JUTLP
Journal of University Teaching & Learning Practice

Reimagining the Artificial Intelligence Assessment Scale (AIAS): A refined framework for educational assessment

Mike Perkins [a], Jasper Roe [b], Leon Furze [c],
[a] British University Vietnam, Vietnam; [b] Durham University, United Kingdom; [c] Deakin University, Australia

Abstract

Higher education institutions, educators, and students continue to grapple with the wide-ranging implications of Artificial Intelligence (AI) and Generative Artificial Intelligence (GenAI). The ability of advanced GenAI models to complete educational assessments continues to be one of the most pressing issues for academia. In early 2024, we published the AI Assessment Scale (AIAS), which describes a practical framework for addressing GenAI use in educational assessment. The AIAS was well received and has been implemented in hundreds of institutions worldwide and translated into 30 languages. Building on our experience using the AIAS and drawing on feedback and critiques received worldwide, in this paper, we provide an updated version of the AIAS. The updated AIAS encompasses several important developments, including clarifying the theoretical basis of the scale through social constructivist principles, developing the AIAS to function as an assessment redesign framework to enhance validity, and adjusting the visual assets of the AIAS to function inclusively and non-hierarchically. Furthermore, in anticipation of future developments in the field of AI and the increasing focus on preparing students to function in an AI-augmented world, we introduce an additional level, 'AI exploration'. Through this paper we also seek to clarify both the intended and unintended applications of the AIAS. We provide implementation guidance through practical examples and explain how the AIAS functions as both a communication tool and a framework for task redesign that strengthens validity.

Editors

Section: Educational Technology
Editor in Chief: Dr Joseph Crawford
Senior Editor: A/Prof Rachel Fitzgerald

Publication

Submission: 15 April 2025
Revised: 12 August 2025
Accepted: 25 September 2025
Online First: 27 September 2025

Practitioner Notes

1. The AIAS provides a structured yet flexible framework for redesigning assessments in the age of ubiquitous Generative AI.
2. The AIAS features five distinct levels ranging from "No AI" to "AI Exploration", allowing educators to select the appropriate level of the scale based on learning outcomes and context and redesign their assessments appropriately.
3. The AIAS should be used as a tool to guide assessment redesign rather than to restrict students' use of GenAI through detection or prohibition.
4. Educational institutions should address GenAI through policy development that embraces transparency and focuses on authentic assessments rather than surveillance approaches.
5. When implementing the AIAS, educators should consider equity of access to GenAI tools and explicitly teach ethics and privacy, including data handling, consent, disclosure of AI assistance and intellectual property.

Citation: Perkins, M., Roe, J., & Furze, L. (2025). Reimagining the Artificial Intelligence Assessment Scale (AIAS): A refined framework for educational assessment. *Journal of University Teaching and Learning Practice, 22*(7). https://doi.org/10.53761/rrm4y757

JUTLP
Journal of University Teaching & Learning Practice

Keywords

AI Assessment Scale (AIAS), Generative Artificial Intelligence (GenAI), Academic Integrity, Assessment Redesign.

