Flawed by design: embedding AI to drive authentic assessment and integrity

Professor Rania Shibl, Southern Cross University, Australia, and Dr Judy Watson, University of the Sunshine Coast, Australia

Background

The rapid and widespread adoption of generative artificial intelligence (AI) tools in higher education has intensified long-standing debates about academic integrity, authenticity and the purpose of assessment. Tools such as ChatGPT and other large language models are now capable of producing syntactically correct code, coherent explanations and complete solutions to many assessment tasks traditionally used in computing and data-related disciplines. While these technologies offer significant opportunities to support learning, they also challenge educators to reconsider what knowledge, skills and attributes assessments are designed to measure.

In disciplines such as information and communications technology (ICT), the impact of generative AI is particularly acute. Tasks that previously served as reliable indicators of individual competence, such as writing SQL queries or designing database schemas, can now be completed, at least superficially, through automated means. This raises critical questions: if AI tools can generate code, what is the educational value of assessing code production alone? How can assessment practices evolve to ensure they continue to measure student understanding, judgement and accountability in an AI-enabled world?

This case study is situated within a second-year undergraduate course in database design (ICT211) at the University of the Sunshine Coast (UniSC), enrolling approximately 200 students. The course focuses on foundational concepts in relational database design, including schema development, data integrity constraints and SQL implementation. Historically, assessment in this course followed a conventional model, requiring students to design and implement database structures based on a provided case study. However, the increasing accessibility of generative AI tools significantly undermined the reliability of these tasks as indicators of genuine learning.

The central challenge addressed in this work was the design of a summative assessment that remained authentic, academically rigorous and scalable, while acknowledging rather than prohibiting the presence of AI tools. Rather than attempting to restrict or detect AI use, the assessment sought to reposition AI as part of the learning environment and to shift the focus of assessment from content generation to content evaluation and responsibility.

The work is informed by several complementary theoretical and conceptual frameworks that collectively support a shift toward authentic, applied and AI-aware assessment design. Authentic assessment theory argues that students should demonstrate learning through tasks that mirror real-world professional practice rather than through decontextualised or purely abstract exercises. Wiggins' (1990) seminal work on authentic assessment

emphasised that learners should engage in the same kinds of complex tasks faced by practitioners, thereby bridging the gap between education and practice. In higher education contexts, authentic assessment has been shown to enhance engagement, motivation and the development of transferable, 21st-century skills by requiring students to integrate knowledge and apply it in meaningful situations (Vlachopoulos and Makri, 2024). In the context of database design, this means evaluating not only whether students can write syntactically correct SQL, but whether they can diagnose errors, interpret requirements and implement solutions that would function in a real organisational setting.

Constructivist learning theory further underpins this approach. From a constructivist perspective, learners actively construct understanding through problem-solving, interpretation, dialogue and reflection rather than passively absorbing information (Tam, 2000). Foundational theorists such as Piaget and Vygotsky highlighted the active and social nature of learning, arguing that knowledge is built as students grapple with authentic problems and reflect on their experiences. In practical terms, constructivist pedagogy privileges inquiry-based and open-ended tasks that require learners to test ideas, receive feedback and refine their understanding (Chand, 2024). The use of deliberately flawed scripts in this case study aligns closely with this paradigm: students must interpret the problem, diagnose inconsistencies and iteratively refine their solutions, thereby constructing knowledge through engagement with complexity rather than reproducing memorised procedures.

Recent scholarship on AI in education further reframes assessment design in light of generative technologies. Rather than positioning AI solely as a threat to academic integrity, emerging research advocates treating AI as a cognitive partner within the learning process (Elshall and Badir, 2025). This perspective calls for assessments that make student thinking visible and foreground distinctly human capabilities such as critical judgement, ethical reasoning, metacognitive oversight and accountability. Since generative AI can now produce plausible code or written responses, sustainable assessment in the AI era must move beyond evaluating surface-level outputs and instead examine how students critique, validate, adapt and take responsibility for AI-assisted work. Sector-wide discussions and contemporary research emphasise that assessment redesign should leverage AI for what it does well while ensuring that human expertise remains central (Corbin, Dawson and Liu, 2025; Khlaif et al, 2025). This case study aligns with these calls by embedding AI within the assessment process in a constrained and purposeful manner, shifting the focus from content generation to evaluation, professional judgement and accountability in an AI-enabled world.

Approach

To address these challenges, a constrained AI-supported assessment model was developed and implemented. Central to this approach was the creation of ‘Liam Chat’, a simulated AI client designed to generate assessment inputs rather than assessment outputs. Liam Chat was built using OpenAI’s API and deployed through a ShinyApps.io

interface, providing a lightweight, web-based tool accessible to students during the assessment period.

The assessment task required students to review, correct and complete a flawed SQL CREATE TABLE script and to populate the database with appropriate sample data based on a fictional business case, *The Smart Café*. Rather than providing all students with the same script and dataset, Liam Chat generated unique, personalised inputs for each student. These inputs were deliberately flawed, containing errors in data types, missing or incorrect constraints, or logical inconsistencies that required diagnosis and correction.

A key design decision was to tightly constrain Liam Chat's capabilities. Although initially envisioned as a tool for interactive requirement gathering, its functionality in this iteration was deliberately limited to two outputs only: (1) a flawed SQL schema and (2) a dynamically generated dataset. Liam Chat did not provide explanations, hints, or corrective feedback. This constraint was essential to maintain alignment with learning outcomes and to prevent the AI from inadvertently completing the task on behalf of students.

