Journal of Learning Development in Higher Education ISSN: 1759-667X
Issue 34 February 2025

Contextual assessment design in the age of generative AI

Chahna Gonsalves
King's College London, UK

Abstract

Generative AI (GenAI) is transforming higher education. It has already challenged the validity of traditional assessment methods and revealed concerns about the authenticity and reliability of conventional approaches. This opinion piece proposes an expanded theoretical framework for contextual learning, incorporating practical, situational, experiential, interdisciplinary, and ethical dimensions to address the limitations of authentic assessment in the face of GenAI's rapid advancements. Whilst the framework is primarily theoretical and not yet empirically validated, it offers a new way of thinking about assessment design, providing educators with a vocabulary and conceptual tools to create more resilient and effective assessments. Learning Developers, who often lead initiatives, engage in collaborative efforts, and support both educators and students, are particularly well-positioned to implement and advocate for these strategies. By integrating diverse contexts into assessment design, this framework aims to promote higher-order thinking and real-world applications, making assessments more adaptable to the complexities of modern professional environments. Future research should explore the empirical validation of this framework across disciplines to ensure its broader applicability and impact on student engagement, learning outcomes, and academic integrity.

Keywords: contextual learning; assessment design; artificial intelligence; generative AI.

Introduction

Generative AI (GenAI) is reshaping higher education through its challenge to the reliability of traditional assessment methods. Its ability to generate text, audio, and visual content casts doubt on the authenticity of conventional essays, reports, case studies, reflective writing, creative assignments, and even some forms of problem-based and project-based

assessments (Cassidy, 2023; Bobula, 2024). With GenAI tools such as ChatGPT, Midjourney, and DALL-E increasingly being integrated into academia, universities must urgently revise their assessment strategies to remain credible.

The need to evolve assessment practices in response to GenAI is urgent (Bobula, 2024). GenAI tools use sophisticated natural language processing and deep learning models to produce human-like responses, analyse data, and generate creative content (McAlister, Alhabash and Yang, 2023). GenAI chatbots, for instance, predict the next word in a sequence based on vast data patterns, creating coherent and contextually relevant responses. Although these capabilities enhance interactive learning, they also pose significant challenges to traditional assessment methods. As AI tools become increasingly prevalent,[1] educational institutions must adapt their strategies to maintain academic integrity and evaluation validity (Bobula, 2024). In this context, 'validity' refers to the extent to which assessment methods measure the intended learning outcomes. Valid assessment ensures that students' knowledge and skills — not just their ability to use AI tools — are accurately evaluated (Messick, 1989). Traditional assessments risk losing their validity if they fail to distinguish between AI-generated and student-generated work. To maintain integrity, universities must adapt their assessment strategies to ensure that they continue to measure genuine student learning and critical engagement.

Universities have explored various strategies to mitigate against AI, with many reverting to in-person, written examinations to bypass AI misuse (Cassidy, 2023); seeking to develop AI-resilient formats that emphasise creativity and critical thinking; and developing ethical guidelines for incorporating AI tools into traditional assessment. Each approach poses challenges and opportunities, highlighting the need for a nuanced strategy to integrate AI effectively into educational practices. However, many of these efforts remain piecemeal and reactive rather than strategic and forward-thinking, risking a cycle of short-term fixes rather than sustainable transformation. Despite widespread concern and numerous pilot initiatives, uncertainty persists about the best path forward (Eager and Brunton, 2023). Nonetheless, inaction is not an option: the evolution of pedagogy and assessment design is necessary to adapt to GenAI realities.

[1] The rapid advancement of GenAI is exemplified by the release of DeepSeek's R1 model in January 2025. Developed in just two months at a cost of under $6 million, R1 demonstrates capabilities comparable to leading models from companies like OpenAI, which reportedly spends over $5 billion annually on similar technologies (Financial Times, 2025). This significant development occurred between the initial submission of this article in June 2024 and its publication in February 2025.

