Tang (2026). *AJBSS, 1*(5)
ISSN 3083 1709 (Online)
ISSN 3083 1717 (Print)

Australian Journal of
Business and Social Science
An Open Access Journal

The AI Classroom: How Generative Tools Are Reshaping Teaching and Assessment

Tang, W. C.
BEng(Hons), MEd, PCEd, MSc, MBA, PhD, MCGI, SMIIM, CMgr, FCMI, FIMA, CPMC, FIMC
Adjunct Professor of Jesselton University College, Sabah, Malaysia
ronald.tang@jesselton.edu.my

Abstract

GenAI has hurtled from experimental curiosity to embedded in education at breakneck speed. This article examines the effects of generative tools (large language models, multimodal generators, and AI tutors) on teaching and assessment and weighs the evidence of pedagogical value against the risks they present. The issue is not whether GenAI is able to produce fluent prose and solve structured problems but rather whether the use of GenAI in education can be regulated in ways that guard epistemic purposes central to education, as evidenced by systematic reviews and meta-analyses, analyses of institutional policy, and investigations of student and faculty perceptions. It points to evidence of some modest benefits to GenAI feedback for thought, to the self-agreement of students that we must reform assessment rather than just care about detection, and to the inequities of connectivity and digital fluency and language/prosody as not side issues but constitutive of the AI classroom. The article acknowledges the detection versus redesign dilemma of evidence-based assessment of GenAI and advances the field in building the capacity to support teaching, assessment, and equity safeguards and reflections on why the field must move toward a "if GenAI works, then under what conditions, by whom, and with what evidentiary warrants?" mindset.

Keywords: academic integrity; AI literacy; assessment redesign; educational equity; generative AI; teaching practice

I. Introduction

Since the release of the November 2022 version of ChatGPT-3.5 (Rojek et al., 2024), institutions have been reading this moment as momentous for education and as triggering the evolution of generative AI (Gen AI) tools reshaping teaching, learning, and assessment. This description captures both the rapidity of adoption and the difficulty of evaluating its impact in a 2025 systematic review of GenAI use in higher education (Belkina et al., 2025). Sugars as a student uses large language models to write essays, generate solutions to problem sets, summarise literature, code; as an instructor uses it to write rubrics, lesson plans, feedback; as an institution ranges from outlawing them to enthusiastic endorsement. The systematic review notes pedagogical uses of GenAI are now handily categorised into learning support, skills development (critical skills), and automated feedback and assessment, whereas challenges are AI literacy and dependency, technical limitations, quality and ethical concerns, and pedagogical challenges.

That terrain is the focus of this article. Its chief contention is that the most significant transformations do not reside in the generation text but in a sort of choreography of epistemic labour (who/what makes claims, who adjudicates, and under what authority). The availability of a new tool is less a change to teaching and assessment than is the adoption of age-old notions of authorship, process, and the evidential value of finished products. For these reasons, the article unfolds thematically, not chronologically. Section 2 discusses pedagogical shifts in the GenAI environment. Section 3 examines assessment redesign as the key institutional response. Section 4 analyses the evidence on AI-generated feedback and learning outcomes. Section 5 considers language representation, equity, and access. Section 6 discusses the changing role of teachers and the need for AI literacy. Section 7 considers policy responses and their limitations. Section 8 discusses research gaps and the research agenda. Section 9 provides conclusions.

II. Pedagogical Shifts: From Content Delivery to Orchestration

The survey of pedagogical uses in higher education (Qian, 2025) reveals a heavy reliance on "text-based tools like ChatGPT" but also an emerging interest in "specialised applications" of GenAI for "automated feedback and assessment, learning supports, and the development of critical skills."

This distribution is telling. Our most popular tools, then, are those that function to reduce the cost of producing plausible (but not necessarily cogent) academic prose. The most obvious pedagogical uses are those that automate labour-intensive ones for instructors: creating practice materials and grading and providing feedback.

A second systematic review of GenAI use in K–12 observes that the technology holds great promise for personalising learning experiences, motivating students, enhancing assessment, and embracing novel and immersive pedagogies (Yusuf et al., 2025). But that review also notes the early evidence base and vast differences in implementation, and opportunity language prevails in the K-12 literature. The review is more tentative regarding the approaches in higher education, concluding with the need for an ethical framework and for innovative assessment strategies and for advocating a collaborative model between GenAI, teachers, and students. Thus, the change is not from one tool to another but from a transmission model of teaching to what is called orchestration: the deliberate design of learning environments in which human and machine inputs are sequenced and evaluated.