JUTLP
Journal of University Teaching & Learning Practice

Introduction

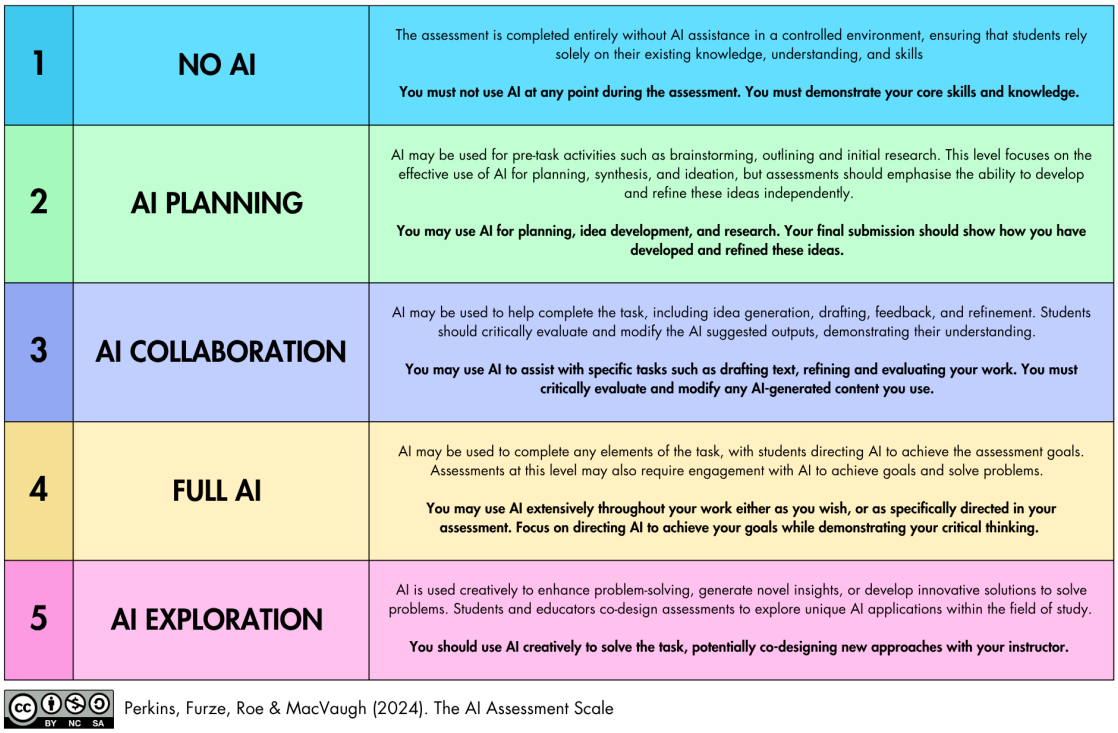
The release of ChatGPT in late 2022 marked a turning point in education. Within months, generative AI (GenAI) tools were producing text, images, and multimodal outputs that challenged established assumptions about academic work. As these technologies spread into classrooms and universities, particularly in relation to assessment, educators and institutions adopted a wide spectrum of responses ranging from prohibition and detection to full integration. With continued advances and broader availability, recognition has grown that both educators and students must learn to engage productively with these tools rather than attempt to resist them.

By 2023, calls for guidelines on how students should ethically use GenAI were proposed (Cotton et al., 2023; Perkins, 2023; Crawford et al., 2023). Multiple frameworks have since been developed to guide educators and students on how to integrate (or restrict) the use of GenAI tools in assessment practices. Our framework, the AI Assessment Scale (AIAS), is one such example (Perkins, Furze, et al., 2024). The goal of this framework was to offer a way for educators to have open dialogue with students about the values of educational integrity, prioritise a transparent approach to how GenAI could be used in assessment, and serve as a way to redesign assessment tasks for the GenAI era. We sought to foreground ethicality, transparency, and a nuanced, rather than a black-and-white understanding of these new technologies.

The AIAS has been adopted globally, demonstrating the need for practical tools to address current educational challenges. The AIAS has been translated into 30 languages and is used in over 300 educational institutions worldwide (Furze, 2024). Since its release, we have sought to gather and reflect on the various implementations and contexts in which the AIAS has been implemented, spanning different educational environments, types of assessment, and considerations of culture and inclusivity (Perkins et al., 2025). This holistic process of feedback gathering, discussion, and iterative development resulted in the decision to update the AIAS which we present and discuss in this paper. The updated version of the AIAS is shown in Figure 1.

JUTLP
Journal of University Teaching & Learning Practice

Figure 1
The Updated AIAS



1	NO AI	The assessment is completed entirely without AI assistance in a controlled environment, ensuring that students rely solely on their existing knowledge, understanding, and skills **You must not use AI at any point during the assessment. You must demonstrate your core skills and knowledge.**
2	AI PLANNING	AI may be used for pre-task activities such as brainstorming, outlining and initial research. This level focuses on the effective use of AI for planning, synthesis, and ideation, but assessments should emphasise the ability to develop and refine these ideas independently. **You may use AI for planning, idea development, and research. Your final submission should show how you have developed and refined these ideas.**
3	AI COLLABORATION	AI may be used to help complete the task, including idea generation, drafting, feedback, and refinement. Students should critically evaluate and modify the AI suggested outputs, demonstrating their understanding. **You may use AI to assist with specific tasks such as drafting text, refining and evaluating your work. You must critically evaluate and modify any AI-generated content you use.**
4	FULL AI	AI may be used to complete any elements of the task, with students directing AI to achieve the assessment goals. Assessments at this level may also require engagement with AI to achieve goals and solve problems. **You may use AI extensively throughout your work either as you wish, or as specifically directed in your assessment. Focus on directing AI to achieve your goals while demonstrating your critical thinking.**
5	AI EXPLORATION	AI is used creatively to enhance problem-solving, generate novel insights, or develop innovative solutions to solve problems. Students and educators co-design assessments to explore unique AI applications within the field of study. **You should use AI creatively to solve the task, potentially co-designing new approaches with your instructor.**

 Perkins, Furze, Roe & MacVaugh (2024). The AI Assessment Scale

The remainder of this article explores the current developments in GenAI, education, and assessment before we reflect on the different applications of the AIAS thus far, and in doing so, explore and address the feedback and critiques that the framework has received. In the third section, we introduce the revised version of the AIAS, explaining how and why the decisions were made to change and augment it. Finally, we consider the future of assessment and attempt to justify why the AIAS may continue to be a useful tool for guiding assessment strategies and reform in the coming years.