The case for contextual learning: bridging research and innovation

The imperative to redesign assessment strategies is underscored by extensive research on the implications of technological advancements. Recent studies, such as by McAlister, Alhabash, and Yang (2023) and Luo (2024), highlight the critical need for evolving assessment practices to meet modern educational demands. Much of the existing literature, however, does not offer practical strategies, leaving a significant gap that this opinion piece aims to contribute towards resolving. Xia et al. (2024) advocate for more innovative assessment methods, such as live presentations and creative projects, which are less susceptible to AI-generated content. However, their recommendations often remain broad and lack specific guidance on practical implementation. Moreover, frameworks like the AI Assessment Scale (AIAS) (Perkins et al., 2024) provide structured approaches for ethical AI use but stop short of offering practical steps for redesigning assessments in response to AI's capabilities. This gap between high-level recommendations and actionable strategies underscores the need for a more nuanced and context-driven approach to assessment design. Research shows that whilst GenAI excels at routine cognitive tasks, it falters in higher-order, complex analysis, highlighting the imperative to weave context into assessment design (Thanh et al., 2023). This limitation of AI underscores the necessity for integrating context into assessments — moving beyond traditional notions of authenticity and realism. Thanh et al. (2023) argue that contextualisation is crucial as it connects theoretical concepts to real-world applications, enriching the educational experience by deepening understanding and enhancing critical thinking.

Given these insights, I propose 'context' as a pivotal factor in redesigning assessments. This approach aligns with the need for assessment designs that incorporate real-world applications and emphasise the contextualisation of tasks, ensuring that AI-supported assessments are meaningful, relevant, and better able to engage students in higher-order thinking and practical problem-solving (Eager and Brunton, 2023; Mills, Bali and Eaton, 2023). To develop this argument further, I draw on Gilbert's models of context as a foundational framework, expanding them to address the specific challenges posed by GenAI and demonstrating how a more nuanced application of context can enhance assessment resilience.

Expanding on Gilbert's models of context

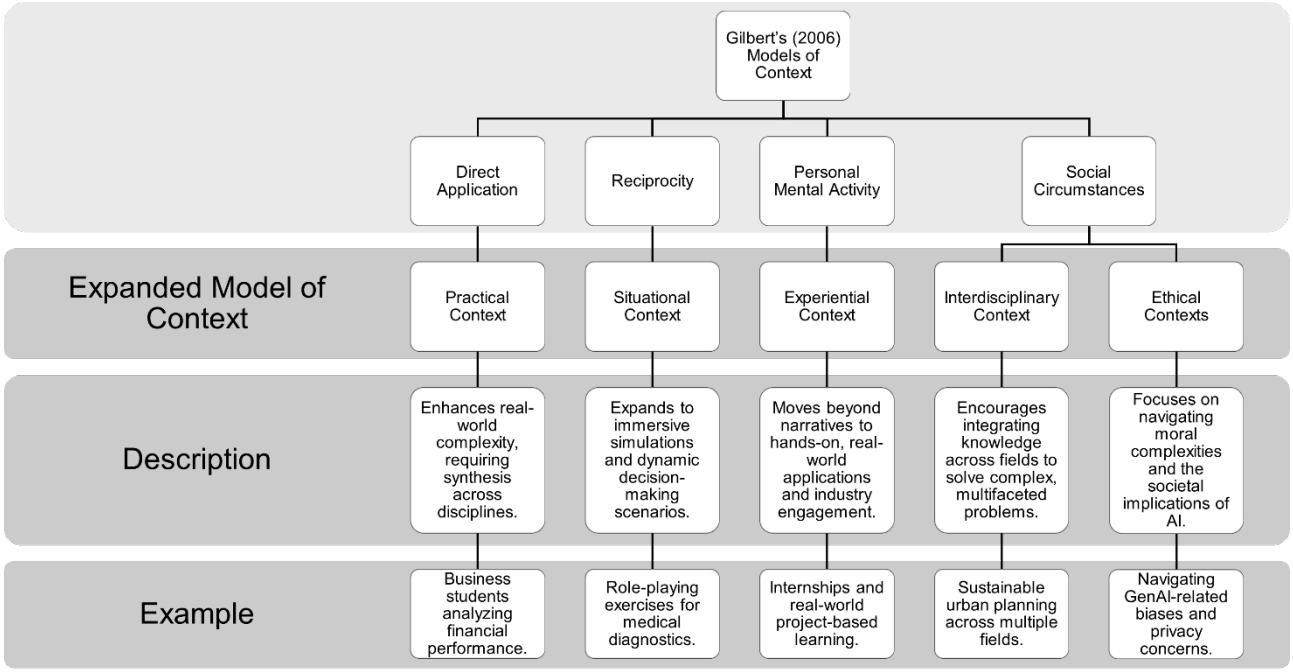
Gilbert's (2006) models of context offer a foundational approach to integrating real-world relevance into educational assessments. His models were initially developed for Chemical Education, but they are applicable to other disciplines. They highlight how context enriches the understanding of complex concepts, making theoretical knowledge more practical and engaging.

