delaware, mcmyers@udel.edu  
Nagasawa, Mirai, The University of Alabama, mnagasawa@crimson.ua.edu  
Nash, Brooke, University of Kansas, bnash@ku.edu  
Nason, Kelsey, James Madison University, nasonkt@jmu.edu  
Nava, Imelda, UCLA, inava@g.ucla.edu  
Naveiras, Matthew David, Riverside Insights, matthew.naveiras@riversideinsights.com  
Neiman, Samantha, American Institutes for Research, sneiman@air.org  
Nemeth, Yvette M, HumRRO, ynemeth@humrro.org  
Nesbitt, Jaylin, WestEd, JNesbit2@wested.org  
Nese, Joseph F. T., University of Oregon, jnese@uoregon.edu  
Newstadt, Michelle, BrainPOP, michellen@brainpop.com  
Nguyen, Cuc Thi Kim, University of Melbourne, cuc@unimelb.edu.au  
Nguyen, Tram-Anh Tran, University of Massachusetts Amherst, anhn@umass.edu  
Nguyen Khoa, Lan Anh, University of Melbourne, la.nguyenkhoa@unimelb.edu.au  
Nickodem, Kyle, University of Minnesota - Twin Cities, nicko013@umn.edu  
Nicola, Matthew, Harvard, mnicola@law.harvard.edu  
Nicola, Tara P., Harvard Graduate School of Education, tara\_nicola@g.harvard.edu  
Noh, Jaehee, Seoul National University, jaeheehoh65@snu.ac.kr  
Novack, Miriam, Northwestern University, miriam.novack@northwestern.edu  
Nowinski, Cindy, Northwestern University, c-nowinski@northwestern.edu  
Nugba, Regina Mawusi, University of Cape Coast, Regina.nugba@ucc.edu.gh  
Nydick, Steven, Duolingo, steven@duolingo.com  
Ober, Teresa, ETS, tober@ets.org  
Odom, Kimberly, University of Alabama, kpodom@ua.edu  
O'Donnell, Francis, National Board of Medical Examiners, fodonnell@nbme.org  
Ogut, Burhan, American Institutes for Research, bogut@air.org  
Oh, Hyeon-Joo, Riverside Insights, joannehj@gmail.com  
Oh, Hyunjee, hyunjeeoh1223@gmail.com  
Okay, Serfie, Hacettepe University, serokyav@gmail.com  
Olivarez-Durden, Sandra, Edmentum, Inc., Sandra.Olivarez-Durden@Edmentum.com  
Olivera Aguilar, Margarita, ETS GLOBAL, margarita.olag@gmail.com  
Oliveri, Maria Elena, Buros Center for Testing-UNL, moliveri2@unl.edu  
Olson, Jacinta, University of Iowa, jacinta.olson18@gmail.com  
Oluwalana, Olasumbo, ETS, ooluwalana@ets.org  
O'Neil, Timothy, Pearson, t66oneil@yahoo.com

## PARTICIPANT EMAIL

(Last Name, First Name, Affiliation, Email)

