# ASCILITE 2024

## Navigating the Terrain:

*Emerging Frontiers in Learning Spaces, Pedagogies, and Technologies*

# "All things are ready, if our mind be so": Attitudes to STEM assessment redesign in the age of genAI

**Christopher Bridge**
University of New England

**Ryan Naylor**
University of Sydney

**Birgit Loch**
Federation University

In the two years since the launch of ChatGPT in November 2022, the higher education sector in Australia has experienced both the great potential of generative artificial intelligence (genAI) to transform learning and teaching, but also the serious threat to assessment security and assurance of learning that it poses. In light of recent regulatory calls for accountability in academic integrity, this article focuses on the issue of assessment security and redesign, given the affordances of readily available large language models. A nation-wide survey of STEM academics revealed that 37% of respondents had not even tested their assessments in a genAI app; that there was little consensus – and possible misunderstandings – concerning approaches for improving assessment security; and that sector recommendations were either not well understood, or not favoured, particularly those advocating a move towards program-level assurance of learning and only securing assessments at key points of a degree program. These results highlight the need for significant support and guidance for teaching academics—in terms of both creating a clear understanding of how best to respond to assessment redesign as future consensus emerges, and institutional workload in implementing those responses —in order for challenges posed by genAI to assurance of learning to be met successfully.

*Keywords:* generative artificial intelligence, assessment redesign, academic integrity, assessment security, STEM, large survey

## Introduction

Two years since the advent of ChatGPT, with a raft of competitors and updates having emerged in the meantime, it is timely to revisit the question: How ready are we for generative artificial intelligence (genAI)? Shakespeare's Henry V found that "all things are ready, if our mind[s] be so" was a motto for success at the Battle of Agincourt, but, two years on, is the higher education sector sufficiently prepared for the new era in which genAI is transforming how we work and learn? We present here potentially concerning data suggesting that in STEM, we may not be.

The fact that to date only the University of Sydney has publicly released its response to the recent TEQSA request for information, *Addressing the risk of artificial intelligence* (TEQSA, 2024; University of Sydney, 2024), speaks of uncertainty in the sector, as do continuing debates about balancing the implications of generative AI as a tool versus as a threat to learning (Lodge et al., 2023). The primacy of this debate is demonstrated by the fact that Lodge et al. (2023) has already attracted 21 scholarly citations (Google Scholar) in under 10 months, but large numbers of academics may still not be engaging with the discussion. We present the results of a survey of STEM academics' attitudes to genAI which show that over one-third of respondents had not even checked the performance of ChatGPT or similar on their assessments, and that STEM academics are at best lukewarm concerning sector recommendations for addressing assessment security.

While genAI poses enormous potential for positively transforming learning and teaching in higher education, in this article we focus on genAI as a threat to academic integrity, and the associated compliance and assurance of learning issues. This positioning is intentional within our study design, as questions of compliance and assessment design may be more grounded than vaguer, philosophical questions about the role of genAI in supporting learning that have not yet neared consensus. In this lens, the advent of genAI is the latest development in a series which includes cheating sites, contract cheating, and paraphrasing software, that has made assessment redesign as urgent as it is unavoidable. Two years into this revolution, with "wait and see" no longer an option, it is apparent our minds may not be sufficiently ready.

## Literature review

Soon after the launch of ChatGPT in November 2022, it became clear that genAI has the ability to pass a wide range of academic assessments. Li et al. (2023) summarized literature reporting on the ability of ChatGPT to pass third-year medical exams, as well as assessments in parasitology, pathology, microbiology and law. Since then, the capabilities of large language models have only improved. On the other hand, the vast potential for genAI to transform education in positive ways has been recognised. Yan et al. (2024), for example, identify 53 educational processes that could benefit from genAI, while Baidoo-Anu and Ansah (2023) highlight the implications for personalised, formative education.