GenAI, Education and Assessment: The Story So Far

Although the terms AI and GenAI are often conflated, we distinguish between Artificial Intelligence (AI) and Generative Artificial Intelligence (GenAI). AI refers to systems that exhibit ‘intelligent behaviour by analysing their environment and taking actions—with some degree of autonomy—to achieve specific goals’ (European Commission, 2018, p. 4). In contrast, GenAI ‘creates new content in response to prompts based on its training data" (Lorenz et al., 2023, p. 8). For

educators, the pressing task is to discern which parts of an assignment reflect a student's own effort and which have been shaped by GenAI, enabling targeted feedback, appropriate integrity checks and informed assessment redesign.

Since the initial release of ChatGPT, GenAI technology has become much more capable across an ever-broadening range of writing, mathematical, and scientific reasoning tasks. Initial academic integrity concerns were related to a focus on written assessments, such as essays (Cotton et al., 2023; Perkins, 2023). However, the integration of vision capabilities into multiple GenAI systems, alongside rapid developments in the ability of these tools to generate high-quality text, images, and videos, meant that this rapidly developed to broader risks to academic integrity with the unchecked use of the tools in various fields. Research has shown that GenAI tools can perform on par with humans in a wide variety of tasks, including high-stakes professional examinations in the fields of medicine, law, and science (Head & Willis, 2024; Newton et al., 2024). GenAI models can now handle multimodal content effectively (Shahriar et al., 2024) and spend more time 'thinking' or 'reasoning', enabling them to achieve top percentile scores in mathematical and coding competitions and challenges (OpenAI, 2024). This means that assessment tasks that were initially thought to be more resistant to GenAI tools are also vulnerable to completion by these tools.

Alongside these models, a growing ecosystem of specialised GenAI tools supports discrete stages of academic work. Examples include Deep Research for literature exploration, Otter.ai for high-accuracy transcription, Scribe for content summarisation, Research Rabbit for iterative literature mapping, GitHub Copilot for code generation, and creative platforms such as Canva AI or Runway for the rapid production of multimodal artefacts. The wide availability of such tools lowers the barrier for students to integrate GenAI throughout the assessment lifecycle, from brainstorming to final submission.

GenAI Detection: A Red Herring?

In response to the advent of GenAI tools and their ability to destabilise many forms of educational assessment, bans on the use of GenAI were quickly followed by the introduction of GenAI text detection tools brought to the market by several private corporations. GenAI text detectors are commonly used as adversarial tools to identify academic integrity violations by students, but these tools are unreliable (Elkhatat et al., 2023; Perkins, Roe, et al., 2024; Sadasivan et al., 2023; Weber-Wulff et al., 2023). This unreliability, and the potentially harrowing consequences for students falsely accused of using these tools that may impact their academic and personal lives (Gorichanaz, 2023; Roe, Perkins, & Ruelle, 2024), means that they cannot be recommended for use in summative assessment.

The development of adversarial discourse is also common in university policies, which frame the use of GenAI as misconduct (Luo (Jess), 2024a), and among instructors, who may now view students' produced texts more critically and with suspicion (Farazouli et al., 2023), leading

students who are already using GenAI tools in ethical and allowed ways to feel the need to hide their usage (Gonsalves, 2024). When educators are not transparent about how GenAI is used or evaluated in the grading and moderation process, students' trust in assessment declines (Luo (Jess), 2024b). However, there seems to be an expanding awareness of the fruitlessness of this approach, as it seems likely that hybrid writing or co-creation may soon become the new norm, and trying to distinguish between AI and non-AI-enhanced writing could become a meaningless task (Eaton, 2023). Ultimately, GenAI text detection is impossible and inefficient (Mao et al., 2024), and leads to a perspective of the assessment process which is overly adversarial and focused on security theatre rather than genuine attempts to support assessment security.

No Turning Back

There is a growing understanding among many educators that, for better or worse, we must learn to live with the reality that GenAI will be used by many members of our society, including students, teachers, and researchers.

