Assessing Assessment in the Age of AI: A Collaborative Framework for Computer Science

Claire Rocks
The University of Warwick
Coventry, UK

Marie Devlin
Steve Riddle
Newcastle University
Newcastle upon Tyne, UK

Lee Clift
University of Strathclyde
Glasgow, UK

James Davenport
Alan Hayes
University of Bath
Bath, UK

Wendy Yanez
University of Birmingham
Birmingham, UK

Patrick McAllister
Ian McChesney
Ulster University
Belfast, UK

Alastair Irons
Walbrook Institute London
London, UK

Alexandros Chrysikos
London Metropolitan University
London, UK

Abstract

Generative AI presents significant challenges for Computer Science assessment, raising questions about validity, authenticity, and inclusivity across diverse assessment formats. This RIPPA brings together computing educators to co-create and apply a practical framework for evaluating assessment in relation to generative AI, generating a shared evidence base and practical guidance to support informed and inclusive assessment design in AI-rich education.

CCS Concepts

• **Social and professional topics → Student assessment**.

Keywords

Assessment, AI, inclusivity, academic integrity

ACM Reference Format:
Claire Rocks, Marie Devlin, Steve Riddle, Lee Clift, James Davenport, Alan Hayes, Wendy Yanez, Patrick McAllister, Ian McChesney, Alastair Irons, and Alexandros Chrysikos. 2026. Assessing Assessment in the Age of AI: A Collaborative Framework for Computer Science. In *Proceedings of The United Kingdom and Ireland Computing Education Research (UKICER) conference (UKICER '26).* ACM, New York, NY, USA, 2 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

The rapid development of generative AI tools has introduced challenges for assessment in Computer Science in Higher Education, extending longstanding concerns around academic integrity and the limitations of detection-based approaches [3, 6, 8]. More recently, computing education research has examined the implications of generative AI for programming education, highlighting both its educational potential and the risks it poses for assessment validity [1, 4]. UK sector bodies have similarly called for assessment to adapt while maintaining standards and supporting learning [2, 5, 7].

Existing approaches largely provide general principles for assessment reform or frameworks for specifying permitted levels of AI use. There is still a gap in relation to a Computer Science-specific framework that can be used to evaluate how generative AI affects the breadth of assessment used across the discipline, from programming coursework to written and oral examinations, presentations and practical tasks. In the absence of such a framework, reactive responses, including increased reliance on invigilated examinations, risk narrowing assessment and undermining inclusive practice.

This RIPPA addresses the gap through a collaborative, multi-institutional process to co-create and refine a framework for evaluating assessment in response to generative AI. It considers contexts in which AI may be restricted, regulated, or incorporated, and will produce a shared evidence base and practical tools to support assessment evaluation and redesign.

2 Objectives

The project has four objectives:

(1) To collaboratively develop a practical framework for evaluating how generative AI affects the validity, authenticity, and integrity of Computer Science assessments.
(2) To apply and iteratively refine this framework across multiple institutions and a diverse range of assessment formats.
(3) To identify common patterns of susceptibility and robustness across Computer Science assessment types.
(4) To develop and disseminate practical strategies for valid, inclusive and future-facing assessment.

3 Research Questions

The project will be guided by the following research questions:

RQ1. How can the impact of generative AI on the validity, authenticity, and integrity of Computer Science assessments be conceptualised and systematically evaluated?
RQ2. How do different Computer Science assessment formats vary in relation to AI-assisted completion?
RQ3. What design features support valid, authentic, and inclusive assessment in AI-rich contexts?
RQ4. How useful is a collaboratively developed framework in supporting educators to evaluate and redesign assessments in response to generative AI?

UKICER '26, Cambridge, UK

ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM
https://doi.org/XXXXXXX.XXXXXXX

4 Methodology and Activities

The project will use a participatory, iterative co-design approach organised in three phases. Participants will help to develop the framework, apply it within their own contexts, and contribute to shared cross-institutional analysis.

Phase 1 – Month 1: Participants will be recruited through an initial workshop at UKICER 2026. Using sample Computer Science assessments, they will identify AI-related challenges and examples of good practice to develop an initial framework. Version 1 will take the form of a structured rubric for evaluating assessments against defined dimensions and recording findings consistently. The outputs will be a taxonomy of Computer Science assessment types and the initial assessment framework.

Phase 2 – Months 2–3: Participants will apply Version 1 within their own institutions by analysing one or two assessments. Using shared templates, they will record assessment characteristics and framework ratings, and reflect on their findings. Where useful, they may also conduct structured generative AI trials and retain examples of relevant interactions or outputs.

Participants will share their findings as short case studies at an online workshop. The project team will combine the ratings, reflections and, where available, AI trial outputs. This analysis will identify common AI-related challenges, compare assessment contexts, and inform Version 2 of the framework.