Ong, Thai Quang, National Board of Medical Examiners, tong@nbme.org  
Oppenzato, Colleen, HMM, Colleen.Oppenzato@hmmco.com  
Oppong, Frank Agyemang, fo242022@ohio.edu  
Oquendo, Jodine, AWS, jodino@amazon.com  
Organisciak, Peter, University of Denver, peter.organisciak@du.edu  
Ormerod, Chris, Cambium Assessment, Inc., christopher.ormerod@cambiumassessment.com  
Ortega, Gabriel, Center for Advanced Research in Education (FB0003), Institute of Education, Universidad de Chile, Santiago, Chile, gabriel.ortega@ciae.uchile.cl  
Osses-Vargas, Alejandra, Australian Council for Educational Research, alejandra.osses@acer.org  
Ouyang, Wenli, National Board of Medical Examiners, wouyang@nbme.org  
Ozer Ozkan, Yesim, Gaziantep University, yozer80@gmail.com  
Pace, Lillian, KnowledgeWorks, pacel@knowledgeworks.org  
Page, Jane, University of Melbourne, j.page@unimelb.edu.au  
Pan, Qianqian, National Institute of Education, Nanyang Technological University, panqianqian2013@gmail.com  
Pan, Tianshu, Pearson, tianshu.pan@pearson.com  
Pan, Yanfang, Beijing Normal University, yanfangpan@outlook.com  
Pan, Yiqin, University of Florida, ypan@coe.ufl.edu  
Paraboni, Ivandre, Universidade de São Paulo, ivandre@usp.br  
Pardos, Zachary A., University of California, Berkeley, pardos@berkeley.edu  
Paris, Joseph, West Chester University, jparis@temple.edu  
Park, John, United Nations University, johnpark2023@utexas.edu  
Park, Jung Yeon, George Mason University, jpark233@gmu.edu  
Park, Seohee, American Board of Internal Medicine, spark@abim.org  
Park, Yena, Duolingo, yena@duolingo.com  
Park, Yooyoung, NBOME, pyooyoung@gmail.com  
Parker, Emily Hope, eparker@Apertureed.com  
Parks, Charles, UCLA CRESST, cparks@cresst.org  
Parra-Martinez, Fabio Andres, University of Arkansas, ap448@uark.edu  
Partchev, Ivailo, CITO, partchev@gmail.com  
Patel, Milan, Cambium Assessment, Inc., milan.patel@cambiumassessment.com  
Patel, Nirmal, Smart Paper, nirmal@playpowerlabs.com  
Patelis, Thanos, Johns Hopkins U & University of Kansas, tpatelis@yahoo.com  
Patterson, Jim, College Board, jpatterson@collegeboard.org  
Patterson, John D., Penn State University, jptrsn@psu.edu  
Patton, Elizabeth Adele, Curriculum Associates, BPatton@cainc.com  
Patz, Richard, University of California, Berkeley, rpatz@berkeley.edu  
Pearson, Jami-Jon, Measurement Incorporated, jpearson@measinc.com  
Peasley, Donald, US Department of Education, Donald.Peasley@ed.gov  
Pei, Francesca, University of California, San Francisco, francesca.pei@ucsf.edu  
Pellegrino, James, University of Illinois at Chicago, pellegjw@uic.edu  
Peng, Fang, Houghton Mifflin Harcourt, pfrenee@gmail.com  
Peng, Luyao, Awada Tech LLC, luyaopeng.cn@gmail.com  
Perez, Leticia, WestEd, leticiaperezhuff@gmail.com  
Perie, Marianne, WestEd mp@measurementinpractice.com  
Petscher, Yaacov, Florida Center for Reading Research, ypetscher@fcrr.org  
Pham, Duy N., University of Massachusetts, Amherst, dpham@umass.edu  
Phillip, Kristy, The University of the West Indies, kristy.phillip@my.uwi.edu  
Piacentini, Mario, OECD, mario.piacentini@oecd.org  
Pila, Sarah, Northwestern University, sarah.pila@northwestern.edu  
Pinsonneault, Laura, National Center for the Improvement of Educational Assessment, lpinsonneault@nceia.org  
Pinto, Wallace Nascimento, University of Florida, wallacenpj@gmail.com  
Pitoniak, Mary, ETS, mpitoniak@ets.org  
Pittman, Jocelynn, Cambium Assessment, jocelyn.robinson@cambiumassessment.com  
Poe, Mya, Northeastern University, m.poe@northeastern.edu  
Por, Han-Hui, ETS, hpor@ets.org  
Portilla, Ximena, MDRC, ximena.portilla@mdrc.org

## PARTICIPANT EMAIL

(Last Name, First Name, Affiliation, Email)

Potgieter, Cornelis, Texas Christian University, c.potgieter@tcu.edu  
Powers, Sonya, Edmentum, Inc., sonya.powers@edmentum.com  
Prasad, Seema, Educational Quality and Assessment Programme (EQAP), seemap@spc.int  
Pratt, Amy, University of Cincinnati, Prattat@ucmail.uc.edu  
Prest, June, University of Alabama, jlpreast@ua.edu  
Prendergast, Caroline, American Board of Surgery, cprendergast@absurgery.org  
Prewitt, Nicole, University of Alabama, nbprewitt@ua.edu  
Primi, Ricardo, University of Sao Francisco, rprimi@mac.com  
Proctor, Thomas, College Board, tproctor@collegeboard.org  
Pugsley, John, Pharmacy Examining Board of CA, jpugsley@pebc.ca  
Qiao, Xin, University of South Florida, xqiao@usf.edu  
Qin, Qi, Cognia, qi.qin@cognia.org  
Qiu, Chen, ACT, INC., chenqiu\_qpm.uky2022@outlook.com  
Quan, Yale, University of Washington, yalequan@uw.edu  
Quanbeck, Mari, University of Minnesota - Twin Cities, quanb016@umn.edu  
Quesen, Sarah, WestEd, sarah.quesen@gmail.com  
Quranda, Milton, NBME, mquranda@nbme.org  
Rabinowitz, Stanley N., EdMetric LLC, snr55aus@gmail.com  
Rabkin, Sasha, Equal Opportunity Schools, sasha@eoschools.org  
Rahimi, Ahmad, University of Florida, srahimi@coe.ufl.edu  
Raimondi, Matt, School District U-46, raimondi.matt@gmail.com  
Rajeb, Mehdi, University of Alabama, mrajeb@crimson.ua.edu  
Ramazan, Onur, Washington State University, onur.ramazan@wsu.edu  
Ramnarine, Nirmala, The University of the West Indies, nirmalaramnarine@hotmail.com  
Ramsawak-Jodha, Nalini, University of the West Indies, St. Augustine, Trinidad, nalini.ramsawak-jodha@sta.uwi.edu  
Randall, Jennifer, University of Michigan, jennrand@umich.edu  
Raymond, Krystina, University of Toronto, krystina.raymond@utoronto.ca  
Razavi, Pooya, Edmentum, Inc. pooya.razavi@protonmail.com  
Reckase, Mark, Psychometric Solutions, reckase@msu.edu  
Redman, Elizabeth, UCLA CRESST, redman@cresst.org  
Reise, Steve, UCLA, reise@psych.ucla.edu  
Re-McGlynn, Leta S., National Board of Medical Examiners, lremcglynn@nbme.org  
Ren, He, University of Washington, heren@uw.edu  
Ren, Shixiu, Beijing Normal university, shixiuren@mail.bnu.edu.cn  
Resanovich, Mary, NWEA, mary.resanovich@nwea.org  
Reshetar, Rosemary, National Conference of Bar Examiners, rreshetar@ncbe.org  
Rhodes, Sally, Edmentum, Inc., Sally.Rhodes@Edmentum.com  
Richardson, Julie, NWEA, julie.richardson@nwea.org  
Richardson, Megan, Harvard, mrichardson@gse.harvard.edu  
Rijmen, Frank, Cambium Assessment, frank.rijmen@cambiumassessment.com  
Rikoon, Sam, American Institutes for Research, srikoon@air.org  
Rios, Joseph A., University of Minnesota, jrios@umn.edu  
Rios, Oscar, osrios@ucdavis.edu  
Roberts, James S., Georgia Institute of Technology, james.roberts@psych.gatech.edu  
Roberts, Jeremy, PBS KIDS, jdroberts@pbs.org  
Robin, Frederic, ETS, frobin@ets.org  
Robitaille, Jennifer, Aperture Education, jrobitaille@apertureed.com  
Robitzsch, Alexander, Leibniz Institute for Science and Mathematics Education, robitzsch@leibniz-ipn.de  
Rocha, Horacio A., CUNY Graduate Center, horaciorocha@yahoo.com  
Rodriguez, Michael C., University of Minnesota, mcrdz@umn.edu  
Roduta Roberts, Mary, University of Alberta, mroberts@ualberta.ca  
Roeters-Solano, Heather, Pearson, heather.roeters-solano@pearson.com  
Rome, Logan, Curriculum Associates, lrome@cainc.com  
Rosales de Veliz, Leslie Vanessa, JML Measurement and Testing Services, LLC, leslie.rosales@gmail.com  
Rose, Mark, markr2321@hotmail.com  
Rosen, Yigal, BrainPop, yigalr@brainpop.com

