

NCME VIRTUAL ANNUAL MEETING

SCHEDULE AT A GLANCE

WEDNESDAY, SEPTEMBER 9

11:00 AM – 12:30 PM ET	Psychometric Innovations and Advances in Medical Educational Assessments	<p><u>Participants:</u></p> <p>Helping Item Writers by Making NLP-Based Suggestions for Item Distractors Peter Baldwin, National Board of Medical Examiners (NBME); Victoria Yaneva, National Board of Medical Examiners; Janet Mee, National Board of Medical Examiners; Brian Clouser, National Board of Medical Examiners; Le An Ha, University of Wolverhampton</p> <p>On the Utility of Using Transfer Learning to Predict Item Characteristics Kang Xue, University of Georgia; Victoria Yaneva, National Board of Medical Examiners; Christopher Runyon, NBME</p> <p>Exploring Automated Assessment of Spoken English Proficiency for Medical Licensure Exams Su G. Somay, National Board of Medical Examiners (NBME); Jessica Salt, Education Commission for Foreign Medical Graduates (ECFMG)</p> <p>An Automated Scoring Routine for Constructed Responses on a Medical Licensure Exam Christopher Runyon, NBME; Polina Harik, National Board of Medical Examiners; Abeed Sarker, Emory University School of Medicine; Graciela Gonzalez-Hernandez, University of Pennsylvania</p> <p>Identifying Response Pattern in Clinical Notes with NLP Feature Identification Janet Mee, National Board of Medical Examiners; Ravi Pandian, National Board of Medical Examiners; Andrew Houriet, National Board of Medical Examiners; Christopher Yang, Drexel University</p> <p>IRTTree Response Style Modeling for Improving Feedbacks about Cognitive Biases Lale Khorramdel, National Board of Medical Examiners; Matthias von Davier, National Board of Medical Examiners; Ann King, National Board of Medical Examiners; Andrew Houriet, National Board of Medical Examiners</p> <p><u>Session Organizers:</u></p> <p>Lale Khorramdel, National Board of Medical Examiners</p> <p>Christopher Runyon, NBME</p> <p><u>Chair:</u></p> <p>Lale Khorramdel, National Board of Medical Examiners</p> <p><u>Discussant:</u></p> <p>Isaac I Bejar, Educational Testing Service</p>
1:00 – 2:30 PM ET	Fireside Chat with Classroom Assessment Task Force: Making Measurement Meaningful with Classroom Assessment	<p><u>Presenters:</u></p> <p>Sue Brookhart, Duquesne University</p> <p>Heidi Andrade, SUNY Albany</p> <p>Mark Wilson, University of California, Berkeley</p>

		<p>E Caroline Wylie, Educational Testing Service</p> <p>Jade Caines Lee, University of New Hampshire</p> <p>Alison Bailey, UCLA</p> <p>Neal Kingston, University of Kansas</p> <p>Dale Whittington, Retired</p> <p><u>Session Organizer:</u></p> <p>Kristen Huff, Curriculum Associates</p>
5:00 – 6:30 PM ET	Using Psychometric Approaches to Improve Test Development	<p><u>Participants:</u></p> <p>Developing Useful Test Blueprints: An Evaluation Study Andrew C. Dwyer, American Board of Pediatrics; Robert Furter, American Board of Pediatrics</p> <p>One Blueprint Fits All? Using Person-centered Methodological Approaches in Practice Analysis Studies Pamela Kaliski, American Board of Internal Medicine; Kelly Foelber, ABIM; Jerome Clouser, ABIM</p> <p>The Effect of Item Writing Experience on Item Quality Bozhidar M. Bashkov, American Board of Internal Medicine; Jerome Clouser, ABIM</p> <p>Modeling Text Complexity Judgments and Their Association with Item Difficulty Brian F. Patterson, Curriculum Associates; Pamela Seastrand, Curriculum Associates</p> <p>Using Item Text to Predict Item Survival: Can It Go Beyond Linguistic Characteristics Victoria Yaneva, National Board of Medical Examiners; Le An Ha, University of Wolverhampton; Peter Baldwin, National Board of Medical Examiners (NBME); Janet Mee, National Board of Medical Examiners</p> <p><u>Session Organizer:</u></p> <p>Bozhidar M. Bashkov, American Board of Internal Medicine</p> <p><u>Discussant:</u></p> <p>Kurt F. Geisinger, Buros Center for Testing</p>
7:30 – 9:00 PM ET	NCME Social Event: Trivia <i>Pre-registration required;</i> <i>N<= 150</i>	<p>Join us for a virtual game night with “The Go Remote: Virtual Game Show!”</p> <p>HOW IT WORKS: Grab some headphones, a beverage of your choice, and your game face (yes-you will want to turn on your laptop video for this!) for an action-packed Happy Hour Game Show. Compete in live games such as: Trivia, Pictionary, Name that Movie/Tune, and GIF Battles. This session will be hosted outside of the event platform. Pre-registration required. Sign up here.</p>