Ensuring academic integrity is nevertheless a regulatory requirement for higher education providers in Australia. The Australian Government Tertiary Education Quality and Standards Agency (TEQSA) requires adherence to, among others, the following standards: that "[s]tudents are only awarded a degree when they have demonstrated achievement of the learning outcomes", that "[p]reventative action is taken to mitigate foreseeable risks to academic integrity", and that "[a]cademic oversight assures the quality of teaching and learning by maintaining oversight of academic integrity and monitoring of potential risks" (Gniel, 2023). This has prompted both regulatory and scholarly guidance on reforming assessment design to assure these requirements.

**Sector recommendations for assessment reform**

Recommendations for addressing academic integrity in an era of readily accessible genAI were discussed in a wide range of Australian higher education forums during 2023. Key amongst these was the document *Assessment reform in the age of AI* (Lodge et al., 2023) published by TEQSA, first as a consultation paper in September 2023, then as a finalised document in November 2023. This publication included the propositions "a systematic approach to program assessment aligned with disciplines/qualifications" (p. 4) and "security at meaningful points across a program to inform decisions about progression and completion" (p. 6).

Programmatic assessment aims to assure program-level learning outcomes using a methodology which requires multiple data points from a range of different assessment activities; an academic panel which makes decisions on attainment of learning outcomes from the collated data, such that high-stakes decisions require more data points than low-stakes judgments; the possibility for students to track their own progress towards attaining learning outcomes, to ensure assessment for learning rather than merely of learning, entailing the provision of high-quality feedback; and regular consultation with mentors to scaffold self-regulation of learning and provide remediation where students are not on track towards attainment of outcomes (van der Vleuten et al., 2012; Van Der Vleuten et al., 2015). However, even in the discipline of medicine where the approach is well-established, moving towards programmatic assessment is recognised as challenging. Khanna et al. (2023) have described the process as a "complex socio-cultural change" (p. 846), which needs to overcome both student and staff preconceptions concerning assessment. Students participating in their study struggled with the concept of different levels of stakes associated with different assessments, and for staff the need for multiple data points came with the risk of overassessing.

Though writing just before the advent of ChatGPT, Dawson (2021) has discussed a range of approaches to assessment security that are relevant in the era of genAI, including programmatic assessment, dialogic feedback, random security checks, and surveillance approaches to detection such as stylometric monitoring and content matching with a cumulative data base of a student's work.

A number of other recommendations emerged in the first half of 2023 on semi-formal university forums such as webinars and learning and teaching blogs, responding to the challenge to assessment security posed by genAI. These included focussing assessment on higher order skills, such as critical thinking (e.g. Liu, Brown, et al., 2023; OpenAI, 2023), incorporating non-text-based artefacts in assessments (e.g. Hillier, 2023; Tang et al., 2023), linking an assessment to individual experience (e.g. Kozar et al., 2023; Liu, Ho, et al., 2023), and ensuring that assessment is authentic (e.g. Liu & Bridgman, 2023; see also the scholarly article Swiecki et al., 2022). Elsewhere, Bearman and Ajjawi (2023) argued in favour of pedagogy in the age of genAI involving the development of evaluative judgement through interactions with genAI.

We refer in this paper to the suggestions and conclusions from these texts as the "sector recommendations" for assessment reform, for want of a better expression, recognizing that many have not received any broad endorsement, and noting that none are specific to STEM education or STEM-specific assessment security threats. The example of programmatic assessment highlights the many levels of preparedness that need to be in place for major assessment reform to take place. We argue in this paper that not even the direction is clear, let alone preparations in place for the transformation in assessment practices than can no longer be delayed.

## Methodology

A survey investigating attitudes of STEM academics in Australian universities towards genAI was run from 29 October 2023 until 17 April 2024. The survey was designed with input from STEM Associate Deans Learning and Teaching attending the 2023 Australian Council of Deans of Science Learning and Teaching Leaders Conference, held at Monash University, Melbourne, 20-21 June. The aim of the study was to gather information to inform institutional decision-making in response to the genAI revolution. Recruitment was via Associate Deans Learning and Teaching forwarding an email invitation to participate in the survey to the STEM academics in their respective faculties. The considerable length of time for which the survey remained open was partly due to it having been launched late in the year, resulting in only a few responses. The decision was taken to hold it open and proceed with follow-up recruitment emails (anticipated in the ethics application) in the new year. There were 99 valid completions in this time. Because recruitment was via intermediaries we do not know how many invitations were sent, hence we are unable to provide a response rate. Ethics approval was granted by the University of New England's Human Research Ethics Committee (approval no. HE23-127).