## PARTICIPANT EMAIL

(Last Name, First Name, Affiliation, Email)

Ross, Linette P., NBME, lross@nbme.org  
Roussos, Louis, Cognia, roussos.louis1@gmail.com  
Rozman, Mojca, IEA Hamburg, mojca.rozman@iea-hamburg.de  
Rozunick, Chris, TEA, christine.rozunick@tea.texas.gov  
Ruan, Chunyi, ETS, CRuan@ets.org  
Rubel, Jennifer, Edmentum, Inc., Jennifer.Rubel@edmentum.com  
Rubright, Jonathan, National Board of Medical Examiners, jrubright@nbme.org  
Runge, Andrew, Duolingo, arunge@duolingo.com  
Runyon, Christopher, NBME, CRunyon@nbme.org  
Rupp, Andre, Center for Assessment / NCIEA, arupp@nciea.org  
Rushkin, Ilia, BrainPOP, iliar@brainpop.com  
Russell, Michael, Boston College, michael.russell@bc.edu  
Rutkowski, Leslie, Indiana University, lrutkows@iu.edu  
Rutstein, Daisy Wise, edCount, LLC, drutstein@edcount.com  
Ruzek, Erik, HMH, Erik.Ruzek@hnhco.com  
Ryoo, Ji Hoon, Yonsei University, ryoo001@yonsei.ac.kr  
Sabatini, John, Institute for Intelligent Systems, University of Memphis, jpsbtini@memphis.edu  
Sabbaghi, Arman, Purdue University Department of Statistics, sabbaghi@purdue.edu  
Sabbah, Jenin, Center for Measurement Justice, jsabbah@measurementjustice.org  
Sachdeva, Ruchi, Pearson, ruchi.sachdeva@pearson.com  
Saklani, Dinesh Prasad, National Council of Educational Research and Training, India, director.ncert@nic.in  
Salas, Jorge, Vanderbilt University, jorge.a.salas@vanderbilt.edu  
Saldivia, Luis, ETS, lsaldivia@ets.org  
Saleh, Yustina, The Burning Glass Institute, ysaleh@burningglassinstitute.org  
Samson, Sally, University of Alaska, Fairbanks, spsamson@alaska.edu  
Sanchez, Edgar, ACT, Suppression305@gmail.com  
Sanders, Elizabeth A., University of Washington, Seattle, lizz@uw.edu  
Sanders, Sara, University of Alabama, sara.sanders@ua.edu  
Sarac, Merve, College Board, msarac@collegeboard.org  
Satchu, Salma, The Pharmacy Examining Board of Canada, ssatchu@pebc.ca  
Satkus, Paulius, GMAC, psatkus@gmac.com  
Sato, Edynn, Sato Education Consulting LLC, edynn@satoeducationconsulting.com  
Savitz-Romer, Mandy, Harvard, mandy\_savitz-romer@gse.harvard.edu  
Sayin, Ayfer, Gazi University, ayfersayin@gazi.edu.tr  
Scarlatos, Alex, UMass Amherst, ajscarlatos@umass.edu  
Schaller, Nils-Jonathan, IPN – Leibniz Institute for Science and Mathematics Education, schaller@leibniz-ipn.de  
Schellman, Madeline, Pearson, madeline.schellman@pearson.com  
Schemel, Lynn, Indiana Department of Education, LSchemel@doe.in.gov  
Schiller, Ronja, Leibniz Institute for Science and Mathematics Education, schiller@leibniz-ipn.de  
Schmidgall, Jonathan, ETS, jschmidgall@ets.org  
Schmidt, Susanne, susanne.schmidt@uni-mainz.de  
Schneider, Christina, Cambium Assessment, Inc., christina.schneider@cambiumassessment.com  
Schneider, Wei, Cambium Assessment, wei.schneider@cambiumassessment.com  
Schnittka Hoskins, Jessica Elizabeth, UCLA, jess.es.hoskins@gmail.com  
Schultz, Matthew, AICPA, Matthew.Schultz@aicpa-cima.com  
Schwantner, Ursula, Australian Council for Educational Research, Ursula.Schwantner@acer.org  
Schwappach, Bethany, Cajon Valley Union School District, schwappachb@cajonvalley.net  
Schweig, Jonathan, RAND Corporation, jschweig@rand.org  
Scott-Smith, Chandler, American Board of Surgery, chscottsmith@absurgery.org  
Setzer, Carl, AICPA, csetzer@aicpa.org  
Shan, Xintong, Beijing Normal University, yimo17@126.com  
Sharairi, Sid, Riverside Insights, sid.sharairi@riversideinsights.com  
Sharpnack, James, Duolingo, james.sharpnack@duolingo.com  
Shaw, Emily, College Board, eshaw@collegeboard.org  
Shaw, Timothy, Forum One Communications, tshaw@forumone.com  
Shelley, Kyna J., The University of Southern Mississippi, kyna.shelley@usm.edu