THURSDAY, SEPTEMBER 10

9:00 – 10:30 AM ET	Assessing Collaborative Problem Solving at Scale: The Status Quo and the Next	<p><u>Participants:</u></p> <p>CPS assessment in ATC21S - learning from the past <i>Esther Care, Brookings Institution</i></p> <p>Updates on CPS Assessment in PISA 2015 <i>Art Graesser, University of Memphis</i></p> <p>Assessing CPS: Conceptualization and Methodological Differences between the CSCL and Assessment Communities <i>Nancy Law, University of Hongkong</i></p> <p>Learning & Assessment of Collaborative Problem Solving at ACT <i>Alina von Davier, ACTNext; Kristin Stoeffler, ACTNext by ACT; Benjamin Deonovic, ACTNext by ACT; Michael Yudelson, ACTNext by ACT; Pravin Chopade, ACTNext by ACT; David Edwards, ACTNext by ACT; Saad Khan, ACTNext by ACT; Yigal Rosen, ACTNext by ACT</i></p> <p>ETS' Efforts towards Learning and Assessments of CPS at Scale <i>Jiangang Hao, Educational Testing Service</i></p> <p><u>Session Organizer:</u></p> <p><i>Jiangang Hao, Educational Testing Service</i></p> <p><u>Discussant:</u></p> <p><i>Patrick Kyllonen, Educational Testing Service</i></p>
11:00 AM – 12:30 PM ET	Assessment of Mathematical and Scientific Reasoning: An alternative to Machine-scoring Open-Ended Items	<p><u>Participants:</u></p> <p>Investigating an Alternative to Machine-scoring of Open-Ended Items <i>Mark Wilson, University of California, Berkeley</i></p> <p>Developing selected response items for a data-based decision-making instrument <i>Amy Arneson, Education Northwest</i></p> <p>An Exploration of Selected-Response Items Compared to Constructed-Response Item Types in Science Education <i>Linda Morell, University of California, Berkeley; Weeraphat Suksiri, University of California-Berkeley; Sara Dozier, Stanford University; Jonathan Osborne, Stanford University; Mark Wilson, University of California, Berkeley</i></p> <p>Comparing Selected Response and Constructed Response Items in Mathematical Problem Solving <i>Yukie Toyama, University of California; Jerred Jolin, University of California, Berkeley; James Mason, University of California, Berkeley; Mark Wilson, University of California, Berkeley</i></p> <p><u>Session Organizer:</u></p> <p><i>Mark Wilson, University of California, Berkeley</i></p> <p><u>Chair:</u></p> <p><i>Karen Draney, University of California Berkeley</i></p> <p><u>Discussants:</u></p>

		Richard Patz , University of California-Berkeley Jonathan Osborne , Stanford University
1:00 – 2:00 PM ET	Presidential Address: “Psychometricians in the Hands of an Angry Mob.”	Stephen G. Sireci , University of Massachusetts Amherst Ye Tong , Pearson
2:30 – 4:00 PM ET	Beyond Accommodations: Intentional Design Methods for Improving Validity When Assessing Diverse Populations	<u>Participants:</u> Improving Fairness: Theories of Action and Intentional Design Methods Melissa L. Gholson , Educational Testing Service Equitable Assessment through Twinning Danielle Guzman-Orth , Educational Testing Service; Cary Supalo , Educational Testing Service; Cinda Parton , WestEd Applying Universal Design Principles to the Accessible Test Development Process Robert P Dolan , Diverse Learners Consulting; Cara Wojcik , CAST; Jenna Gravel , CAST; Allison Posey , CAST; Elizabeth Hartmann , Lasell College; Kimberly Ducharme , CAST; Jose Blackorby , CAST <u>Session Organizer:</u> Robert P Dolan , Diverse Learners Consulting <u>Discussant:</u> Sheryl Lazarus , National Center on Educational Outcomes
4:30 – 6:00 PM ET	Principled Item Design: State-of-the-Art	<u>Participants:</u> Test and Item Development with Cognitive Design Systems: Some Examples Susan Embretson , Professor The Interplay of Validity and Efficiency Isaac I Bejar , Educational Testing Service Considerations for Training Subject Matter Experts to Write Item Models Audra Kosh , Edmentum The Challenges of Principled Item Design Richard Luecht , University of North Carolina Automated Item Generation Using Deep Learning Matthias von Davier , National Board of Medical Examiners <u>Session Organizers:</u> Richard Luecht , University of North Carolina Isaac I Bejar , Educational Testing Service <u>Chair:</u> Richard Luecht , University of North Carolina