Phase 3 – Month 4: Following the Phase 2 workshop, participants will work in collaborative groups to apply Version 2 by redesigning at least one assessment or critically evaluating a proposed redesign.. They will consider implications for learning outcomes, validity, authenticity, feasibility, and inclusivity, documenting their decisions using a shared template. A final online workshop will facilitate discussion across groups. The outputs will include a refined framework, a set of practical assessment design patterns, and a collection of case studies demonstrating how Computer Science assessments can be adapted in response to generative AI.

Evaluation: Evaluation will be embedded throughout the project rather than conducted separately at the end. During co-design, participants will identify the characteristics of a useful framework for their assessment contexts. These discussions will shape evaluation measures which may include clarity, usability, relevance across assessment formats, support for inclusive redesign, and practical value.

After applying Version 2, participants will complete a short questionnaire combining Likert-scale and open-ended items. This will assess these measures alongside willingness to reuse or recommend the framework. Qualitative feedback will explore how the framework supported redesign and identify strengths, limitations and priorities for refinement. Findings will be analysed across participants and assessment contexts to assess its practical usefulness and inform the final iteration.

5 Project Team and Participant Engagement

The RIPPA will be led by a team of computing education researchers and practitioners from eight UK HEIs, with expertise in assessment, pedagogy, generative AI, and curriculum design. The Project Team will coordinate workshops, support participants in applying and refining the framework within their own contexts, and lead the synthesis, evaluation and dissemination of findings.

A wider group of computing educators will be recruited at UKICER 2026 and involved throughout the project. By applying and refining the framework through their own assessment practice, participants will contribute to its refinement across diverse institutional contexts and assessment formats, while producing outputs that are directly applicable to their own teaching practice. Through participation, individuals will become members of a Community of Practice dedicated to designing and refining Computer Science assessment in response to generative AI. Through workshops, peer review, and cross-institutional sharing of assessment redesigns, participants will apply a structured evaluation framework, receive constructive feedback on their own assessment practices, and learn from colleagues across diverse institutional contexts.

6 Outputs and Dissemination

The project will produce a co-created, practice-informed framework for evaluating the impact of generative AI on Computer Science assessment, alongside a set of practical assessment design principles and a collection of cross-institutional case studies illustrating their application in different contexts. Together, these outputs will provide both a shared evidence base and practical guidance to support educators in reviewing and adapting assessment in response to AI. The project will lead to a co-authored publication submitted to UKICER 2027. The framework, case studies, and supporting materials will also be shared openly with the wider computing education community to support ongoing discussion and development of assessment practice in Computer Science education.

References

[1] Brett A. Becker, Paul Denny, James Finnie-Ansley, Andrew Luxton-Reilly, James Prather, and Eddie Antonio Santos. 2023. Programming Is Hard - or at Least It Used to Be: Educational Opportunities and Challenges of AI Code Generation. *SIGCSE 2023 - Proceedings of the 54th ACM Technical Symposium on Computer Science Education* 1 (3 2023), 500–506. doi:10.1145/3545945.3569759

[2] Russell Group. 2024. Russell Group principles on the use of generative AI tools in education. https://www.russellgroup.ac.uk/news/russell-group-hosts-dutch-universities-discuss-ai-education

[3] Sathiamoorthy Manoharan and Ulrich Speidel. 2020. Contract cheating in computer science: A case study. *Proceedings of 2020 IEEE International Conference on Teaching, Assessment, and Learning for Engineering, TALE 2020* (12 2020), 91–98. doi:10.1109/TALE48869.2020.9368454

[4] Carrie Anne Philbin and Sue Sentance. 2025. Student Use and Teacher Practice: A Scoping Review of Generative AI in Computing Education Using PICRAT. *UK and Ireland Computing Education Research Conference (UKICER 2025), September 2025, Edinburgh, United Kingdom* 1 (9 2025), 1–7. doi:10.1145/3754508.3754513

[5] QAA. 2023. Reconsidering assessment for the ChatGPT era: QAA advice on developing sustainable assessment strategies. https://www.qaa.ac.uk/docs/qaa/members/reconsidering-assessment-for-the-chat-gpt-era.pdf

[6] Judy Sheard, Simon, Matthew Butler, Katrina Falkner, Michael Morgan, and Amali Weerasinghe. 2017. Strategies for maintaining academic integrity in first-year computing courses. *Annual Conf. Innovation and Technology in Computer Science Education, ITiCSE* Part F128680 (6 2017), 244–249. doi:10.1145/3059009.3059064

[7] Rose Stephenson and Charlotte Armstrong. 2026. *Student Generative Artificial Intelligence Survey 2026.* Technical Report 199. Higher Education Policy Institute (HEPI). https://www.hepi.ac.uk/reports/student-generative-ai-survey-2026/

[8] Debora Weber-Wulff, Alla Anohina-Naumeca, Sonja Bjelobaba, Tomáš Foltýnek, Jean Guerrero-Dib, Olumide Popoola, Petr Šigut, and Lorna Waddington. 2023. Testing of Detection Tools for AI-Generated Text. *International Journal for Educational Integrity* 19 (7 2023). Issue 1. doi:10.1007/s40979-023-00146-z