In addition to demographic data and a question designed to gauge general familiarity with genAI in a higher education setting, "Have you tried putting (at least some of) your assessment questions into a generative AI app, to see how well the app performs?", this study primarily makes use of the results of one section of the survey, related to academic integrity. In this section the first question asked, "Has generative AI made you rethink your approach to securing your assessments against cheating? How?" Textual responses were categorized under the themes in Table 1, determined inductively by the authors. While the first question was bound to the academic's institutional context, the second question sought a general response, asking, "In your opinion, what assessment formats are acceptably secure against cheating in your unit(s), given that generative AI apps are now widely available?" Textual responses were likewise categorized by theme. The final question considered here asked, "Which of the following might be effective strategies for securing your unit(s) and/or degree(s) against cheating, given that generative AI apps are now widely available?" and respondents were asked to provide an ordinal rating alongside a selection of recommended strategies arising from formal and semi-formal sources in the first half of 2023 (described in the literature review, above). Respondents were also given the option of selecting "Don't know what this is". This question included the possibility of selecting "other" and providing a textual response.

## Results

The 99 respondents represented a broad range of STEM disciplines: in descending order these were Biological Sciences (20), Mathematical Sciences (14), Chemical Sciences (12), Physics and Astronomy (10), a total of 9 from a range of Engineering disciplines, Computer Science (9), Information Systems (6), Environmental Studies (6), Other Information Technology (4), Other Natural and Physical Sciences (2), Earth Sciences (1), and Other Agriculture, Environmental and Related Studies (1). Twenty-four had 25 or more years of teaching experience, 23 had under 10, 26 had 10-15, and 20 had 16-24 years of experience. Sixty-one held Teaching and Research positions, 29 were Teaching Focused, and 2 each held Research Only and other positions. Thirty-seven were academic level B, 20 were D, 16 were E, 14 were C, and 7 were A, and 4 were other. Twenty-three Australian universities were represented, though not evenly, with 61 respondents coming from 6 universities.

### Familiarity

Over one-third (37%) of respondents had not tried putting at least some of their assessment questions into a genAI app. Accepting that some assessments, such as skill-based evaluations, may not be amenable to genAI completion, we still found this rate to be high. That the survey was open for 5½ months provided the opportunity for a longitudinal comparison, to see whether this engagement rate had perhaps increased over time. But this rate was the same (37%) for the first 50% of respondents as for the second 50%. Given that this was a voluntary survey on genAI for STEM academics, and that those engaged with the issue were therefore more likely to participate, we found this rate of non-engagement concerning.

### Rethinking assessment security

In response to the question asking whether genAI had led to a rethink of assessment security and if so how, 19 respondents (23%) answered no. A further 10 respondents (12%) answered yes but either provided no elaboration concerning how, or indicated that they had not yet found a solution. Of those who answered yes and described how, textual responses were coded into the categories listed in Table 1. The most common response was invigilation (13, i.e. 23% of those who responded with yes and offered a description of how). A further 5 (9%) mentioned oral assessment.

Table 1
*Approaches Put Forward by Those Agreeing That GenAI Had Made Them Rethink Their Approach to Securing Their Assessments Against Cheating.*

| | n | % of those agreeing and describing how |
|---|---|---|
| Invigilation | 13 | 23% |
| Assess higher order skills | 7 | 12.5% |
| Greater specificity | 7 | 12.5% |
| Detect AI | 5 | 9% |
| Oral | 5 | 9% |
| Use class/project/individual data | 4 | 7% |
| Incorporate genAI | 4 | 7% |
| Assess process not outcome | 3 | 5% |
| Authentic assessment | 3 | 5% |
| Non-text-based assessment | 2 | 4% |
| Practical assessment | 2 | 4% |
| Raise awareness of genAI | 2 | 4% |
| Groupwork | 1 | 2% |