## PARTICIPANT EMAIL

(Last Name, First Name, Affiliation, Email)

Shelton, Stephanie, The University of Alabama, sashelton@ua.edu  
Shen, Yawei, Pearson, shen920501@gmail.com  
Shenavai, Kevin, Johannes Gutenberg University Mainz, kshenava@uni-mainz.de  
Shepard, Lorrie, University of Colorado Boulder, lorrie.shepard@colorado.edu  
Sherer, David, Carnegie Foundation, dsherer@carnegiefoundation.org  
Shermis, Mark D., Performance Assessment Analytics, LLC, mshermis@gmail.com  
Shi, Dingjing, University of Oklahoma, dshi@ou.edu  
Shi, Qingzhou, Northwestern University, qshi7@crimson.ua.edu  
Shih, Ching-Lin, National Sun Yat-sen University, educls@g-mail.nsysu.edu.tw  
Shimizu, Yuki, s.y.koyomi@gmail.com  
Shin, David, Pearson, cshin0803@gmail.com  
Shin, Hyesung, 7castle42@naver.com  
Shin, Hyo Jeong, Sogang University, hshinedu@sogang.ac.kr  
Shin, Jinnie, jinnie.shin@coe.ufl.edu  
Shipan, Rebecca, American Institutes for Research, rshipan@air.org  
Shivraj, Pooja, American Board of OBGYN, pshivraj@abog.org  
Shotts, Bruce L., NBCE, bshotts@nbce.org  
Siebert, Julian Maximilian, Stanford University, jms312@stanford.edu  
Siegel, Danielle, University of California, Davis, dpsiegel@ucdavis.edu  
Siefert, Paula, Northwestern University, paula@heliumfoot.com  
Sikali, Emmanuel, NCES, Emmanuel.Sikali@ed.gov  
Sinclair, Andrea, HumRRO, asinclair@humrro.org  
Sinharay, Sandip, ETS, ssinharay@ets.org  
Sireci, Stephen G, University of Massachusetts, Amherst, sireci@umass.edu  
Sirganci, Gozde, Southern Methodist University, gsirganci@smu.edu  
Skoblow, Hanamori, University of Missouri, skoblowh@mail.missouri.edu  
Slomp, David, University of Lethbridge, david.slomp@uleth.ca  
Smith, Charity, Fetterman & Associates, charitysmith.99@comcast.net  
Smith, Digory, Eedi, digory.smith@eedi.co.uk  
Smith, Jessalyn, DRC, jsmith@datarecognitioncorp.com  
Smith, Larissa, lsmith009@ets.org  
Smith, Mireya Carmen-Martinez, Amplify Education, mart1799@umn.edu  
Smith, Nate, nrsmith@umass.edu  
Soland, James, University of Virginia, jgs8e@virginia.edu  
Solano-Flores, Guillermo, Stanford University, gsolanof@stanford.edu  
Someshwar, Shonai, UNC Greensboro, s\_somesh@uncg.edu  
Somsong, Sarunya, Srinakharinwirot University & Southern Methodist University, ssomsong@smu.edu  
Song, Dan, University of Iowa, dan-song@uiowa.edu  
Song, Naiqing, Southwest University, songnq@swu.edu.cn  
Soni, Parth Tusharbai, Indian Institute of Management Ahmedabad, phd19parths@iima.ac.in  
Sotelo, Jose de Jesus, ETS, josetoselo2023@u.northwestern.edu  
Soto, Consuelo Andrea, consuelo.soto.allende94@gmail.com  
Southerland, Joshua, Pearson, josh.southerland@pearson.com  
Soyoye, Olushola Olufemi, University of Delaware, olusholasoyoye@gmail.com  
Sparks, Anthony, American Board of OBGYN, asparks@smu.edu  
Sparks, Jesse R., ETS, jsparks@ets.org  
Spikings, Desiree, NWEA, desiree.spikings@nwea.org  
Stark, Stephen, University of South Florida, sestark@usf.edu  
Steed, Teneka, UNC Greensboro, tcsteed@uncg.edu  
Steedle, Jeffrey, Curriculum Associates, jtsteedle@gmail.com  
Stethen, Kelley, ACS Ventures, LLC, kelleyrwheeler@gmail.com  
Stewart, Sheridan, Harvard, sheridan\_stewart@gse.harvard.edu  
Stickney, Eric, Renaissance, Eric.Stickney@renaissance.com  
Still, Cathryn, ELPA21, still@cresst.org  
Stopek, Joshua, AICPA, Joshua.stopek@aicpa-cima.com  
Strietholt, Rolf, IEA Hamburg, rolf.strietholt@iea-hamburg.de