1. **Context as the direct application of concepts.** This model emphasises the practical utility of theoretical knowledge, such as using an understanding of acidity to treat a wasp sting with household solutions. Although it emphasises links between theory and practice, this model may lack the depth needed to encourage critical thinking about broader implications, particularly in complex real-world scenarios.
2. **Context as reciprocity between concepts and applications.** Gilbert's second model focuses on the bidirectional relationship between theory and practice, where each informs and enhances the other. For instance, 'pure water' may refer to chemically pure water in a chemistry context and might denote drinkable water in an environmental context. This dual perspective enriches understanding, yet it may not fully prepare students to navigate interdisciplinary and ethical complexities that arise in today's interconnected world.
3. **Context provided by personal mental activity.** This model encourages deep personal engagement by having students create personal narratives that connect their learning to broader themes, such as interpreting ecological issues through environmental science. Yet, without broader contextual integration, it may fall short in preparing students for the interdisciplinary and ethical challenges they will encounter in their professional lives.
4. **Context as social circumstances.** This model situates learning within broader social, cultural, and ethical frameworks. For example, understanding the chemistry of global warming requires not only scientific knowledge but also an appreciation of societal and policy implications. Although this model promotes holistic understanding, it may not sufficiently address the practical or interdisciplinary engagement necessary to tackle today's multifaceted challenges.

Gilbert's models provide a strong foundation for understanding context in learning, but they do not fully address the complexities introduced by GenAI. To bridge this gap, I propose

an expanded framework with five key dimensions — practical, situational, experiential, interdisciplinary, and ethical — drawn from my experience in contextual assessment design. These speculative yet practical ideas aim to spark discussion on adapting assessment practices in the GenAI era. My framework offers educators a structured yet flexible approach to embedding context, ensuring assessments remain resilient against AI-generated outputs whilst fostering critical thinking and real-world application. Rather than claiming to be definitive, it serves as a conceptual toolset for those exploring ways to maintain assessment integrity in a rapidly evolving landscape. Whilst speculative, its strength lies in its cross-disciplinary relevance and potential for further empirical exploration. By introducing these dimensions, I hope to encourage dialogue on preserving meaningful, valid, and AI-resistant assessments that continue to promote deep learning. Figure 1 provides a visual representation of how these five dimensions expand upon Gilbert's models, illustrating their integration into a more comprehensive framework for contextual assessment.

Figure 1. Expanding Gilbert's (2006) Models of Context: a framework for contextual assessment.



1. **Practical context.** Building on Gilbert's direct application model, this dimension emphasises complex, real-world scenarios that require students to synthesise and apply knowledge within a discipline. This is particularly relevant in fields where GenAI tools can perform basic tasks but struggle to offer context-dependent decision-making in complex, novel scenarios (Frenkel and Emara, 2024). Business students, for instance, might analyse a company's financial performance by integrating

accounting principles, market trends, and financial ratios. This approach encourages students to evidence deeper learning and critical thinking by challenging them with the complexities of their specific disciplines.