This framing of orchestration is in line with the broader literature that argues that the value of GenAI is in the pedagogical design, not the tool. As an upcoming meta-analysis states: GenAI produces half-way decent outcomes in either high school or university settings, but usage differs according to learner plus learning goal (Huang et al., 2026) with the corollary that some (not all) learning goals might be prerequisites when the GenAI-driven pedagogy is designed at various education levels. The implication is that the pedagogical question is not about the use of GenAI, but rather the kind of cognitive labour it displaces. If GenAI is used to produce a first draft for students to assess, then the displaced labour may be low-value transcription. If it is used to generate the answer, then the displaced labour may be the very reasoning the assessment was designed to elicit.

III. Assessment Redesign: The Dominant Institutional Response

GebAI is more likely to drive redesigning assessment than detection. A 2025 study on faculty attitudes toward GenAI in Saudi higher education (Bamasoud et al, 2025) find that redesign of assessment practices will be the best approach to maintain academic integrity, rather than restrictive measures. The same study observes that faculty concerns are more about academic integrity and assessment practices than about technology. So, the issue is not the tool, but an assessment regime (how we assess you, student).

In a 2025 study of 40 universities in six world regions, researchers find that institutions are proactively taking steps to integrate GAI by focusing on enhancing academic integrity, improved teaching and learning, and equity (Dai et al., 2025) and also note common aspects that include: developing guidelines for ethical GAI usage, creating assessments that would be resistant to appropriate misuse, and designing training programs (for faculty and students) to engage the growth of GAI literacy. Authentic assessments (McArthur, 2023) are a shorthand for those tasks that demand process evidence and contextual judgement or embodied performance, which are harder to do with a prompt and a fluent model. Another policy paper calls for redesigning assessments to be AI-proof (Temimi et al., 2025), additional training for staff and students, multilayered enforcement methods, defining acceptable use, while acknowledging that almost half of students (47%) use LLMs in their coursework and that detection software is up to around a high 88% accuracy, which leaves an error margin of 12% (Qian, 2025). The 88% detection accuracy also makes a case against detection: 12% margin for error is not a rock-solid ground for high-stake disciplinary decisions.

But rethinking assessment practice is not a monolithic affair. The range runs from the minor (e.g., in-class handwritten exams, oral defenses, annotated drafts) to the major (e.g., portfolio-based assessment, staged project work, reflective process journals, group work with negotiated judgement). So little is known about redesigns in terms of maintaining validity or reducing cheating. In the Saudi faculties paper, Variation across academic rank suggests that expectations of academic role and academic responsibility contribute to faculty perceptions, with Associate Professors most concerned with 'assessment credibility and student performance' and Professors least concerned. This variation

suggests that assessment redesign is both an institutional problem and a technical problem: who is responsible for credibility (and for student performance) matters, as does who can change standards.

Kooli et al. (2025) report one (quasi-)experimental case of redesign in practice. They present a 'colloquial engagement theory with AI awareness' that integrates informal, contextually situated questioning into the design of assessments to combat AI-assisted cheating and authentically motivate learners to compare students and AI performance on traditional versus colloquial multiple-choice assessments. The idea is that questions that call for idiomatic, situated judgement cannot easily be answered correctly by a generalist model, as readily as questions that promote recall of stock phrases. It is a nice case study of the concept that assessment redesign ought to target the specific type of intellectual work that GenAI is unable to approximate, although the generalisability of any given redesign is still to be tested.

IV. AI-Generated Feedback and Learning Outcomes

Compared to other GenAI literature, feedback is the most fully developed in terms of volume. But a 2024 meta-analysis of 31 experimental and quasi-experimental studies (Hu, 2024) discovered that GenAI feedback probably has a moderately positive effect on academic achievement and identified large and statistically significant moderation of effects by contextual factors. That same analysis also finds GenAI feedback has a more positive effect on cognitive outcomes and promising (but less conclusive evidence of an impact on metacognitive development) as well as modest effects on noncognitive factors. A wider meta-analysis of 53 studies reports positive effects of generative artificial intelligence on academic achievement (Liu et al., 2025), higher-order thinking, motivation, and writing. Another meta-analysis finds GenAI feedback can achieve significantly higher educational outcomes than non-GenAI feedback and that GenAI tools (Dong, 2026) could achieve higher academic achievement, higher order thinking skills, and writing skills compared to traditional technologies.

