

AI Research Library & Resource Hub

Artificial Intelligence (AI) is rapidly transforming language learning and assessment. This page brings together AAPPL presentations, publications, and resources that explore how AI can support world language educators while maintaining high standards for language proficiency assessment.

1. Automated Scoring of Language Assessments

How AI is used to score Spanish AAPPL speaking and writing tasks.

Publications

Voss, E., Sallee, K., Son, Y. A., Malone, M. E., Marshall, C., & Chomon-Zamora, C. (2026). An Automated Scoring System for the AAPPL Spanish Presentational Writing Tasks. *Foreign Language Annals*, 59(1), 63-81.

Presentations

Automated Scoring System for a Spanish Writing Test: Feature-Engineered Machine Learning vs. Large Language Model Approach.

2024 AAAL Paper presentation.

Erik Voss, Young-A Son. Houston, TX. March 16, 2024.

Abstract:

This presentation compares two approaches to automated scoring for Spanish writing assessments: traditional feature-engineered machine learning and emerging large language models (LLMs). The presenters examine how each method performs in scoring written responses, highlighting differences in accuracy, reliability, and practical implementation. The session explores the potential of AI-driven scoring technologies to support valid, consistent, and scalable language assessment while identifying important considerations for their use in educational settings.

An Argument for an Automated Scoring System for Spanish Written Responses: Supporting an Evaluation Inference.

2023 ECOLT Conference Presentation.

Erik Voss, Young-A Son. Georgetown University, Washington, D.C., October 27, 2023.

Abstract:

This presentation examines the evidence needed to support the use of an automated scoring system for Spanish written responses. Focusing on the evaluation inference, the presenters discuss how scoring quality, consistency, and alignment with assessment criteria can help determine whether automated scores are meaningful and appropriate. The session highlights the role of validity evidence in building trust in AI-supported language assessment.

An Operational Automated Scoring System for Written Constructed Responses in Spanish.

2023 LTRC Conference Presentation.

Erik Voss, Dario Dathaguy, David Luongo. New York City, June 9, 2023.

Abstract:

This presentation highlights the development and implementation of an operational automated scoring system for Spanish written constructed responses. The presenters discuss how the system was designed to deliver reliable and consistent scoring at scale while supporting high-quality language assessment. The session also shares insights into the practical considerations and benefits of integrating AI-powered scoring into operational testing programs.

A Validity Argument to Support Automated Scoring Models for a Test of Written Spanish.

2022 LTRC Conference Presentation.

Erik Voss, Hung Phan & Xijia Wang. Language Testing Research Colloquium (LTRC), (Tokyo) online, March 7-8, 2022

Abstract:

This presentation explores the validity evidence needed to support the use of automated scoring models for Spanish writing assessments. The presenters examine how automated scores align with established scoring criteria and discuss the importance of reliability, fairness, and evaluation in building confidence in AI-assisted language assessment. The session provides a framework for ensuring that automated scoring systems produce meaningful and defensible assessment results.

Automated Scoring of Written Constructed-Responses in Spanish.

2021 ECOLT Conference (Online) Presentation.

Erik Voss, Hung Phan & Xijia Wang. October 22, 2021.

Abstract:

This presentation introduces the development of an automated scoring system for Spanish written constructed responses. The presenters discuss the use of machine learning techniques to evaluate student writing efficiently while maintaining consistency with human scoring standards. The session highlights the potential of AI to support scalable, reliable, and valid language assessment for proficiency testing.

2. AI Models & Emerging Technologies in Language Assessment

Exploring Technologies—including Large Language Models (LLMs) and Automatic Speech Recognition (ASR)—for language assessment.

Presentations

Adapting an Automatic Speech Recognition (ASR) Model for Spanish Language Assessment.

2026 AAAL Conference Presentation

Kim Sallee, Young-A Son & Meg Malone. Chicago, Illinois. March 22, 2026.

Abstract:

This presentation explores the adaptation of an Automatic Speech Recognition (ASR) model to support Spanish language assessment. The presenters discuss methods for improving speech recognition accuracy for learner language and examine how ASR technology can enhance the scoring of spoken responses. The session highlights both the opportunities and challenges of integrating AI-driven speech recognition into reliable and valid language assessment.

Automated Topic Detection in Spoken Language Assessment: Comparing Human-Labeled Training and Prompt-Based Category Prediction.

LTRC 2026 Poster Presentation.

Son, Y-A., Voss, E., Sallee, K., and Malone, M. Montreal, Canada. June 2-6, 2026.

Abstract:

This poster presentation examines automated topic detection in spoken language assessment by comparing two approaches: human-labeled training data and prompt-based category prediction. The presenters explore how AI can identify and classify topics in spoken responses, supporting more efficient analysis of learner language. The poster highlights the potential of automated topic detection to strengthen spoken assessment research while considering accuracy, consistency, and practical use in language testing.

Affordances and Challenges of Developing an Automated Spoken Spanish Assessment

2025 AAAL Conference Presentation

Sallee, K., Voss, E., Son, Y-A., Malone, M. Denver, Co. 2025, March 22-25.