## PARTICIPANT EMAIL

(Last Name, First Name, Affiliation, Email)

Student, Sanford R, University of Delaware, srstu@udel.edu  
Su, YuLan, Center for Applied Linguistics, YuLan.Su@ascendlearning.com  
Suarez-Alvarez, Javier, University of Massachusetts Amherst, suarezj@umass.edu  
Subedi, Dipendra, Pearson, dipendra.subedi@pearson.com  
Subhiyah, Raja G., National Board of Medical Examiners, rsubhiyah@nbme.org  
Sudheesh, Athul, University of Texas at Dallas, Athul.Sudheesh@utdallas.edu  
Suh, Hongwook, Cambium Assessment, Inc., hongwooks@gmail.com  
Suh, Yon Soo, NWEA, yon.soo.suh@nwea.org  
Suk, Youmi, Teachers College Columbia University, ysuk@tc.columbia.edu  
Sul, David, Sul & Associates International, dsul@sulandassociates.com  
Suleiman, Mahmoud, The Pharmacy Examining Board of Canada, msuleiman@pebc.ca  
Sun, Huaping, American Board of Anesthesiologists, sonkahe@hotmail.com  
Sun, Ting, Ting.Sun@hsc.utah.edu  
Sun, Xiaojian, sun.xiaojian@outlook.com  
Sung, Youjin, University of Maryland, College Park, yjsung@umd.edu  
Supovitz, Jonathan, University of Pennsylvania, jons@upenn.edu  
Sussman, Joshua, University of California, Berkeley, jsussman@berkeley.edu  
Svetina Valdivia, Dubravka, Indiana University, dsvetina@indiana.edu  
Swain, Matthew, American Board of Internal Medicine, mswain@abim.org  
Sweeney, Sandra, Cognia, sandra.sweeney@cognia.org  
Swygert, Kimberly, National Board of Medical Examiners, kswygert@nbme.org  
Talreja, Vinita, AWS, vgtalrej@amazon.com  
Tan, Bin, University of Alberta, btan4@ualberta.ca  
Tan, Cheng Yong, The University of Hong Kong, tancy@hku.hk  
Tan, Ling, Australian Council for Educational Research, Ling.Tan@acer.org  
Tan, Yanyan, Pearson, yt58945@uga.edu  
Tang, Alexander, University of Hawaii at Mānoa, aftang@hawaii.edu  
Tang, Nai-En, National Board of Chiropractic Examiners, ntang@nbce.org  
Tang, Xiuxiu, Purdue University, tang469@purdue.edu  
Tao, Sisi, The University of Hong Kong, Faculty of Education, tss2046@hotmail.com  
Taylor, Catherine, ctaylor@uw.edu  
Taylor, CJ, AWS, etaylorc@amazon.com  
Taylor, Darius D., UMASS Amherst, ddtaylor@umass.edu  
Templin, Jonathan, University of Iowa, jonathan-templin@uiowa.edu  
Tenison, Caitlin, ETS, ctenison@ets.org  
Thacker, Arthur, HumRRO, athacker@humrro.org  
Thier, Ben, Cognitive ToyBox, ben@cognitivetoybox.com  
Thissen, David, University of North Carolina, dthissen@email.unc.edu  
Thompson, W. Jake, University of Kansas, wjakethompson@gmail.com  
Thurlow, Martha, NATIONAL CENTER ON EDUCATIONAL, thurl001@umn.edu  
Tian, Wei, Beijing Normal University, tianwei65396@163.com  
Todd, Jessica Andrews, ETS, jandrewstodd@ets.org  
Tong, Ye, yetong@nbme.org  
Toton, Sarah Linnea, Caveon Test Security, sarah.toton@caveon.com  
Toutkoushian, Emily Karen, The American Board of Anesthesiology, toutkoushian.1@gmail.com  
Traynor, Anne, Purdue University, atraynor@purdue.edu  
Trüb, Ruth, FHNW School of Education, ruth.trueb@fhnw.ch  
Tsai, Chia-Lin, University of Northern Colorado, chialin.tsai@unco.edu  
Tsey, Enoch, enoch.tsey001@stu.ucc.edu.gh  
Tu, Naidan, University of South Florida, naidantu@usf.edu  
Tyack, Lillian, Boston College, tyack@bc.edu  
Ulitzsch, Esther, University of Oslo, ulitzsch@inp.uni-kiel.de  
Valdivia Medinaceli, Montserrat B, UC DAVIS, montse.bea.v.m@gmail.com  
Van der Ark, Andries, RESEARCH INSTITUTE OF CHILD DE, L.A.vanderArk@uva.nl  
Vanlwaarden, Adam, Center for Assessment, avaniwaadren@nceia.org  
Van Orman, Dustin, Western Washington University, vanormd2@wwu.edu