There were a range of other approaches to securing assessments against genAI put forward. These included assessing higher order skills, such as focussing on “critically evaluating and social skills”; not merely providing an answer but also justifying it and discussing other options; and requiring “some additional thinking or

research even if they are still using AI as a part of that process". Another approach could be described as demanding greater specificity in assessment tasks and rubrics, for example "[c]hanging topic focuses to specific things, writing marking rubrics (so that is what the student uses to identify where marks come from), including correct referencing as a requirement", and designing assessment tasks to be "very context specific". A further approach put forward was to design assessment that made use of data generated in class, or as the result of a project, or individual data. Responses coded into this category included assessments that are "practical-based/application based using data generated in class" and "individual student projects where they collect their own data". This category is distinguished from the greater specificity classification in that the latter refers to modifying an existing assessment through tightening assessment requirements, whereas the former encompasses changing to an alternative assessment format. Assessing process rather than outcome was also identified as an approach to securing assessments against genAI use, for example in a coding assessment not just requiring the program to function correctly, but also to meet design and structuring standards. Another approach in this category required students to bring drafts to class in the weeks prior to submission. Other approaches put forward included detecting AI use; incorporating AI use; authentic assessment; non-text-based and practical assessment; groupwork; and raising awareness of academic integrity related to genAI use.

The low rate of interest in incorporating genAI into assessments may also be indicative of low engagement by STEM academics with genAI, or it may be a sign that this is not considered a way of improving assessment security. It is also worth noting that of the 13 respondents who opted for invigilation, 5 mentioned invigilation in combination with other measures, for example invigilated exams alongside written assessments that are "practical-based/application based using data generated in class". One respondent described a combination of major assessments that were invigilated in-person, and minor formative tasks that remained unsecured.

**Secure assessment formats**

The above responses encapsulated realistic plans within institutional constraints, however the survey also gave respondents the opportunity to provide a hypothetical, unconstrained opinion, through the question: "What assessment formats are acceptably secure against cheating in your unit(s)?" In answer to this question there was a clear preference in favour of invigilation (39 or 49% of respondents) and oral assessments (21 or 26%). The remaining categories ranged in descending order from non-text-based assessments (11 or 14%), practical assessments, greater specificity in assessment questions, projects, assessing higher skills, incorporating genAI into the assessment, dialogic feedback, presentations, reflections, authentic assessment, using class/project/individual data, detecting AI use, and finally assessing process not outcome, with only 1 respondent (1%).

**Response to sector recommendations**

Another aim of the survey was to test the receptiveness of STEM academics to emerging ideas concerning assessment security in the age of genAI, not so much as a means of influencing institutional directives, but rather as a way of scoping the attitudinal landscape in which institutional policy will need to be implemented, to identify where staff development resources might need to be concentrated.

These sector recommendations, outlined in the literature review above, were compiled and consolidated (without further evaluation of veracity), and their receptiveness amongst STEM academics probed using the question: "Which of the following might be effective strategies for securing your unit(s) and/or degree(s) against cheating, given that genAI apps are now widely available?" Response options were: "Don't know what this is", and a 5-point ordinal scale ranging from *Not effective at all* to *Extremely effective*. These responses are presented in Table 2.

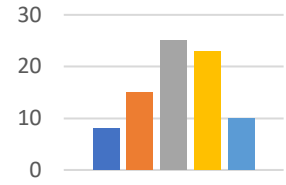
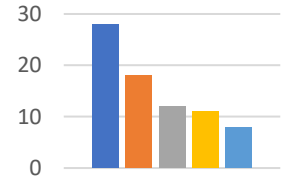
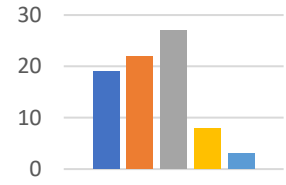
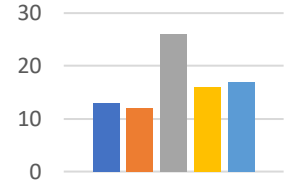
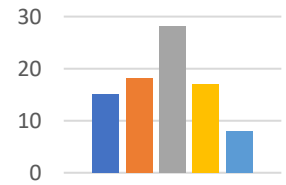
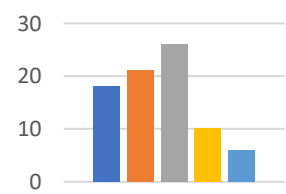
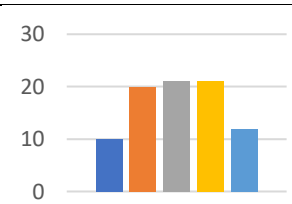
Table 2

