AIME-Con 2025

Artificial Intelligence in Measurement and Education Conference (AIME-Con)

Volume 3: Coordinated Session Papers

October 27-29, 2025

The AIME-Con organizers gratefully acknowledge the support from the following sponsors.

Platinum

Gold

Silver

Supporters


FOCALPOINT
Teaching & Learning. Measured. Managed


MEASUREMENT
INCORPORATED


University of
Pittsburgh
Learning Research and
Development Center


WestEd
WestEd.org

duolingo english test

The future of language assessment is here

The Duolingo English Test is a computer adaptive test powered human-in-the-loop AI and supported by rigorous validity research. The test measures speaking, writing, reading, and listening skills, providing a deeper insight into English proficiency.

Built on the latest language assessment science

- Accessible by design, supporting test takers wherever they are for just $70
- Built on rigorous research and industry- leading security
- Integrates the latest assessment science and AI for accurate results
- Accepted by over 5,800 programs worldwide

englishtest.duolingo.com

Evidence-based approach to AI in Measurement & Learning

At the intersection of artificial intelligence and educational measurement, Pearson stands as your trusted partner—delivering clarity, confidence, and innovation in every assessment moment.

Why Pearson?

- AI-Enhanced Accuracy: Using automated scoring and predictive analytics to provide insights that are accurate, fair, and timely.
- Future-Ready Solutions: Platforms that evolve with policy, pedagogy, and technology.
- Personalized Learning Journeys: Multi-lingual access and adaptive item generation to support each student's unique growth trajectory.
- Ethical AI Practices: Commitment to data security, transparency, explainability, and bias mitigation.
- Collaborative Innovation: Partnering with educators, researchers, and technologists to shape the future of assessment.

Human-Centric AI	Pearson believes AI's highest purpose is to elevate and empower human capabilities.
Assessment as a Learning Continuum	We reimagine assessments not as endpoints, but as integral parts of the learning journey.
AI as an Environment	Pearson is exploring how this shift impacts our approach to assessment—ensuring our tools are adaptive and future-ready.
Balancing Vision and Capabilities	We deliver reliable solutions today while building toward the future of AI in education.

Pearson

Order copies of this and other NCME proceedings from:

National Council on Measurement in Education (NCME)
520 S. Walnut St. Box 2388
Bloomington, IN 47402
USA
Tel: +1-812-245-8096
ncme@ncme.org

ISBN 979-8-218-84230-7

Preface

Introduction

The inaugural NCME-sponsored Artificial Intelligence in Measurement and Education Conference (AIME-Con) brought together an interdisciplinary community of experts working at the intersection of artificial intelligence (AI), educational measurement, assessment, natural language processing, learning analytics, and technological development. As AI continues to transform education and assessment practices, this conference provided a critical platform for fostering cross-disciplinary dialogue, sharing cutting-edge research, and exploring the technical, ethical, and practical implications of AI-driven innovations in measurement and education. By bringing together experts from varied domains, the conference fostered a rich exchange of knowledge to enhance the collective understanding of AI's impact on educational measurement and evaluation.

Conference Theme - Innovation and Evidence: Shaping the Future of AI in Educational Measurement

The NCME-Sponsored AIME-Con focused on how rigorous measurement standards and innovative AI applications can work together to transform education. With sessions spanning summative large-scale assessment, formative classroom assessment, automated feedback, and informal learning tools, this conference fostered both the advancement and evaluation of AI technologies that are effective, reliable, and fair.

The National Council on Measurement in Education

The National Council on Measurement in Education is a community of measurement scientists and practitioners who work together to advance theory and applications of educational measurement to benefit society. A professional organization for individuals involved in assessment, evaluation, testing, and other aspects of educational measurement, our members are involved in the construction and use of standardized tests; new forms of assessment, including performance-based assessment; program design; and program evaluation. Learn more about NCME, including our goals and our leadership, at www.ncme.org. We are grateful to the NCME.

NCME Special Interest Group on Artificial Intelligence in Measurement and Education

The AIME SIGIMIE seeks to advance the theoretical and applied research into AI of educational measurement by bringing together data scientists, psychometricians, education researchers, and other interested stakeholders. The SIGIMIE will discuss current practices in using Generative AI, approaches to evaluate their precision/accuracy, and areas where more foundational research is required into the way we test and measure educational outcomes. This group seeks to create a strong professional identity and intellectual home for those interested in the use of AI in many areas, including automated scoring, item evaluation, validity studies, formative feedback, and generative AI for automated item generation.

Proposal Requirements and Review Process for Coordinated Paper Sessions

AIME-Con invited submissions of coordinated paper sessions, which brought together 4–5 papers on a common theme within a 90-minute session. Each proposal included both session-level information (title, abstract, keywords, chair/moderator, and discussant where applicable) and paper-level details (title, short abstract, topic of interest, and either a 1,000-word structured summary or a six-page paper). **All contributors were identified at submission, as the review process was not blind.**

Submissions were evaluated by members of the review committee using a rubric that evaluated the following dimensions:

- **Relevance and community impact:** pertinence to the AI in measurement and education community, and potential contribution to current discussions and challenges in the field
- **Significance and value:** scholarly merit or practical importance of the work, and potential impact on theory, practice, or policy
- **Methodological rigor:** coherence and appropriateness of the proposed methods, techniques, and approaches; and soundness of the overall research design
- **Quality of expected outcomes:** whether the proposed analysis and interpretation methods are appropriate, and the potential contribution to knowledge in the field
- **Feasibility and timeline:** the realistic likelihood that the proposed work can be completed by the conference date

For the purposes of this conference, "AI" was defined broadly to include rule-based methods, machine learning, natural language processing, and generative AI/large language models. Reviewers provided constructive feedback and overall recommendations to ensure that accepted sessions reflected both scholarly merit and practical value to the AI in measurement and education community.

Organizing Committee

NCME Leadership

Amy Hendrickson, Ph.D. (President)
Rich Patz, Ph.D. (Executive Director)

Conference Chairs

Joshua Wilson, University of Delaware
Christopher Ormerod, Cambium Assessment
Magdalen Beiting Parrish, Federation of American Scientists

Proceedings Chair

Nitin Madnani, Duolingo

Proceedings Committee

Jill Burstein, Duolingo
Polina Harik, NBME

Program Committee

Conference Chairs

Joshua Wilson, University of Delaware
Christopher Ormerod, Cambium Assessment
Magdalen Beiting Parrish, Federation of American Scientists

Reviewers

Hope Adegoke, University of North Carolina, Greensboro
Magdalen Beiting-Parrish, Federation of American Scientists
Peter Foltz, University of Colorado, Boulder
Hudson Golino, University of Virginia
Hongli Li, Georgia State University
Sheng Li, University of Virginia
Jianyuan Ni, Juniata College
Christopher Ormerod, Cambium Assessment
Corey Palermo, Measurement Incorporated
Shaila Quazi, Drexel University
Andrew Runge, Duolingo
Christopher Runyon, National Board of Medical Examiners
Raashi Sangwan, Miller School of Medicine, University of Miami
Khem Sedhai, University of Albany
Mark Shermis, Performance Assessment Analytics, LLC
Kimberly Swygert, National Board of Medical Examiners
Joshua Wilson, University of Delaware
Jiawei Xiong, University of Georgia
Xinhui Maggie Xiong, ExamRoom AI

Table of Contents