These are promising findings, but they should be taken with caution. Firstly, many of the included studies use outcome measures that are not standardised measures of durable learning but rather researcher-designed tests or rubric-scored writing tasks. Second, feedback comes in quite a few different varieties: some studies use GenAI to generate feedback aligned with rubrics, others to generate model answers, and others to generate iterative dialogue. Of course, this gives rise to a lot of heterogeneity and therefore the meta-analytic average masks a lot of variability. Thirdly, instead of good human feedback, the preceding conditions tend to be delayed feedback or no feedback. In addition, a meta-analysis of AI versus human feedback points toward a hybrid approach, combining the scalability of AI and the deep, empathetic or contextual features of human feedback. This is a far less exciting finding than a bold claim of AI advantage: the GenAI feedback advantage is costless and timely, not that it is replacing relational pedagogy.

The evidence for metacognitive and non-cognitive outcomes is weaker. This meta-analysis of many studies describes metacognition benefits as promising but less established and non-cognitive outcome benefits as modest. Because the best reasons for GenAI in education concern self-regulated learning (Jin et al., 2026) and reflection and student autonomy. The experimental evidence is not yet quite so flattering.

V. Equity, Access, and Representation

Equity in the AI classroom (Dakakni & Safa, 2023) is more often an afterthought, but there is evidence of how constitutive it is. Analysis of multination qualitative interview data of K–12 teachers in 2026 highlighted three mechanisms of inequity. First, unstable Internet access prevented some students from using GenAI tools outside school even when they could on a device. Second, unequal and limited access to technologies-maintained inequities in students' digital literacy levels, putting low-income students at a disadvantage for not having daily chances to learn basic skills. Third, the culturally and linguistically constituted design of Gen AI systems often reproduced inequities (Zajko, 2022), and teachers stated the outputs tended to reflect dominant linguistic norms, creating Western or U.S.-centric views.

These are not marginal implementation issues. The first is that all assignments that utilise GenAI outside of class are structurally unfair in settings where internet access is dicey. The second means that even where access is equal in principle, the complementary experience students have of digital practice prior to GenAI usage shapes their effective use. The third is that GenAI outputs carry inherent representational meaning that can further disadvantage students whose linguistic or cultural contexts are not well represented in the training data. Access also is uneven because of cost, language, connection, devices as barriers, and policy, as this analysis of access notes: Access to AI learning tools has also been uneven, likely deepening inequity in open learning. Equity is acknowledged in policy literature as a goal but tends to be constructed as fair access to GAI tools, not the knowledge and language encoded in the tools.

The implications for equity in assessment are glaring (Fields & Green, 2025). If assessment redesign involves GenAI-mediated tasks, then students with restricted access to connectivity or devices will be learning unsupervised with students who access GenAI full time. Or if the assessment redesign forbids GenAI altogether, the students who were allowed to become fluent may mysteriously be out of luck, and the ban may not even be enforceable. The fair answer is not a one-size-fits-all policy but a policy of differentiation: to ensure that foundational learning not always rely on GenAI; to provide obvious opportunities for all students to develop forms of literacies around the AI; to confront the representational politics of the devices themselves.

VI. The Changing Role of Teachers and the Demand for AI Literacy

Teacher capacity is consistently mentioned in literature as a binding constraint. A 2025 review of GenAI use in higher education advocates a collaborative model of teachers, students, and GenAI and highlights AI literacy (Tzirides et al., 2024) and dependency issues as core challenges: The pedagogics of teacher professional development also credibly points to enhanced confidence and preparation by training: a study of intentionally designed GenAI literacy modules finds enhanced teachers' confidence and preparedness to ethical and effective GenAI use and reports qualitative findings consistent with a higher AI literacy level. A second qualitative case study of preservice teacher education reports that GenAI has highlighted the necessity of preservice teachers learning technological pedagogical content knowledge (TPACK) and self-regulated learning (SRL) strategies.

These are positive but partial findings. In the literature, the term "AI literacy" is used inconsistently, ranging from basic operational knowledge to critical evaluation of the outputs of models to ethical and political critique of artificial intelligence systems. A policy analysis of 30 Asian universities characterised institutional policies as demonstrating a desire to maintain traditional academic values and practices and to take informed advantage of these models, specifying a focus on applications in text generation, student management, and maintaining academic integrity. The focus on academic integrity is understandable but perhaps obscures attention to pedagogical transformation. In the case of staff-facing guidance, a separate review of 16 UK universities notes that this side of the policy landscape is still much less developed than the student-facing side. Teachers are asked to embed GenAI and redesign assessment, but not always with the conceptual tools to distinguish between and assess uses that can support learning and uses that can substitute it.