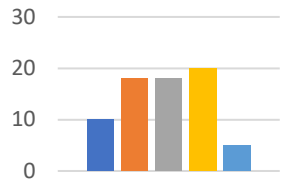
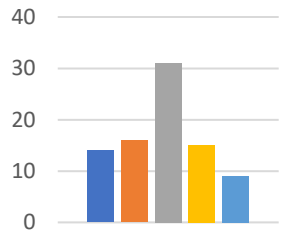
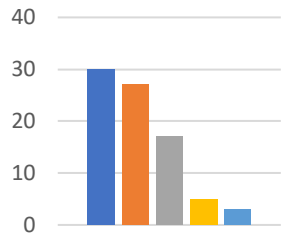
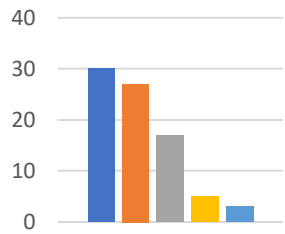
*Sector Recommendations Rated for Effectiveness at Securing Assessments Against Cheating in the Age of GenAI*

# ASCILITE 2024

## Navigating the Terrain:

*Emerging Frontiers in Learning Spaces, Pedagogies, and Technologies*

| | Don't know what this is | Ordinal distribution<br>Not effective at all · Slightly effective · Moderately effective · Very effective · Extremely effective | Item weighted average |
|---|---|---|---|
| 1. Assessing higher order skills (and allowing students to use AI and other tools for lower order skills) | 6.9% |  | 3.1 |
| 2. Requiring students to write entire assignments on a platform which records keystroke analytics (e.g. Cadmus) | 10.5% |  | 2.4 |
| 3. Moving towards programmatic assessment* and only securing assessment items that evaluate degree-level learning outcomes | 9.2% |  | 2.4 |
| 4. Authentic assessments (assessment tasks modelled on typical workplace tasks) | 3.4% |  | 3.1 |
| 5. Assessing evaluative judgement* of one's own or peers' work, rather than just production of the work | 1.1% |  | 2.8 |
| 6. Saving students' written work as they progress through their degree in a database to enable cross-checking of personal stylistic metrics in their latest submission (e.g. through Turnitin's Authorship tool) | 6.9% |  | 2.6 |
| 7. Redesigning assessments to incorporate non-text-based artefacts | 2.3% |  | 3.1 |

| | | | |
|---|---|---|---|
| 8. Using dialogic feedback* both as a formative teaching tool and for monitoring authorship | 18.4% |  | 2.9 |
| 9. Linking an assessment to individual experience (e.g. writing a reflection on a class discussion, keeping a logbook of progress on an assignment over time) | 2.3% |  | 2.9 |
| 10. Random security checks, i.e. using some of the methods described in this and the previous question but only on randomly selected students | 5.7% |  | 2.1 |
| 11. Only securing assessment items related to progression and completion of a degree | 14.0% |  | 2.3 |

*Further explanation provided in the survey (see below).

A number of terms used in the survey were also supplied with definitions. Programmatic assessment was defined as “assessment that is conceptualised not as a series of one-off, disconnected and haphazard tasks, but as a continuous, connected and planned programme … under a programmatic approach all degree outcomes must be met to pass the programme” (Dawson, 2021, p. 137). Evaluative judgement was defined as an “understanding of what quality work looks like and the ability to make decisions about the quality of their own work and the work of others” (Dawson, 2021, p. 21). And dialogic feedback was given the definition: “an ongoing conversation between the student and other people, such as teachers and peers, towards improvement of the student’s work” (Dawson, 2021, p. 138).