## PARTICIPANT EMAIL

(Last Name, First Name, Affiliation, Email)

Van Rijn, Peter, ETS GLOBAL, pvanrijn@etsglobal.org  
Van Schijndel, Tessa, University of Amsterdam, T.J.P.vanSchijndel@uva.nl  
Vasu, Aveline A., National University of Singapore, paeaav@nus.edu.sg  
Vele, Veronica, Australian Dental Council, veronica.vele@adc.org.au  
Ventura, Claudia J., University of Connecticut, claudia.j.ventura@uconn.edu  
Verkuilen, Jay, City University of New York GR, jverkuilen@gc.cuny.edu  
Vispoel, Walter, University of Iowa, walter-vispoel@uiowa.edu  
Vivas Corrales, Pablo, University of Minnesota, vivas014@umn.edu  
Vo, Thao, Washington State University, thao.vo@wsu.edu  
Vo, Yen, University of Iowa, yen-vo@uiowa.edu  
Von Davier, Alina A., Duolingo, avondavier@duolingo.com  
Von Davier, Matthias, Boston College, vondavim@bc.edu  
Wackerle-Hollman, Alisha, University of Minnesota, wacke020@umn.edu  
Wainer, Howard, National Board of Medical Examiners, hwainer@nbme.org  
Walker, Marcus, National Commission on Certification of Physician Assistants, marcusw@nccpa.net  
Walker, Michael E., memwalker@gmail.com  
Wall, Nathan, EMETRIC, nwall@emetric.net  
Wan, Lei, College Board, lwan@collegeboard.org  
Wan, Siyu, ABIM, siyuwan93@outlook.com  
Wang, Aijun, FSBPT, wajlm2003@gmail.com  
Wang, Bowen, University of Florida, bowen.wang@ufl.edu  
Wang, Changjiang, PEARSON, Changjiang.Wang@Pearson.com  
Wang, Chuang, University of North Carolina at Charlotte, cwang15@uncc.edu  
Wang, Chun, University of Washington, wang4066@uw.edu  
Wang, Chunxin, ACT INC., annwang728@gmail.com  
Wang, Dongwei, UMass Amherst, dongweiwang@umass.edu  
Wang, Hongling, ACT, INC., hongling.wang@act.org  
Wang, Jue, University of Science and Technology of China, juewang01@ustc.edu.cn  
Wang, Kuo, Southern Methodist University, wangp@smu.edu  
Wang, Min, College Board, wangminemail@gmail.com  
Wang, Nixi, University of Washington, niciblessed@gmail.com  
Wang, Shiyu, University of Georgia, swang44@uga.edu  
Wang, Shuang, Beijing Normal University, marvelousariel@gmail.com  
Wang, Songtao, OISE/University of Toronto, songtao.wang@mail.utoronto.ca  
Wang, Xi, The University of Iowa, xwang405@uiowa.edu  
Wang, Xiyu, Purdue University, wang5608@purdue.edu  
Wang, Xuan, National Conference of Bar Examiners, xwang@ncbex.org  
Wang, Yang Caroline, Education Analytics, cwang@edanalytics.org  
Wang, Yu, University of Minnesota - Twin Cities, wang7919@umn.edu  
Wang, Yuan, ETS, ywang@ets.org  
Wang, Zhen, Cambium Assessment, zhen.wang@cambiumassessment.com  
Wang, Zhuoran, National Council of State Boards of Nursing (NCSBN), wzhranran@gmail.com  
Warner, Zachary, New York State Education Department, zachary.warner@nysed.gov  
Waters, Charlotte, Australian Council for Educational Research, Charlotte.Waters@acer.org  
Watts, Ashley, Vanderbilt University, ashley.watts@vanderbilt.edu  
Webb, Blue, American Institutes for Research, bwebb@air.org  
Weeks, Jonathan, ETS, jweeks@ets.org  
Wei, Hsin-Ro, Riverside Insights, Hsin-ro.wei@riversideinsights.com  
Weiner, Drew, National Conference of Bar Examiners, dweiner@ncbex.org  
Weiner, John, Lifelong Learner Holdings, john.weiner@lifelonglearner.com  
Weiss, David, University of Minnesota, djweiss@umn.edu  
Welch, Catherine, University of Iowa, catherine-welch@uiowa.edu  
Wells, Craig, UMass Amherst, cswells@umass.edu  
Wendel, George, American Board of Obstetrics and Gynecology, gwendel@abog.org  
Westrick, Paul, College Board, westrick\_3@hotmail.com  
Wheeler, Jordan, University of Nebraska Lincoln, jwheeler21@unl.edu