Abstract:

This presentation explores the opportunities and challenges involved in developing an automated assessment for spoken Spanish. The presenters discuss how AI technologies can support the analysis and scoring of spoken responses while addressing issues such as accuracy, reliability, learner language variation, and validity. The session highlights key considerations for using automated tools responsibly in spoken language assessment.

Artificial Intelligence, Automated Essay Scoring, Large Language Models (EALTA 2024)

2024 EALTA Conference Presentation

Voss, E., Malone, M., Sallee, K. Belfast, Northern Ireland. June 4-9, 2024

Abstract:

This presentation explores the evolving role of artificial intelligence, automated essay scoring, and large language models (LLMs) in language assessment. The presenters examine how these technologies can enhance the scoring of written responses while discussing their potential benefits, limitations, and implications for validity and fairness. The session provides an overview of emerging AI applications and their impact on the future of language testing.

3. Ethics, Fairness & Responsible AI

Ensuring AI systems are ethical, transparent, secure, fair, and align with international testing standards.

Presentations

Applying the ILTA Code of Ethics to Responsible Language Testing.

2026 AAAL Presentation

Malone, M, Deygers, B. Chicago, IL 2026

Abstract:

This presentation examines how the International Language Testing Association (ILTA) Code of Ethics can guide the responsible development and use of language assessments in an era of artificial intelligence. The presenters discuss key ethical principles—including fairness, transparency, accountability, and test validity—and their application to AI-supported language testing. The session provides practical guidance for ensuring that innovation in assessment remains aligned with professional and ethical standards.

Accountability and Ethics in Language Assessment: Examining Approaches to Machine Scoring
2025 ECOLT Conference Presentation

Malone, M., Sallee, K., Gravina, S., Son, Y-A. Washington, DC. September 2026

Abstract:

This presentation explores the ethical and accountability considerations involved in using machine scoring for language assessment. The presenters examine approaches to ensuring fairness, transparency, validity, and responsible AI implementation while maintaining the quality and integrity of assessment outcomes. The session highlights best practices for balancing technological innovation with ethical decision-making in language testing.

Maintaining Ethical and Secure Data Practices in Machine Scoring of the AAPPL Test

2025 AIRiAL Conference Presentation

Gravina, S., Sallee, K., Sun, S. September 26-27, 2025.

Abstract:

This presentation examines best practices for maintaining ethical and secure data management in AI-powered scoring of the AAPPL assessment. The presenters discuss strategies for protecting learner data, ensuring privacy and security, and promoting transparency and fairness throughout the machine scoring process. The session highlights the importance of responsible data practices in supporting trustworthy and reliable language assessment.

4. AI Research & Machine Learning Methods

Research comparing different AI and machine learning approaches used in language assessment.

Related Presentations

Automated Scoring System for a Spanish Writing Test: Feature-Engineered Machine Learning vs. Large Language Model Approach.

2024 AAAL Paper Presentation.

Erik Voss, Young-A Son. Houston, TX. March 16, 2024.

Abstract:

This paper presentation compares two AI approaches for scoring Spanish writing assessments: feature-engineered machine learning and large language models (LLMs). The presenters evaluate the strengths and limitations of each method, focusing on scoring accuracy, consistency, and practical implementation. The session offers insights into how emerging AI technologies can support reliable, valid, and scalable language assessment.

Artificial Intelligence, Automated Essay Scoring, Large Language Models (EALTA 2024)

2024 EALTA Conference Presentation

Erik Voss, Meg Malone, and Kim Sallee. Belfast, Northern Ireland. June 4-9, 2024

Abstract:

This presentation explores how artificial intelligence, automated essay scoring, and large language models (LLMs) are transforming language assessment. The presenters examine emerging AI technologies, their potential to improve the scoring of written responses, and key considerations related to validity, reliability, and responsible implementation. The session offers insights into the evolving role of AI in supporting high-quality language assessment.

5. Foundational Research & Background

Resources for readers who want to understand the broader context of AI and technology in language assessment.

Publications

Voss, E. (2026). *An Automated Scoring System for the AAPPL Spanish Presentational Writing Tasks.* *Foreign Language Annals*, 59(1), 63–81.

Voss, E. (2025). *Comparison of Traditional Machine Learning and Neural Network Approaches for Automated Scoring of Second Language English Essays.* *Language Testing*, 42(4), 369–396.

Voss, E. (2024). *Artificial Intelligence in Language Assessment.* In A. J. Kunnan (Ed.), *The Concise Companion to Language Assessment* (pp. 112–125). Wiley-Blackwell.

Voss, E., Cushing, S. T., Ockey, G. J., & Yan, X. (2023). *The Use of Assistive Technologies Including Generative AI by Test Takers in Language Assessment: A Debate of Theory and Practice.* *Language Assessment Quarterly*, 20(4–5), 520–532.

Chapelle, C. A., & Voss, E. (2008). *Utilizing Technology in Language Assessment.* In N. H. Hornberger (Ed.), *Encyclopedia of Language and Education* (2nd ed., Vol. 7, *Language Testing and Assessment*, pp. 123–134). Springer.