Respondents also had the opportunity to select “Other” in response to this question and offer a textual response. While many of these responses fell within the categories already given above, the most common exceptions were various forms of invigilation (14 responses), which was not one of the options which could be selected.

## Discussion

The overall impression from these data is one of directionlessness. There is no clear indication from STEM academics in favour of one or a set of preferred options: the most common effectiveness rating for sector recommendations was *Moderately effective*, with relatively little differentiation between the different items.

No options reached consensus as "very" or "extremely" effective. Although the sample size is relatively small (n=99), we speculate that those STEM academics who were more concerned about or interested in genAI as of early 2024 were more likely to participate. Consequently, our findings may overestimate the overall level of understanding and engagement with genAI among the national population of STEM academics.

It is also clear that pragmatic concerns influenced the responses, with 23% indicating a change in their assessments to invigilation in their institutional context, but 49% selecting invigilation as an acceptably secure format when the question was posed in a general sense. At the same time it is noteworthy that even this latter figure represents only half of the STEM academics who participated in the survey. Note, however, that keystroke analytics—which may be regarded as a highly specific form of invigilation—was one of the least supported sector recommendations (item average 2.4).

Though support for various approaches to assessment security was at best lukewarm, two tentative conclusions may characterize attitudes towards assessment reform amongst our sample of STEM academics. Firstly, intrusive security methods such as Cadmus, Turnitin's Authorship tool and random security checks were not favoured (item averages 2.1-2.4); secondly a range of more progressive assessment alternatives, including assessing higher order skills, authentic assessment, assessing evaluative judgement, incorporating non-text-based artefacts, dialogic feedback, and linking assessment to individual experience, were somewhat more favoured (item averages 2.8-3.1). Having said that, two of the TEQSA recommended approaches, introducing program-level assessment and only securing assessments at key points in the program, were either not well understood or disfavoured (item averages 2.3-2.4).

As noted, this survey does not evaluate the validity of any of these approaches in improving assessment security. Regarding item 4 of Table 2, authentic assessment, the work of Bretag et al. (2019) in demonstrating that assessment authenticity is not a guarantee of academic integrity is especially noted. The increasing sophistication of genAI in interpreting photos and diagrams also calls into question the effectiveness of item 7, incorporating non-text-based artefacts, in improving assessment security. This highlights the issue that in a rapidly evolving technological context, some perceived approaches to assessment reform may only have limited shelf-life, potentially indicating increased importance of training, professional development, and partnership with specialist expertise.

The particularly high level of professed ignorance concerning dialogic feedback (item 8), despite a definition having been supplied, is also noteworthy. For those who understood it, it scored comparatively well, although a number of respondents commented elsewhere in the survey that they felt approaches requiring more workload were unrealistic. One comment in the "other" category of this question on rating sector recommendations was "More contact with students. This is expensive but basically the only thing that really works."

We acknowledge that our conclusions are drawn from a limited data set, and may not be generalisable to STEM academics across Australia. We also recognize that the attitudes reported here do not necessarily identify approaches to assessment reform that will ultimately prove to be effective regarding academic integrity and assurance of learning. However, we hope that these results may provide indicative information about the attitudinal landscape in which STEM assessment reform will need to take place, and point towards the type of professional development and support that will be necessary.

## Conclusion

To return to our question posed in the introduction, how ready is the higher education sector for the genAI revolution, the findings from this study have not provided much cause for optimism, particularly from an academic integrity and assurance of learning perspective. Amongst the STEM academics who participated in our survey, there appears to be little agreement on the path forward: a significant proportion have not even engaged with genAI, let alone formed a conclusion on how to secure their assessments. Of those who had

formed opinions, there seemed to be scant confidence in the robustness of current assessment reform recommendations, as evidenced by the finding that most items were only seen to be slightly to moderately effective. It is particularly concerning that two of the TEQSA recommendations related to assessment (Lodge et al., 2023) were amongst the least favoured – or least understood – options for assessment reform amongst STEM academics. As an emerging field, without substantial consensus on effective assessment reform (Lodge et al., 2023), this confusion may be understandable, but it highlights the importance of effective and ongoing professional development and training as a clearer understanding emerges. Additionally, a significant minority are not engaged with the need for assessment redesign sufficiently to test their own assessment tasks with genAI. Two years into the genAI age, it should be clear that "wait and see" is no longer an option.