Although we are aware of the considerable environmental costs of AI development (Driessens & Pischetola, 2024) and the known existential risks of unregulated AI (OECD, 2024), pragmatically speaking, we do not believe that resisting the development of AI is a tenable solution. Research in the open-source community (Luccioni et al., 2024) shows promise in identifying and mitigating environmental concerns, and we are hopeful that practices such as smaller, more efficient models, on-premises 'local' language models, and a higher degree of accountability in the sector will improve the future of AI. We contend that avoiding technology reduces our ability to realise the potential benefits of GenAI in learning and to prepare students for the critique of the technologies that they will encounter to varying extents throughout their personal and professional lives. This new awareness in scholarship, policy, and practice reflects the idea that GenAI is not always a threat to assessment (Pearce & Chiavaroli, 2023) and that not all uses of GenAI tools are automatically considered misconduct (Bjelobaba et al., 2025; Eaton, 2024).

Assessment redesign has become a major focus in preparing for a future AI-enabled world (Mao et al., 2024; Thanh et al., 2023; Thompson et al., 2023; Xia et al., 2024), and must consider two competing imperatives. On the one hand, educators must consider how to incorporate GenAI to ensure that students develop skills relevant to the future AI-enabled workforce. However, it remains essential to limit GenAI and other forms of outside influence in assessments aimed at verifying students' genuine mastery of specific learning outcomes. Balancing these two imperatives is extremely challenging, both from an institutional and individual educator perspective, especially when it comes to designing assessments that incorporate or recognise the reality of GenAI support, but that are also able to adequately test learning outcomes. To support educators in this challenge, common themes have emerged in the assessment design literature. These include focusing on student-centred (Hsiao et al., 2023), authentic (Rasul et al., 2024), learning processes instead of the end product, testing higher-order thinking skills

JUTLP
Journal of University Teaching & Learning Practice

(Smolansky et al., 2023), and evaluative judgement (Bearman et al., 2024) wherever possible, and avoiding assessment via single examination or essay (Gorichanaz, 2023).

This recognition has helped advance frameworks for integrating GenAI into assessments, such as the AIAS. However, the AIAS is not the only framework available for integrating GenAI into education, and multiple approaches have been proposed. In medical education for example, Pearce & Chiavaroli (2023) suggest that medical education relies on an 'assisted' and 'unassisted' dichotomy, one which allows external resources including GenAI and one which does not, with a refocusing on neglected assessment types such as oral examinations. Other useful frameworks include the two-lane (Liu & Bridgeman, 2023), six-lane (Steel, 2024) and traffic light (Cotterell, 2024) approaches, with the University of Sydney's two-lane model gaining prominence in the Australian sector.

The Original AI Assessment Scale

The core premise of the AIAS is that banning GenAI is not a productive or tenable approach. Lodge's (2023) assertion that the future of assessment requires students to be able to participate in an AI-reliant society sums up this position well: whether we personally use or refuse GenAI in our classrooms, students will encounter the technology in many aspects of their lives. However, a practical problem that causes significant distress for practitioners and colleagues around the world is how to support academic integrity and ensure that students' submitted work reflects their true abilities. Any framework or potential solution must also be flexible, easy to understand and explain, and adaptable across different educational contexts, given the significant variety of assessment tasks across disciplines, levels, and institutional norms.

The AIAS was designed to address these challenges in two key ways. First, it cultivates sustained, explicit dialogue with learners about why, when, and how GenAI may be used. Transparency and dialogue regarding GenAI usage from both an educator and student perspective normalise responsible practice and make covert misuse harder to rationalise. Second, it supports the redesign of assessment tasks so that GenAI is integrated with a clear purpose, always anchored to the learning outcomes being measured. This integration should also include an awareness of when GenAI use is not helpful for an assessment task and, therefore, when a controlled assessment environment may be required to achieve this.

Rather than adopting a binary approach of either allowing or prohibiting GenAI use, and given that GenAI was an emerging technology, we believed that a more nuanced framework was needed to help both educators and students understand the technology's potential strengths and limitations. The framework was structured into five distinct levels of use, allowing educators to structure their conversations with students and design assessment tasks around these five levels.

The AIAS, as originally conceived, is shown in Figure 2.