So, the role of the teacher is shifting, but not in a single direction. In some, the teacher becomes an architect of learning worlds, designing human and machine input. In others, the teacher is the chief curator of AI-synthesized artefacts and the manager of AI-augmented process (Martino et al., 2025). In still others, the teacher becomes a critical conversational partner who helps students interrogate the outputs of systems whose training data and optimisation targets are opaque. These roles overlap, but each requires different forms of professional knowledge. The evidence about what kinds of professional development work best for what kinds of teachers under what kinds of conditions still does not exist.

Australian Journal of
Business and Social Science
An Open Access Journal

VII. Policy Responses and Their Limits

Policy has certainly advanced at institutional level, but not always in a uniform way. A review of policies across 40 universities internationally (Jin et al., 2025) found that there are still gaps in comprehensive policy frameworks, regarding data privacy concerns and equal access to GAI tools. It also highlights the need for straightforward communication, stakeholder engagement, and continued oversight. Likewise, the Asian policies are points of focus for integrity and adoption, as would be anticipated by such countries as Singapore, at the cost of more emphasis on 'text generation applications' than on pedagogical transformation.

One thread in the policy literature can be summarised as tension between restriction and redesign. A policy guide for school leaders suggests moving away from a punitive mindset toward viewing AI as a tool for supporting learning and teaching and developing a common understanding of what constitutes "human work" and what type of originality. This corresponds to the empirical finding that faculty concerns are not about the technology but the credibility of assessments. As it turns out, the appropriate policy position is not to prohibit nor leave it open but the position expressed through nuanced policy with clear expectations for when and what kind of GenAI use is permitted, assessments designed so the work will produce an artefact of the learning processes, and teacher support for that moment of judgement that determines when GenAI is assisting the learning process and when the use of GenAI is taking place of that learning.

We must also be mindful of the limitations of policy. Detection tools are not without issues, and overreliance on such can result in false accusations and loss of trust. A policy paper points out that detection tools can be 88% accurate and have a 12% error rate (Sun, Liao & Ma, 2026). An error of 12% is unacceptable in high stakes. Hence, detection-oriented policies are both epistemically and ethically troubling. A more promising alternative is to design assessment systems that do not presuppose the possibility of satisfactorily performing the impossible task of distinguishing human and machine authorship.

VIII. Gaps in the Evidence and a Research Agenda

There are some major gaps. First, the evidence base is largely composed of short-term studies of individual courses or interventions. The moderate effects noted in meta-analyses raise questions about the sustainability of such effects and their transferability/interactive effects with prior achievement. Second, there is little definition of GenAI feedback. Few studies describe the prompts, model version, rubric alignment, or human-in-the-loop degree in detail. This clouds what features of feedback stimulate learning. Third, a frequently overlooked outcome: equity. Many studies report an average effect, not effects by access, language, and prior digital fluency. Fourth, the assessment-redesign literature is filled with proposals but lacks validation. There is little evidence about which redesigns are valid, which can constrain misconduct, and which inadvertently penalise certain groups. Fifth, AI literacy among teachers is poorly measured (Zainal et al., 2026), and little is known about the impact of professional development on changes in practice.

A proper research agenda would address these gaps by (a) using longitudinal designs that follow learning beyond a course, (b) providing disaggregated analyses that determine who thrives and who is left behind, (c) validating particular redesigns of assessments, including their effect on student experience and equity, (d) providing rich reporting of GenAI interventions, including prompts, versions of models, and human supervision, and (e) using mixed methods that combine outcome measures with sensitivity to the institutional and cultural conditions in which GenAI is being adopted.

IX. Conclusion

The AI classroom is not a future but a present, and it is a present that will be distributed and evaluated unevenly. The evidence reviewed here suggests some cautiously optimistic visions of GenAI's potential for supporting feedback and some cognitive outcomings, a vociferous position of fear regarding detection-based responses to integrity issues in academia, and a strong warning that things such as equity cannot be an afterthought in these discussions. The question of whether GenAI "works"

is too simple; the most pedagogically defensible formulation asks under what circumstances and with whom, with what evidence warrants. It requires designs that view assessment as learning, teacher capacity as something other than operational fluency, and institutional policies in which equity is a design principle rather than a compliance checkbox. The danger is not that GenAI will displace teachers or render assessment impossible; the danger is that GenAI will be adopted quickly before we have designed the assessment frameworks, professional knowledge, and equitable conditions that excellence in integration demands.

References

Bamasoud, D. M., Mohammad, R., & Bilal, S. (2025). Adopting generative AI in higher education: A dual-perspective study of students and lecturers in Saudi universities. *Big Data and Cognitive Computing*, *9*(10), 264.

Belkina, M., Daniel, S., Nikolic, S., Haque, R., Lyden, S., Neal, P., ... & Hassan, G. M. (2025). Implementing generative AI (GenAI) in higher education: A systematic review of case studies. *Computers and Education: Artificial Intelligence*, *8*, 100407.