2. **Situational context.** Expanding Gilbert's reciprocity model, situational context involves dynamic and immersive simulations and role-playing exercises that mirror real-life decision-making and negotiation. These immersive approaches are crucial for developing adaptability and critical thinking — skills that are difficult for AI to replicate. Studies show that whilst AI can assist in technical problem-solving, it often fails when asked to apply knowledge to situationally dependent scenarios (Frenkel and Emara, 2024). This approach promotes the development of adaptability and resilience whilst requiring students to identify and respond to situational cues, such as recognising symptoms and contextual clues when diagnosing a patient. It enhances their perceptiveness and responsiveness in dynamic, emergent contexts.
3. **Experiential context.** Extending Gilbert's focus on personal engagement, experiential context emphasises the importance of hands-on learning and real-world engagement, bridging academic theory with lived experiences through internships, project-based learning, and industry collaboration. Unlike situational context, which immerses students in controlled simulations, experiential learning requires students to interact directly with complex, unpredictable, real-world environments. This pushes them to apply theoretical knowledge, exercise judgment, and adapt to context-specific variables. These are all skills that GenAI tools struggle to replicate. Recent studies show that whilst AI can streamline routine tasks, it often shifts human roles from creation to oversight, reducing situational awareness and impairing human judgment in complex, real-world contexts (Simkute et al., 2024). Students must identify and integrate relevant contextual factors — such as industry standards, cultural dynamics, and stakeholder expectations — into their work. By engaging in experiential learning, students confront the inherent uncertainties of real-world scenarios, where critical thinking, adaptability, and professional judgment are essential. This process deepens their understanding, fosters adaptability, and prepares them for the complex demands of diverse professional environments.
4. **Interdisciplinary context.** Moving beyond the single-disciplinary focus of Gilbert's models, this dimension encourages the integration of knowledge across disciplines to solve multifaceted problems. Designing sustainable urban infrastructure, for example, might involve input from Civil Engineering, Environmental Science, public policy, and Sociology. This dimension fosters comprehensive problem-solving skills that are

crucial for addressing the interconnectedness of global issues, thereby helping students navigate complex, cross-disciplinary challenges. This also aligns with educational theories of situated learning and knowledge transfer, where learning is most effective when students can apply knowledge across contexts, promoting adaptability and innovation in complex scenarios (Lang, 2024). This expansion offers a deeper and more relevant application of Gilbert's models by supporting students to navigate and address the interconnectedness of global issues in ways that single-discipline approaches cannot.

5. **Ethical context.** Whilst Gilbert's model of social circumstances addresses social and cultural frameworks, incorporating the ethical context adds a critical layer focused on navigating moral complexities in both societal ethics and the specific context of GenAI use. The widespread adoption of AI, including generative models, requires students to develop ethical awareness, particularly regarding privacy, bias, and accountability, alongside the skills to respond effectively (Bobula, 2024). Incorporating discussions on AI's limitations and societal implications into the curriculum, along with structured support for understanding these issues, helps students develop ethically sound decision-making skills. This dimension equips students with the critical-thinking skills necessary to engage with ethical dilemmas in professional settings and consider AI's broader societal impact with nuance and responsibility.

My expanded framework builds on Gilbert's models by categorising and integrating the practical, situational, experiential, interdisciplinary, and ethical dimensions, offering a comprehensive approach to contextual learning. It directly confronts the challenges posed by AI, ensuring assessments stay relevant, engaging, and intricately aligned with the evolving complexities of modern education. Although GenAI tools can synthesise information from various contexts, the process fundamentally depends on the user's ability to identify, gather, and understand relevant contextual information before feeding it into AI tools (Thanh et al., 2023). Active engagement with the material reflects student awareness and their ability to integrate these contexts effectively, making contextual assessment a vital component of evolved assessment design. Ultimately, this framework ensures that assessment remains focused on evaluating the student's critical thinking and contextual understanding rather than their ability to generate AI-assisted responses, reinforcing the importance of human judgment and interpretation in learning.