## PARTICIPANT EMAIL

(Last Name, First Name, Affiliation, Email)

White, Lauren, Peason, law03k@gmail.com  
Whitfield, Erik, University of Colorado Boulder, Erik.Whitfield@colorado.edu  
Whitmer, John, Federation of American Scientists, jwhitmer@fas.org  
Whitney, Sarah, NWEA, sarah.whitney@nwea.org  
Williams, Kevin, ETS, kmwilliams@ets.org  
Williams, Natasha Jayne, Curriculum Associates, njwilliams087@gmail.com  
Williamson, Shelby, National Strength and Conditioning Association, shelby.williamson@nsca.com  
Willse, Jada, University of North Carolina, jadagw@email.unc.edu  
Wilmurth, Gina, NWEA, gina.wilmurth@nwea.org  
Wilson, Joshua, University of Delaware, joshwils@udel.edu  
Wilson, Mark, Berkeley School of Education, University of California, Berkeley, markw@berkeley.edu  
Wilson, Nicole, James Madison University, wilso2rn@jmu.edu  
Wind, Stefanie A., University of Alabama, swind@ua.edu  
Winter, Sonja D., University of Missouri, Columbia, sdwinter@missouri.edu  
Wise, Steven, stevewise23@gmail.com  
Withycombe, Adam, NWEA, Adam.Withycombe@nwea.org  
Wolf, Benjamin, University of Zurich, Benjamin.Wolf@ibe.uzh.ch  
Wolfe, Edward W., Pearson, ed.wolfe@pearson.com  
Wolff, Sarah, UNLV, sarah.wolff@unlv.edu  
Wollack, James, University of Wisconsin, jwollack@wisc.edu  
Wong, Yun Leng, wong0620@umn.edu  
Wongvorachan, Tarid, University of Alberta, wongvora@ualberta.ca  
Woo, Ada, Ascend Learning, adawoo811@gmail.com  
Woo, Yejin, Ewha Womans University, wooye6093@ewhain.net  
Woodhead, Simon, Eedi, simon.woodhead@eedi.co.uk  
Woods, Scott, ACT, Inc., scott.woods@act.org  
Wolf, Sherri, Consultant, sherriwoolf@gmail.com  
Workman, Trent, Pearson, Trent.Workman@Pearson.com  
Worthington, Michelle, ETS, mworthington@ets.org  
Wright, Daniel B., UNLV, dbrookswr@gmail.com  
Wu, Hao, Vanderbilt University, wu.498@osu.edu  
Wu, Lizzy, University of Illinois, Champaign, lixinwu2@illinois.edu  
Wu, Tong, Riverside Insights, tong.wu@riversideinsights.com  
Wu, Yi-Fang, Cambium Assessment, Inc., wuyifang91@gmail.com  
Wu, Zebing, Cambium Assessment, zebing.wu@cambiumassessment.com  
Wuyts, Celine, OECD, celine.wuyts@oecd.org  
Wylie, E. Caroline, Center for Assessment, ecwylie@nciaea.org  
Wyse, Adam E., Renaissance, adam.wyse@renaissance.com  
Xia, Yan, yx18@illinois.edu  
Xiao, Xingyao, Berkeley School of Education, University of California, Berkeley, xiaoxg@berkeley.edu  
Xie, Qing, University of Iowa, xieqingemma@gmail.com  
Xin, Tao, Beijing Normal University, xintao@bnu.edu.cn  
Xiong, Jiawei, Pearson, jiawei.xiong@pearson.com  
Xiong, Xinhui Maggie, ETS, maggiexio8888@gmail.com  
Xiong, Yimei, Purdue University, xiong137@purdue.edu  
Xu, Xin, Beijing Normal University, xux636@nenu.edu.cn  
Xu, Yangmeng, University of Alabama, yxu81@crimson.ua.edu  
Xue, Mingfeng, University of California - Berkeley, mingfengxue@berkeley.edu  
Yan, Duanli, ETS, dyan@ets.org  
Yancey, Kevin, Duolingo, kyancey@duolingo.com  
Yaneva, Victoria, NBME, vyaneva@nbme.org  
Yang, Huidi, NYU, hy1331@nyu.edu  
Yang, Ji Seung, University of Maryland, jsyang@umd.edu  
Yang, Xiangdong, Department of Educational Psychology, Faculty of Education, East China Normal University, xdyang@dep.ecnu.edu.cn  
Yang, Xiaotong, Florida State University, xy15@my.fsu.edu

## PARTICIPANT EMAIL

(Last Name, First Name, Affiliation, Email)