The limitations of this study mean that any recommendations we derive for the higher education sector can only be tentative. Nevertheless, policy development, even in a new and still fluid context, must be informed by the outcomes of initial experience, research and evidence, and sensitive to preconceptions and misconceptions on the ground. Professional development of teaching academics, and indeed, cultural change concerning understandings of curriculum and assessment design, may also be required. These changes could be effected through partnerships between academics and different fields of expertise, including educational design and genAI professionals, but also through nurturing discussion and collaboration among academics at departmental and institutional levels, as well as inter-institutionally, through communities of practice, conferences and other forums. Though it was not tested in this study, it is also possible that policy will need to accommodate discipline variation, which STEM-specific solutions unlikely to correspond one-on-one to other disciplines.

Finally, we must keep in mind that assurance of learning and ensuring our students learn are importantly different goals. Assessment design drives student behaviour—increased invigilation, for example, may have different implications to programmatic assessment in students adopting transactional versus relational or learning community-based approaches to learning. Clear understandings among academics, in STEM and beyond, are essential to ensure assessment reform provokes the desired learning culture. The near future will determine whether "our mind is ready." Whether we will echo the triumphant tone of Shakespeare's history, Henry V, or lament a missed opportunity, in the words of Hamlet: "The readiness is all."

## References

Baidoo-Anu, D., & Ansah, L. O. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI*, *7*(1), 52-62. https://doi.org/10.61969/jai.1337500

Bearman, M. & Ajjawi, R. (2023). Learning to work with the black box: Pedagogy for a world with artificial intelligence. *British Journal of Educational Technology, 54*(5), 1160-1173. https://doi.org/10.1111/bjet.13337

Bretag, T., Harper, R., Burton, M., Ellis, C., Newton, P., van Haeringen, K., Saddiqui, S., & Rozenberg, P. (2019). Contract cheating and assessment design: exploring the relationship. *Assessment & Evaluation in Higher Education, 44*(5), 676-691. https://doi.org/10.1080/02602938.2018.1527892

Dawson, P. (2021). *Defending assessment security in a digital world: Preventing e-cheating and supporting academic integrity in higher education*. Routledge. https://doi.org/10.4324/9780429324178

Gniel, H. (2023). *AI: A regulatory perspective*. Tertiary Education Quality and Standards Agency. https://www.teqsa.gov.au/sites/default/files/2023-04/AI-a-regulatory-perspective-2023.pdf

Hillier, M. (2023, March 3). Assessment tweaks in response to generative artificial intelligence. *TECHE*. https://teche.mq.edu.au/2023/03/assessment-tweaks-in-response-to-generative-artificial-intelligence/

Khanna, P., Roberts, C., Burgess, A., Lane, S., & Bleasel, J. (2023). Unpacking the impacts of programmatic approach to assessment system in a medical programme using critical realist perspectives. *Journal of Critical Realism, 22*(5), 840-858. https://doi.org/10.1080/14767430.2023.2279805

Kozar, O., Kirwan, M., Tse, N., Hess, N., Bradford, W., Kumar, N. K., Bodis, A., & Mackaway, J. (2023, March 6). From quizzes to essays: A campus-wide assessment update. *TECHE*. https://teche.mq.edu.au/2023/03/from-quizzes-to-essays-a-campus-wide-assessment-update/

Li, Y., Sha, L., Yan, L., Lin, J., Raković, M., Galbraith, K., Lyons, K., Gašević, D., & Chen, G. (2023). Can large language models write reflectively. *Computers and Education: Artificial Intelligence, 4*, 100140. https://doi.org/10.1016/j.caeai.2023.100140

