

Artificial Intelligence in Measurement and Education Conference (AIME-Con 2025)

Volume 3: Coordinated Session Papers

Pittsburgh, Pennsylvania, USA
27-29 October 2025

ISBN: 979-8-3313-2856-6

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2025) by the Association for Computational Linguistics
All rights reserved.

Printed with permission by Curran Associates, Inc. (2026)

For permission requests, please contact the Association for Computational Linguistics
at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006
Fax: 1-570-476-0860

acl@aclweb.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2633
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

<i>When Does Active Learning Actually Help? Empirical Insights with Transformer-based Automated Scoring</i>	
Justin O Barber, Michael P. Hemenway and Edward Wolfe	1
<i>Automated Essay Scoring Incorporating Annotations from Automated Feedback Systems</i>	
Christopher Ormerod	9
<i>Text-Based Approaches to Item Alignment to Content Standards in Large-Scale Reading & Writing Tests</i>	
Yanbin Fu, Hong Jiao, Tianyi Zhou, Nan Zhang, Ming Li, Qingshu Xu, Sydney Peters and Robert W Lissitz	19
<i>Review of Text-Based Approaches to Item Difficulty Modeling in Large-Scale Assessments</i>	
Sydney Peters, Nan Zhang, Hong Jiao, Ming Li and Tianyi Zhou	37
<i>Item Difficulty Modeling Using Fine-Tuned Small and Large Language Models</i>	
Ming Li, Hong Jiao, Tianyi Zhou, Nan Zhang, Sydney Peters and Robert W Lissitz	48
<i>Operational Alignment of Confidence-Based Flagging Methods in Automated Scoring</i>	
Corey Palermo, Troy Chen and Arianto Wibowo	56
<i>Pre-Pilot Optimization of Conversation-Based Assessment Items Using Synthetic Response Data</i>	
Tyler Burleigh, Jing Chen and Kristen Dicerbo	61
<i>When Humans Can't Agree, Neither Can Machines: The Promise and Pitfalls of LLMs for Formative Literacy Assessment</i>	
Owen Henkel, Kirk Vanacore and Bill Roberts	69
<i>Beyond the Hint: Using Self-Critique to Constrain LLM Feedback in Conversation-Based Assessment</i>	
Tyler Burleigh, Jenny Han and Kristen Dicerbo	79
<i>Investigating Adversarial Robustness in LLM-based AES</i>	
Renjith Ravindran and Ikkyu Choi	86
<i>Effects of Generation Model on Detecting AI-generated Essays in a Writing Test</i>	
Jiyun Zu, Michael Fauss and Chen Li	92
<i>Exploring the Interpretability of AI-Generated Response Detection with Probing</i>	
Ikkyu Choi and Jiyun Zu	99
<i>A Fairness-Promoting Detection Objective With Applications in AI-Assisted Test Security</i>	
Michael Fauss and Ikkyu Choi	107
<i>The Impact of an NLP-Based Writing Tool on Student Writing</i>	
Karthik Sairam, Amy Burkhardt and Susan Lottridge	115