Yang, Yanyun, Florida State University, yyang3@admin.fsu.edu  
Yao, Lihua, Northwestern University, lihua.yao@northwestern.edu  
Yavuz, Sinan, University of Wisconsin-Madison, syavuz@wisc.edu  
Ye, Daisy, Houghton Mifflin Harcourt, daisy.ye@hmhco.com  
Ye, Shaojie, Pinterest, sye@pinterest.com  
Yen, Shu Jing, Center for Applied Linguistics, syen@cal.org  
Yi, Karen, American Institutes for Research, kyi@air.org  
Yildirim-Erbasli, Seyma N., Concordia University of Edmonton, seyma.yildirim-erbasli@concordia.ab.ca  
Yin, Yue, yueyin@uic.edu  
Yoo, Hanwook, Ascend Learning, hanuki82@gmail.com  
Yoon, Hyejin, Michigan State University, hyoon12@gmu.edu  
Yu, Haoyang, Accessible Teaching, Learning, and Assessment Systems (ATLAS) at Achievement & Assessment Institute, haoyang.yu@ku.edu  
Yuan, Ye, University of Georgia, yeyuan0106@gmail.com  
Zapata-Rivera, Diego, ETS, dzapata@ets.org  
Zavitkovsky, Paul, Center for Urban Education Leadership, University of Illinois at Chicago, pzavit@uic.edu  
Zechner, Klaus, ETS, kzechner@ets.org  
Zegers, Mónica, University of California, San Francisco, Monica.ZegersLarrain@ucsf.edu  
Zehner, Fabian, DIPF | Leibniz Institute for Research and Information in Education, Centre f. Int. Student Assessm., fabian.zehner@dipf.de  
Zeng, Biao, Beijing Normal University, biaozen@gmail.com  
Zeng, Ji, Michigan Department of Education, zengj@michigan.gov  
Zenisky, April, University of Massachusetts Amherst, azenisky@educ.umass.edu  
Zhai, Xiaoming, University of Georgia, Xiaoming.Zhai@uga.edu  
Zhan, Peida, pdzhan@gmail.com  
Zhang, Anna Yinqi, Penn State University, ybz5148@psu.edu  
Zhang, Bo, UIUC, bozhang3@illinois.edu  
Zhang, Danhui, Beijing Normal University, danhuizhang@bnu.edu.cn  
Zhang, Jin, azjemail67@gmail.com  
Zhang, Jinming, University of Illinois at Urbana-Champaign, jmzhang@illinois.edu  
Zhang, Lijin, Stanford Graduate School of Education, lijinzhang@stanford.edu  
Zhang, Litong, DRC, LZhang@DataRecognitionCorp.com  
Zhang, Mengyao, National Conference of Bar Examiners, mzhang@ncbex.org  
Zhang, Mingqin, National Council of State Boards of Nursing (NCSBN), mqziowa@gmail.com  
Zhang, Mo, ETS, mzhang@ets.org  
Zhang, Oliver, College Board, ozhang@collegeboard.org  
Zhang, Susu, University of Illinois at Urbana-Champaign, szhan105@illinois.edu  
Zhang, Ting, American Institutes for Research, tzhang@air.org  
Zhang, Xiao, DRC, xzhang@datarecognitioncorp.com  
Zhang, Yichi, University of Southern California, yzhang97@usc.edu  
Zhang, Yu, Federation of State Boards Of Physical Therapy, yzhang@fsbpt.org  
zhang, yuan, American Institutes for Research, yzhang@air.org  
Zhang, Yuxiao, Purdue University, zhan3971@purdue.edu  
Zhang, Zhonghua, University of Melbourne, zhonghua.zhang@unimelb.edu.au  
Zheng, Xiaying, American Institutes for Research, xzheng@air.org  
Zheng, Yi, Arizona State University, yi.isabel.zheng@asu.edu  
Zhong, Xiaoting, University of Iowa, xiaoting-zhong@uiowa.edu  
Zhong, Yang, University of Pittsburgh, YA2118@pitt.edu  
Zhou, Xinchang, UIUC, xz77@illinois.edu  
Zhou, Xuechun, Ascend Learning, xuechun.zhou@ascendlearning.com  
Zhu, Danqi, Fordham University, dzhu17@fordham.edu  
Zlatkin-Troitschanskaia, Olga, Johannes Gutenberg University, troitschanskaia@uni-mainz.de  
Zopluoglu, Cengiz, University of Oregon, cengiz@uoregon.edu  
Zou, Chunrui, University of Pennsylvania, Chunrui.Zou@Pennmedicine.upenn.edu  
Zou, Tongtong, Columbia University, tz2345@tc.columbia.edu  
Zou, Yuqing, University of Iowa, yuqing-zou@uiowa.edu

## **PARTICIPANT EMAIL**

(Last Name, First Name, Affiliation, Email)

Zoucha, James, University of Northern Colorado, zouc5751@bears.unco.edu  
Zu, Jiyun, ETS, jzu@ets.org

National Council on Measurement in Education is very grateful to the following organizations for their generous financial support of our 2024 Annual Meeting.

## PLATINUM



## GOLD



## SILVER



## FRIEND