Learning Developers play a vital role in advancing the integration of these contextual frameworks into assessment design. By leading initiatives, engaging in collaborative efforts with educators, and providing support to both faculty and students, they are instrumental in driving the adoption of innovative strategies (Eager and Brunton, 2023). Their involvement ensures that learning activities and assessments are crafted with a focus on personalised examples, interdisciplinary knowledge, and real-time contextualisation. This collaborative approach not only keeps educational content relevant and engaging in AI-supported environments, but it also empowers students to apply and manage contextual knowledge in ways that AI alone cannot achieve.

An additional layer of context is provided by involving university instructors and industry professionals, each contributing distinct yet complementary perspectives. Instructors bring pedagogical expertise and subject knowledge, whilst industry professionals offer real-time insights and practical applications, grounding learning in current industry practices. This collaboration enriches the classroom experience in ways AI cannot easily replicate.

Whilst Learning Developers may not design assessments themselves, they play a crucial role in supporting educators and ensuring students are equipped with the academic skills necessary to meet these evolving assessment demands. To do so effectively, they must first understand the contextual dimensions that shape assessments — how instructors integrate subject-specific knowledge and how industry professionals provide real-world relevance. With this understanding, Learning Developers can support educators in structuring assessments that effectively blend AI tools with students' class-derived knowledge and real-world applications whilst maintaining academic integrity and engagement.

At the same time, Learning Developers help students develop critical thinking, academic writing, and problem-solving skills tailored to these context-rich assessments. By embedding nuanced, context-specific elements into assessment strategies, they ensure that students' personal understanding remains central — making it significantly more difficult for GenAI to generate meaningful responses. This fosters a more authentic, intellectually rigorous educational experience, reinforcing the irreplaceable value of human-led learning and meaningful student engagement with both theoretical and applied knowledge.

Differentiating contextual learning from authentic assessment

Authentic assessment has been endorsed due to its ability to simulate professional scenarios, equipping students with the skills needed for the workforce by emphasising realism, contextualisation, and problem-solving (Wiggins, 1998; Villarroel et al., 2017; Stankov, 2024). Despite its popularity, however, the concept of authentic assessment lacks a clear definition, with interpretations ranging from promoting work readiness to fostering personal growth and addressing societal challenges (Arnold and Croxford, 2024). The emphasis on 'realism' has been central to its perceived value (Bosco and Ferns, 2014), often leading to the assumption that assessments closely mirroring real-world tasks are inherently more resistant to academic misconduct (Arnold and Croxford, 2024). Emerging research has challenged this assumption, questioning whether authentic tasks are truly more cheat-resistant (Nikolic et al., 2023; Arnold and Croxford, 2024), especially with the increasing sophistication of GenAI tools and the rapid evolution of the Large Language Models that underpin them.

Whilst realism remains valuable, traditional authentic assessments often confine this concept to specific professional contexts (Villarroel et al., 2017; Stankov, 2024) which can be limiting. The integrity of such assessments is increasingly undermined by GenAI's ability to generate outputs that meet conventional assessment criteria, such as scientific reports or business plans (Nikolic et al., 2023). As a result, there is a pressing need for a more flexible and expansive approach to assessment design: one that moves beyond professional realism alone and incorporates multiple layers of contextual complexity. By embedding diverse contextual dimensions into assessment tasks, educators can create more resilient assessments that challenge students in ways AI cannot easily replicate.

Examinations, for instance, can be adapted by integrating contextual elements that require students to apply disciplinary knowledge to real-world dilemmas or reflect on personal experiences. This challenges the notion of a strict dichotomy between traditional and authentic assessments, suggesting that the boundaries are more fluid than previously thought (Arnold and Croxford, 2024). With appropriate contextualisation, even structured exams can achieve a level of authenticity that aligns with both disciplinary traditions and evolving educational demands (Stankov, 2024). By embedding contextual complexity, educators can design assessments that remain resilient against AI-generated responses, ensuring student input rather than AI assistance remains central to the evaluation process.