Students were explicitly informed that the scripts generated by Liam Chat would be incorrect, but that the nature and location of the errors would vary. This uncertainty mirrored real-world professional contexts, where system specifications are often incomplete or inaccurate and where practitioners must exercise judgement rather than follow instructions blindly. Importantly, students were not prohibited from using external generative AI tools; instead, the task was designed such that superficial reliance on AI-generated solutions would be ineffective unless students could evaluate and adapt those solutions appropriately.

The assessment was delivered over a four-week period. Students accessed Liam Chat to generate their individual scripts and datasets, worked independently to correct and implement the database, and submitted their final SQL code via the learning management system. Marking criteria were deliberately simplified and aligned with professional standards: whether the code executed without errors, whether the data was correctly inserted, and whether the final implementation adhered to the business rules specified in the case study.

Staff involvement focused primarily on assessment design, tool configuration and quality assurance rather than ongoing intervention during the assessment period. Once deployed, the automated generation of scripts and data significantly reduced administrative overhead and ensured consistency in assessment difficulty across the cohort, despite the individualisation of inputs.

In terms of implementation, the development of Liam Chat can be achieved through two primary approaches, each with different trade-offs. The approach adopted in this case used OpenAI's API integrated within a ShinyApps.io interface, allowing students to access the tool without needing individual accounts. This supported equitable access but required initial technical expertise, ongoing maintenance and minimal costs associated with API token usage. The initial development was completed over approximately two to three weeks, equating to around 20 to 30 hours of focused development time, including prompt design,

interface setup, testing and refinement to ensure stable and constrained outputs. An alternative approach is to create a custom GPT within ChatGPT, which removes the need for deployment infrastructure and significantly reduces maintenance overhead. However, this requires students to have access to a (free) ChatGPT account. While this may raise initial access considerations, the increasing prevalence of student use of such tools suggests this is becoming less of a barrier. In both approaches, reliable internet access remains a prerequisite. These options provide flexible pathways for implementation depending on institutional context, technical capability and equity considerations.

Outcomes

The implementation of the constrained AI-supported assessment yielded several important outcomes for both students and staff.

Student outcomes

The assessment revealed substantial variation in student understanding and capability. Students who possessed a strong grasp of database design principles were able to diagnose errors, adapt to the flawed inputs and produce valid, functional database implementations. In contrast, many students who relied heavily on generative AI tools without sufficient conceptual understanding struggled to complete the task successfully. This outcome was notable, as AI use was not restricted; rather, the task design itself exposed the limitations of uncritical AI reliance.

The authenticity of the task played a significant role in these outcomes. Students were required to take responsibility for the correctness and consequences of their implementations, reflecting the realities of professional practice. While this resulted in higher failure rates than in previous iterations of the course, it also provided a more accurate representation of student readiness and highlighted areas requiring further instructional support. Informal observations and student discussions indicated that many found the task more challenging than previous assessments, particularly due to the variability and lack of a single “correct” solution path. However, students who engaged deeply with the task reported that it more closely resembled real-world problem-solving and highlighted the importance of understanding, rather than reproducing, SQL logic.

Staff outcomes

From a teaching perspective, the assessment significantly simplified marking and moderation processes. The focus on executable code and adherence to business rules reduced subjectivity and ambiguity in grading decisions. The personalised nature of the assessment inputs effectively eliminated opportunities for collusion, and no cases of plagiarism or academic misconduct were reported.

The model proved highly scalable, accommodating a large cohort without a proportional increase in staff workload. Once established, the assessment design was reusable and adaptable, offering long-term sustainability benefits.

Key messages and transferability

This case study demonstrates that constrained AI can be used to enhance, rather than undermine, assessment authenticity and integrity. By shifting the focus from production to evaluation, educators can design assessments that focus on human judgement, accountability and professional responsibility.

The approach is transferable to other disciplines and levels, particularly where learning outcomes involve interpreting, validating, or correcting complex artefacts. Potential applications include data analytics, programming, systems analysis, business rule modelling, and even non-technical fields where students must critique AI-generated content.

Future development

Future iterations of this work will explore extending Liam Chat's functionality to support requirement gathering through natural language interaction, enabling students to ask clarifying questions and experience simulated client interviews. Further research will also examine how scaffolded AI-supported assessments can be integrated across programmes to progressively develop students' AI literacy, ethical awareness and professional judgement.

References

Chand, S P (2023) 'Constructivism in education: Exploring the contributions of Piaget, Vygotsky and Bruner', *International Journal of Science and Research*, 12 (7): 274-278.

Corbin, T, Dawson, P and Liu, D (2025) 'Talk is cheap: why structural assessment changes are needed for a time of GenAI assessment and evaluation in higher education', *Assessment & Evaluation in Higher Education*, 50 (7): 1087-1097.

Elshall, A S and Badir, A (2025) 'Balancing AI-assisted learning and traditional assessment: the FACT assessment in environmental data science education', *Frontiers in Education*, 10: 1596462.

Khlaif, Z N, Alkouk, W A, Salama, N and Abu Eideh, B (2025) 'Redesigning assessments for AI-enhanced learning: a framework for educators in the generative AI era', *Education Sciences*, 15 (2): 174.

Tam, M (2000) 'Constructivism, instructional design, and technology: Implications for transforming distance learning', *Educational Technology and Society*, 3 (2): 50-60.

Vlachopoulos, D and Makri, A (2024) 'A systematic literature review on authentic assessment in higher education: best practices for the development of 21st century skills, and policy considerations', *Studies in Educational Evaluation*, 83: 101425.

Wiggins, G (1990) 'The case for authentic assessment', *Practical Assessment, Research and Evaluation*, 2 (2).