		<p>Discussant:</p> <p>Steve Ferrara, Measured Progress</p>
6:00 – 7:00 PM ET	DREAM ON! (DRinking, Eating, And Measurement ONline Party)	<p>Stephen G. Sireci, University of Massachusetts Amherst</p> <p>Bruno D. Zumbo, University of British Columbia</p> <p>Saskia Wools, CITO</p>
FRIDAY, SEPTEMBER 11		
11:00 AM – 12:30 PM ET	Feasibility of an Embedded Field-Test Model in the Enhancement of Preequating	<p><u>Participants:</u></p> <p>An Overview of the Embedded Field-Test Model Wei Tao, ACT</p> <p>Evaluating Impacts on Operational Item Performance in the Embedded Field-Test Model Troy Chen, ACT</p> <p>Comparing CTT Postequating and IRT Preequating in the Embedded Field-Test Model Yi-Fang Wu, ACT</p> <p>Investigating Potential Confounding Testing Mode Effects in Transition from Postequating to Preequating Yong He, ACT</p> <p><u>Session Organizer:</u></p> <p>Yi-Fang Wu, ACT</p> <p><u>Chair:</u></p> <p>Yi-Fang Wu, ACT</p> <p><u>Discussant:</u></p> <p>Deborah Harris, University of Iowa</p>
1:00 PM – 2:30 PM ET	Separate but (Un)Equal? Measurement Expertise vs. Policy Knowledge in Testing for Accountability	<p><u>Participants:</u></p> <p>Does the Measurement Community ‘Own’ the Problems on Interpretation and Use of Test Scores? Ellen Forte, edCount, LLC</p> <p>Ensuring that Test Use Leads to the Intended, Positive Consequences, while Minimizing Unintended, Negative Consequences Suzanne Lane, University of Pittsburgh</p> <p>Towards Inclusive Assessments for all Students, and Inclusive Accountability Systems Martha Lynn Thurlow, NCEO/University of Minnesota</p> <p>Mapping the Relationship Between Policy and Measurement Considerations in Supporting Accountability Testing Neal Kingston, University of Kansas</p> <p>Separate and Unequal? Lessons from Selective High Schools Uses of Standardized Test Scores Howard Everson, City University of New York</p>

		<p>Why Assessment Professionals Should Cultivate Policy Analysis Skills Aaron Pallas, <i>Columbia University, Teachers College</i></p> <p>Conflicting Theories of Action? Reconciling Mindsets of Measurement vs. K-12 Stakeholders to Improve Validity and Consequences of Testing Madhabi Chatterji, <i>Columbia University, Teachers College</i></p> <p><u>Session Organizer:</u></p> <p>Madhabi Chatterji, Columbia University, Teachers College</p>
3:00 PM – 4:30 PM ET	New Research Findings on Understanding and Managing Test-Taking Disengagement	<p><u>Participants:</u></p> <p>Aggregate-Level Ability Estimation Accuracy Under Varying Noneffortful Responding Types and Rates Joseph A. Rios, <i>University of Minnesota</i>; Chelsey Legacy, <i>University of Minnesota</i></p> <p>Using Retest Data to Evaluate and Improve Effort-Moderated Scoring Steven L. Wise, <i>NWEA</i>; Megan R. Kuhfeld, <i>NWEA</i></p> <p>How to Design a Drag-and-Drop Item for Motivated Responding Blair Lehman, <i>ETS</i>; Burcu Arslan, <i>ETS</i></p> <p>Test Value and Emotions: Predicting Examinee Effort and Performance on Low-Stakes Tests Paulius Satkus, <i>James Madison University</i>; Sara J. Finney, <i>James Madison University</i></p> <p>Do Students Rapidly Guess Repeatedly Over Time? A Longitudinal Analysis James Soland, <i>NWEA</i>; Megan R. Kuhfeld, <i>NWEA</i></p> <p><u>Session Organizer:</u></p> <p>Steven L. Wise, <i>NWEA</i></p> <p><u>Chair:</u></p> <p>Dena Pastor, <i>James Madison University</i></p>
5 PM ET – 6:30 PM ET	Insights on Text-to-Speech as a Universal Design Feature: NAEP Mathematics Process Data	<p><u>Participants:</u></p> <p>Evaluation of the Text-to-Speech (TTS) Tool Using NAEP Process Data Soo Lee, <i>American Institutes for Research</i></p> <p>Exploration of TTS Using the Differential Boost Framework Juanita Hicks, <i>American Institutes for Research</i></p> <p>Exploring Read-Aloud Sentences to Understand Learners' TTS Use Tendency Ruhan Ciri, <i>AIR</i></p> <p><u>Session Organizer:</u></p> <p>Juanita Hicks, <i>American Institutes for Research</i></p> <p><u>Discussant:</u></p> <p>Heather Buzick, <i>Educational Testing Service</i></p>
6:30 PM – 7:30 PM ET	GSIC Social Event	<p><u>Hosts:</u></p> <p>Maura O'Riordan, <i>University of Massachusetts Amherst</i></p>

Scott Holcomb, UNC Charlotte

For the full line up, including sessions descriptions and call in information, visit the online portal – [CLICK HERE!](#)