Conclusion

In this opinion piece, I have argued that assessment design must undergo a fundamental transformation to remain fit for purpose in the era of GenAI. This commentary has advocated for a shift beyond the limited notion of authentic assessment toward a more nuanced framework of contextual learning. Authenticity, as traditionally conceived, often centres on professional realism, yet GenAI's capabilities have exposed the vulnerabilities of this approach. A broader perspective is needed — one that embeds practical, situational, experiential, interdisciplinary, and ethical dimensions into assessment design. Contextual learning enhances assessment integrity by introducing complexity, nuance, and personal interpretation — elements that AI struggles to replicate. This approach is not just a response to AI's growing presence but a way to foster deeper engagement, higher-order thinking, and meaningful real-world application.

For this framework to be effective, assessors must evaluate how well students engage with and apply contextual elements. This is particularly relevant in assessments that involve professional exposure, such as placements, work-integrated learning, and industry collaborations, where students must demonstrate their ability to reflect critically and apply disciplinary knowledge in dynamic environments. Establishing clear assessment criteria and rubrics will be essential in ensuring that contextual engagement is measurable and meaningful.

Scaling contextual learning requires more than incremental adjustments. Institutions must prioritise professional development, cross-disciplinary collaboration, and pedagogical innovation to integrate context-rich assessments across disciplines. Whilst the application of this framework will differ across fields — with STEM, social sciences, and the humanities presenting distinct challenges — the overarching goal remains the same: assessments must reflect the complexities of real-world problem-solving and require students to engage beyond AI-generated outputs.

Although this commentary is theoretical, its implications warrant empirical validation. Future research should examine how contextual learning influences student engagement, critical thinking, and academic integrity across disciplines. Empirical enquiry will refine this framework, ensuring that assessment strategies remain both rigorous and adaptable within an evolving educational landscape. As AI continues to shape education, assessment design must evolve beyond the narrow view that equates realism with integrity. Contextual learning offers a more robust alternative by embedding complexity, relevance, and critical

engagement into assessment tasks. This ensures that evaluation remains intellectually demanding, personally meaningful, and resistant to AI-generated shortcuts, reinforcing the irreplaceable value of human thought and judgment in education.

Acknowledgements

The author did not use generative AI technologies in the creation of this manuscript.

References

Arnold, L. and Croxford, J. (2024) 'Is it time to stop talking about authentic assessment?', *Teaching in Higher Education*, pp.1–9. Available at: https://doi.org/10.1080/13562517.2024.2369143

Bobula, M. (2024) 'Generative artificial intelligence (AI) in higher education: a comprehensive review of challenges, opportunities, and implications', *Journal of Learning Development in Higher Education,* 30. Available at: https://doi.org/10.47408/jldhe.vi30.1137

Bosco, A.M. and Ferns, S. (2014) 'Embedding of authentic assessment in work-integrated learning curriculum', *Asia-Pacific Journal of Cooperative Education,* 15(4)**,** pp.281–290. Available at: https://eric.ed.gov/?id=EJ1113553

Cassidy, C. (2023) 'Australian universities to return to "pen and paper" exams after students caught using AI to write essays', *Guardian*, 10 January. Available at: https://www.theguardian.com/australia-news/2023/jan/10/universities-to-return-to-pen-and-paper-exams-after-students-caught-using-ai-to-write-essays (Accessed: 19 August 2024).

Eager, B. and Brunton, R. (2023) 'Prompting higher education towards AI-augmented teaching and learning practice', *Journal of University Teaching and Learning Practice,* 20(5). Available at: https://doi.org/10.53761/1.20.5.02

Financial Times (2025) 'OpenAI's Sam Altman vows "better models" as China's DeepSeek disrupts global race', *Financial Times*, 28 January. Available at: https://www.ft.com/content/b98e4903-ac05-4462-8ad1-eda619b6a9c4 (Accessed: 2 February 2025).