Dai, Y., Lai, S., Lim, C. P., & Liu, A. (2025). University policies on generative AI in Asia: Promising practices, gaps, and future directions. *Journal of Asian Public Policy*, *18*(2), 260-281.

Dakakni, D., & Safa, N. (2023). Artificial intelligence in the L2 classroom: Implications and challenges on ethics and equity in higher education: A 21st century Pandora's box. *Computers and Education: Artificial Intelligence*, *5*, 100179.

Dong, Y. (2026). Generative AI technologies and educational outcomes: a comprehensive meta-analysis comparing traditional and AI-driven approaches. *Humanities and Social Sciences Communications*.

Fields, N. I., & Green, E. (2025). Ensuring equity, access, and opportunity for all youth to thrive. In *Positive youth development: Integrating research and practice* (pp. 209-224). Cham: Springer Nature Switzerland.

Hu, S. (2024). The effect of artificial intelligence-assisted personalized learning on student learning outcomes: A meta-analysis based on 31 empirical research papers. *Science Insights Education Frontiers*, *24*(1), 3873.

Huang, Y., Chen, S., Zhang, W., & Chen, M. (2026). Can Generative AI Feedback Effectively Enhance Learning Outcomes? A Meta-Analysis of 36 Experimental and Quasi-Experimental Studies. *Education Sciences*, *16*(6), 816.

Jin, Y., Osman, K., & Arsad, N. M. (2026). The effects of generative artificial intelligence on self-regulated learning: a meta-analysis. *Information*, *17*(10), 1028.

Jin, Y., Yan, L., Echeverria, V., Gašević, D., & Martinez-Maldonado, R. (2025). Generative AI in higher education: A global perspective of institutional adoption policies and guidelines. *Computers and Education: Artificial Intelligence*, *8*, 100348.

Kooli, C., Yusuf, N., & Sarhan, M. Y. (2025). Colloquial engagement theory with AI awareness (CET-AIA): A new creative pedagogical framework for ethical assessment in the age of artificial intelligence. *Thinking Skills and Creativity*, 102051.

Liu, X., Guo, B., He, W., & Hu, X. (2025). Effects of generative artificial intelligence on K-12 and higher education students' learning outcomes: A meta-analysis. *Journal of Educational Computing Research*, *63*(5), 1249-1291.

Martino, D., Perlangeli, C., Grottoli, B., La Rosa, L., & Pacella, M. (2025). A Knowledge-Driven Framework for AI-Augmented Business Process Management Systems: Bridging Explainability and Agile Knowledge Sharing. *AI*, *6*(6), 110.

McArthur, J. (2023). Rethinking authentic assessment: work, well-being, and society. *Higher education*, *85*(1), 85-101.

Qian, Y. (2025). Pedagogical applications of generative AI in higher education: A systematic review of the field. *TechTrends*, *69*(5), 1105-1120.

Rojek, M., Kufel, J., Bielówka, M., Mitręga, A., Kaczyńska, D., Czogalik, Ł., ... & Bartnikowska, W. (2024). Exploring the performance of ChatGPT-3.5 in addressing dermatological queries: a research investigation into AI capabilities. *Dermatology Review*, *111*(1), 26-30.

Sun, Y., Liao, Y., & Ma, X. (2026). Trusting AI to detect AI? A systematic evaluation of the reliability and robustness of current AIGC detection tools for student academic work. *Computers & Education*, 105616.

Temimi, A., Rebai, R., Aldamen, H., Souki, K., & Amiri, K. (2025). Education at a crossroads: will AI fix learning or break it?. *Management & Sustainability: An Arab Review*, 1-25.

Tzirides, A. O. O., Zapata, G., Kastania, N. P., Saini, A. K., Castro, V., Ismael, S. A., ... & Kalantzis, M. (2024). Combining human and artificial intelligence for enhanced AI literacy in higher education. *Computers and Education Open*, *6*, 100184.

Yusuf, H., Money, A., & Daylamani-Zad, D. (2025). Pedagogical AI conversational agents in higher education: a conceptual framework and survey of the state of the art. *Educational technology research and development*, *73*(2), 815-874.

Zainal, M. A., Matore, M. E. E. M., & Maat, S. M. (2026). Assessing teachers' AI literacy: a systematic review of measurement tools. *Interactive Learning Environments*, 1-29.

Zajko, M. (2022). Artificial intelligence, algorithms, and social inequality: Sociological contributions to contemporary debates. *Sociology Compass*, *16*(3), e12962.