Liu, D. & Bridgman, A. (2023, January 23). How can I update assessments to deal with ChatGPT and other generative AI? *Teaching@Sydney*. https://educational-innovation.sydney.edu.au/teaching@sydney/how-can-i-update-assessments-to-deal-with-chatgpt-and-other-generative-ai/

Liu, D., Brown, M., & Miller, B. (2023, March 8). How Sydney academics are using generative AI this semester in assessments. *Teaching@Sydney*. https://educational-innovation.sydney.edu.au/teaching@sydney/how-sydney-academics-are-using-generative-ai-this-semester-in-assessments/

Liu, D., Ho, E., Weeks, R., & Bridgman, A. (2023, January 20). How AI can be used meaningfully by teachers and students in 2023. *Teaching@Sydney*. https://educational-innovation.sydney.edu.au/teaching@sydney/how-ai-can-be-used-meaningfully-by-teachers-and-students-in-2023/

Lodge, J. M., Howard, S., Bearman, M., Dawson, P., & Associates (2023). *Assessment reform for the age of Artificial Intelligence*. Tertiary Education Quality and Standards Agency. https://www.teqsa.gov.au/sites/default/files/2023-09/assessment-reform-age-artificial-intelligence-discussion-paper.pdf

OpenAI (2023, August 31) *Teaching with AI*. https://openai.com/index/teaching-with-ai

Swiecki, Z., Khosravi, H., Chen, G., Martinez-Maldonado, R., Lodge, J. M., Milligan, S., Selwyn. N., & Gašević, D. (2022). Assessment in the age of artificial intelligence. *Computers and Education: Artificial Intelligence*, 3, 100075. https://doi.org/10.1016/j.caeai.2022.100075

Tang, P., Yu, S., Wright, E., & Rennie, M. (2023, March 14). 4 UTS case studies of Generative AI in assessments. *LX at UTS*. https://lx.uts.edu.au/blog/2023/03/14/ai-in-assessments-4-case-studies/

TEQSA (2024, July 5). *Request for information: Addressing the risk of artificial intelligence*. https://www.teqsa.gov.au/guides-resources/higher-education-good-practice-hub/artificial-intelligence/request-information-addressing-risk-artificial-intelligence

University of Sydney (2024). *TEQSA request for information – artificial intelligence: The University of Sydney action plan*. https://educational-innovation.sydney.edu.au/teaching@sydney/wp-content/uploads/2024/07/USYD-TEQSA-RFI-response.pdf

van der Vleuten, C. P., Schuwirth, L. W. T., Driessen, E. W., Dijkstra, J., Tigelaar, D., Baartman, L. K. J., & Van Tartwijk, J. (2012). A model for programmatic assessment fit for purpose. *Medical Teacher, 34*(3), 205-214. https://doi.org/10.3109/0142159X.2012.652239

Van Der Vleuten, C. P., Schuwirth, L. W. T., Driessen, E. W., Govaerts, M. J. B., & Heeneman, S. (2015). Twelve tips for programmatic assessment. *Medical Teacher, 37*(7), 641-646. https://doi.org/10.3109/0142159X.2014.973388

Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., & Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology, 55*(1), 90-112. https://doi.org/10.1111/bjet.13370



Bridge, C., Naylor, R., & Loch, B. (2024). "All things are ready, if our mind be so": Attitudes to STEM assessment redesign in the age of genAI. In Cochrane, T., Narayan, V., Bone, E., Deneen, C., Saligari, M., Tregloan, K., Vanderburg, R. (Eds.), *Navigating the Terrain: Emerging frontiers in learning spaces, pedagogies, and technologies*. Proceedings ASCILITE 2024. Melbourne (pp. 234-243). https://doi.org/10.14742/apubs.2024.1426

Note: All published papers are refereed, having undergone a double-blind peer-review process.


The author(s) assign a Creative Commons by attribution license enabling others to distribute, remix, tweak, and build upon their work, even commercially, as long as credit is given to the author(s) for the original creation.

© Bridge, C., Naylor, R., & Loch, B. 2024