

Proceedings of PEAFF 2026: The First International Workshop on Pedagogical Evaluation of Automated Feedback

Marcus Messer^{1,*}, Peter B. Johnson¹, Alexandra Neagu¹, Camille Kandiko Howson¹, Jaromir Savelka², Prarthana Bhattacharyya³ and Simon Woodhead³

¹Imperial College London, London, UK

²Carnegie Mellon University, Pittsburgh, USA

³Eedi, London, UK

1. Preface

The 1st Workshop on Pedagogical Evaluation of Automated Feedback (PEAF 2026) took place on the 28th of June 2026. The workshop was co-located at the Festival of Learning, which consisted of three main co-located conferences: ACM Learning@Scale, AI in Education, and Education Data Mining, and was hosted as an in-person-only event.

PEAF 2026 is centred on how we, as researchers, developers, and educators, evaluate the quality of automated feedback. Automated feedback is a commonly used method to provide students across education with information about their goals, their progress towards those goals, and how they should progress, typically using complex software systems and, more recently, generative AI. While automated feedback can be beneficial, evaluating these tools is often limited to accuracy and small-scale surveys. PEAFF aimed to extend these evaluation approaches by focusing on pedagogical evaluation: how does the feedback help the student learn, not just whether it is correct.

This year's inaugural workshop brought together participants to present results, approaches and techniques related, but not limited to, our topics:

- Establish shared practical applications of *education theory* to evaluate the pedagogical quality of feedback, beyond accuracy.
- Methods for *evaluating* the pedagogical quality of feedback, including computable metrics and qualitative methods.
- How *personalising* feedback to the learner, and to their preferences on how to interact with automated feedback, can affect the pedagogical quality of the feedback.
- How *culture and disciplines* factor into the pedagogical quality of automated feedback, especially accounting for biases in generative AI tools.
- *Ethical and societal considerations* of automating feedback in education, including, but not limited to, fairness, bias and equity across learner populations, socialisation, sense of belonging, curiosity and critical thinking.
- How automated feedback *impacts* teaching and learning. Examples include, but are not limited to, classroom activities, learning objectives, attitudes to learning, and societal and cultural changes.

PEAF 2026: The First International Workshop on Pedagogical Evaluation of Automated Feedback, June 28 2026, Seoul, South Korea

*Corresponding author.

✉ m.messer@imperial.ac.uk (M. Messer); peter.johnson@imperial.ac.uk (P. B. Johnson); alexandra.neagu20@imperial.ac.uk (A. Neagu); c.howson@imperial.ac.uk (C. K. Howson); jsavelka@andrew.cmu.edu (J. Savelka); prarthana.bhattacharyya@eedi.com (P. Bhattacharyya); simon.woodhead@eedi.com (S. Woodhead)

ORCID 0000-0001-5915-9153 (M. Messer); 0000-0001-7841-691X (P. B. Johnson); 0009-0004-0840-3776 (A. Neagu); 0000-0003-1045-3655 (C. K. Howson); 0000-0002-3674-5456 (J. Savelka); 0009-0009-0191-7148 (P. Bhattacharyya); 0000-0002-2192-9797 (S. Woodhead)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

The deadline for submissions was the 21st May 2026. After this deadline, two program committee members reviewed and rated the submissions from must reject to must accept. The chair then read the reviews and the papers and ultimately decided whether the submission was accepted or rejected for the workshop. To facilitate clear decisions for the authors, especially for borderline papers, the chair provided meta-reviews for each submission.

This year, PEAf received a total of 14 submissions 9 papers were accepted for publication. The program included interactive sessions, lightning talks, and discussions on the problems with evaluating automated feedback and how to begin addressing and researching them.

The organising committee would like to sincerely thank the members of the program committee for their thoughtful reviews, the authors for their interesting and novel work on the area of pedagogical evaluation of automated feedback, and the participants for ensuring that the workshop facilitated in-depth and meaningful discussion.

2. Organisation

2.1. Organising Committee

Marcus Messer	Imperial College London, London, UK
Peter B. Johnson	Imperial College London, London, UK
Alexandra Neagu	Imperial College London, London, UK
Camille Kandiko Howson	Imperial College London, London, UK
Jaromir Savelka	Carnegie Mellon University, Pittsburgh, USA
Prarthana Bhattacharyya	Eedi, London, UK
Simon Woodhead	Eedi, London, UK

2.2. Program Committee

Kamil Akhuseyinoglu	Carnegie Mellon University, Pittsburgh, US
Prarthana Bhattacharyya	Eedi, London, UK
Christopher Bogart	Carnegie Mellon University, Pittsburgh, US
Neil C C Brown	King's College London, London, UK
Andrew Garrard	University of Sheffield, Sheffield, UK
Peter B. Johnson	Imperial College London, London, UK
Camille Kandiko Howson	Imperial College London, London, UK
Huiyong Li	Kyushu University, Japan
Fun Siong Lim	Nanyang Technological University, Singapore
Alexandra Neagu	Imperial College London, London, UK
Simon Woodhead	Eedi, London, UK

3. Table of Contents

<i>An Introduction to Pedagogical Evaluation of Automated Feedback (PEAF)</i>	4–13
Johnson, Messer, Neagu, Kandiko Howson, Savelka, Bhattacharyya, Woodhead	

PEAF 2026 Workshop Papers	
<i>A Combined Strategy for the Pedagogical Evaluation of Automated Feedback: Generation, Decision and Fairness</i>	14–18
Kirouchenassamy, Yessad, Jolivet, Lallé, Branthôme, Luengo	
<i>Beyond Surface Human-Likeness: AI-Mentor Feedback Alignment, Pedagogical Adaptation, and Student Engagement in Longitudinal Learning Data</i>	19–24
Fan, Kim, Abedini, Koch-Truhponen, Koutcheme, Leinonen	
<i>Evaluating and Interpreting Gender Bias in LLM Feedback: Span-Level Embedding-Based Evidence from Automated Essay Feedback</i>	25–31
Du, Perez Ortiz, Cukurova	
<i>Evaluating the Pedagogical Quality of LLM-Generated Feedback: A Criterion-Based and Comparative Study</i>	32–37
Ngoe Kolle, Demmans Epp, Liaqat, Cutumisu	
<i>Revision-Loop Behavior and Learning Outcomes under Voluntary AI Formative Feedback in an Undergraduate Statistics Course</i>	38–43
Han	
<i>Supporting Tutors in the Gig Economy with Automated Feedback: A Case Study on Ringle</i>	44–49
Park, Kim, Overbay, Kim, Wee, Jung, Kim	
<i>The Correct Answer Trap: Pedagogically-Grounded Detection and Feedback for Hidden Misconceptions</i>	50–55
Imran, Bulathwela	
<i>Using a Learning Progression Framework to Guide LLM-Based Formative Assessment in STEM Education</i>	56–60
Wang, Li, Wieman, Kaldaras	
<i>Does Caring Cost Precision? Evaluating Anxiety-Framed LLM Misconception Feedback for Undergraduate Mathematics</i>	61–65
La Hadi, Alibasa, Asyhari	