Frenkel, M.E. and Emara, H. (2024) 'ChatGPT-3.5 and -4.0 and mechanical engineering: examining performance on the FE mechanical engineering and undergraduate exams', *Computer Applications in Engineering Education,* 32(6). Available at: https://doi.org/10.1002/cae.22781

Gilbert, J.K. (2006) 'On the nature of "context" in chemical education', *International Journal of Science Education,* 28(9), pp.957–976. Available at: https://doi.org/10.1080/09500690600702470

Lang, J.C. (2024) 'Embracing generative AI for authentic learning', *Creative Education,* 15(1), pp.1–20. Available at: https://doi.org/10.4236/ce.2024.151001

Luo, J. (2024) 'A critical review of GenAI policies in higher education assessment: a call to reconsider the "originality" of students' work', *Assessment and Evaluation in Higher Education***,** 49(5), pp.1–14. Available at: https://doi.org/10.1080/02602938.2024.2309963

Mcalister, A.R., Alhabash, S. and Yang, J. (2023) 'Artificial intelligence and ChatGPT: exploring current and potential future roles in marketing education', *Journal of Marketing Communications***,** 30(2), pp.1–22. Available at: https://doi.org/10.1080/13527266.2023.2289034

Messick, S. (1989) 'Validity', in R.L. Linn (ed.) *Educational measurement*. 3rd edn. New York: Macmillan/American Council on Education, pp.13–103.

Mills, A., Bali, M. and Eaton, L. (2023) 'How do we respond to generative AI in education? Open educational practices give us a framework for an ongoing process', *Journal of Applied Learning and Teaching,* 6(1), pp.16–30. Available at: https://doi.org/10.37074/jalt.2023.6.1.34

Nikolic, S., Daniel, S., Haque, R., Belkina, M., Hassan, G.M., Grundy, S., Lyden, S., Neal, P. et al. (2023) 'ChatGPT versus engineering education assessment: a multidisciplinary and multi-institutional benchmarking and analysis of this generative artificial intelligence tool to investigate assessment integrity', *European Journal of Engineering Education,* 48(4), pp.559–614. Available at: https://doi.org/10.1080/03043797.2023.2213169

Perkins, M., Furze, L., Roe, J. and Macvaugh, J. (2024) 'The Artificial Intelligence Assessment Scale (AIAS): a framework for ethical integration of generative AI in educational assessment', *Journal of University Teaching and Learning Practice,* 21(6). Available at: https://doi.org/10.53761/q3azde36

Simkute, A., Tankelevitch, L., Kewenig, V., Scott, A.E., Sellen, A. and Rintel, S. (2024) 'Ironies of generative AI: understanding and mitigating productivity loss in human-AI interaction', *International Journal of Human–Computer Interaction*, pp.1–22. Available at: https://doi.org/10.1080/10447318.2024.2405782

Stankov, P. (2024) 'Are exams authentic assessment? The case of economics', *Practical Assessment, Research, and Evaluation,* 29(1), pp.1−20. Available at: https://doi.org/10.7275/pare.1984

Thanh, B.N., Vo, D.T.H., Nhat, M.N., Pham, T.T.T., Trung, H.T. and Xuan, S.H. (2023) 'Race with the machines: assessing the capability of generative AI in solving authentic assessments', *Australasian Journal of Educational Technology,* 39(5), pp.59–81. Available at: https://doi.org/10.14742/ajet.8902

Villarroel, V., Bloxham, S., Bruna, D., Bruna, C. and Herrera-Seda, C. (2017) 'Authentic assessment: creating a blueprint for course design', *Assessment and Evaluation in Higher Education,* 43(5), pp.840–854. Available at: https://doi.org/10.1080/02602938.2017.1412396

Wiggins, G. (1998) *Educative assessment: designing assessments to inform and improve student performance.* San Francisco, CA, Jossey-Bass.

Xia, Q., Weng, X., Ouyang, F., Lin, T.J. and Chiu, T.K. (2024) 'A scoping review on how generative artificial intelligence transforms assessment in higher education', *International Journal of Educational Technology in Higher Education,* 21(40). Available at: https://doi.org/10.1186/s41239-024-00468-z

Author details

Chahna Gonsalves is a Senior Lecturer in Marketing (Education) at King's College London, where her research and scholarship focus on assessment and feedback, generative AI, communication, and message impact. As Department Education Lead, she oversees curriculum development and pedagogical strategy within the Marketing Department. She is also a Senior Fellow of Advance HE (SFHEA).

Licence