JUTLP
Journal of University Teaching & Learning Practice

Figure 2
The Original AIAS

1	**NO AI**	The assessment is completed entirely without AI assistance. This level ensures that students rely solely on their knowledge, understanding, and skills. **AI must not be used at any point during the assessment.**
2	**AI-ASSISTED IDEA GENERATION AND STRUCTURING**	AI can be used in the assessment for brainstorming, creating structures, and generating ideas for improving work. **No AI content is allowed in the final submission.**
3	**AI-ASSISTED EDITING**	AI can be used to make improvements to the clarity or quality of student created work to improve the final output, but no new content can be created using AI. **AI can be used, but your original work with no AI content must be provided in an appendix.**
4	**AI TASK COMPLETION, HUMAN EVALUATION**	AI is used to complete certain elements of the task, with students providing discussion or commentary on the AI-generated content. This level requires critical engagement with AI generated content and evaluating its output. **You will use AI to complete specified tasks in your assessment. Any AI created content must be cited.**
5	**FULL AI**	AI should be used as a 'co-pilot' in order to meet the requirements of the assessment, allowing for a collaborative approach with AI and enhancing creativity. **You may use AI throughout your assessment to support your own work and do not have to specify which content is AI generated.**

JUTLP
Journal of University Teaching & Learning Practice

Theoretical Underpinning of the AIAS

Broadly, our view of the AIAS is grounded in social constructivist principles, particularly Vygotsky's (1978) understanding that learning is mediated by social interaction. From this perspective, GenAI tools can be viewed as mediating technologies that may support students' knowledge construction, similar to Vygotsky's conceptualisation of language and cultural tools as mediators of higher mental functions. The different levels of the AIAS reflect varying degrees of mediated learning, from independent demonstrations of knowledge to the full integration of GenAI as a collaborative partner in the learning process. Importantly, this perspective of GenAI use centres on the social experiences of the learner and the educator and avoids what we consider to be problematic understandings of GenAI chatbots as tutors for personalised learning, where students learn in isolation from their peers (Bewersdorff et al., 2023). Simultaneously, we acknowledge that students may use GenAI in isolation and beyond institutional oversight; therefore, the AIAS focuses on purposeful task design and redesign rather than attempts at control.

Central to our theoretical framing is Vygotsky's (1978) concept of the Zone of Proximal Development (ZPD), the gap between what learners can achieve independently and what they can achieve with support from others. GenAI tools can function within this zone by providing scaffolding to help bridge the gap between current and potential performance. However, unlike traditional scaffolding, which is gradually removed over time, AI tools remain available as part of the learner's toolkit, requiring careful consideration of how they can support rather than replace learning. This aligns with the AIAS's emphasis on transparency and appropriate technology use rather than restriction and control.

The social constructivist framework also helps explain why the AIAS emphasises support rather than restriction of GenAI use. From this perspective, learning occurs through interactions with educators, peers, technologies and wider environments, and knowledge is actively constructed rather than received. The AIAS operationalises this by, at higher levels, setting tasks that ask students to critique model capabilities and outputs, use GenAI to create and refine new artefacts, and collaborate in groups where GenAI has an explicit role. This is not unrestricted use; tasks are structured so evaluation can focus on specified elements such as judgement, verification processes, and collaborative decision-making, supported by brief evidence like prompt rationales, checking notes, or short peer or oral reviews.

This social constructivist approach aligns with the view of GenAI content as "digital plastic", simultaneously enabling and potentially harmful. Critical AI literacies (CAIL) are central in this framing. CAIL is defined here as the ability to critically analyse and engage with AI systems by understanding their technical foundations, societal implications, and embedded power structures, while recognising their limitations, biases, and broader social, environmental, and economic impacts (Roe et al., 2025, p. 2). Framed this way, CAIL, like the AIAS, moves beyond binary perspectives on GenAI use and emphasises the contextual nature of these technologies.

JUTLP
Journal of University Teaching & Learning Practice

Critique of the AIAS

The first iteration of the AIAS, as a flexible and useful tool to challenge an urgent threat, has been demonstrated to be effective in empirical studies for improving student outcomes, supporting thoughtful assessment redesign, and reducing misconduct through reframing of GenAI use (Furze et al., 2024). However, through practical use among multiple contexts, from K-12 to Higher Education, and across five continents, we have been able to develop a deep understanding of the areas in which the AIAS can be further developed considering the changing nature of GenAI.

One of the areas identified for development of the AIAS is the separation of the framework to differentiate between K-12 and higher education (Kılınç, 2024). While we recognise that the AIAS is broad in scope, the intention was to provide a tool that could form a basis for adaptation in multiple contexts at different educational levels. To this end, while it would be unlikely at a younger level to encourage the unscaffolded use of GenAI tools to meet assessment tasks, it is dependent on the educator to make such decisions and use the framework in line with their context and goals. Furthermore, adaptations of the AIAS can be developed freely in different contexts, such as in an adaptation to the English for Academic Purposes (EAP) classroom (Roe, Perkins, & Tregubova, 2024) and to structure classroom activities (not just assessment). To clarify this point, we contend that the AIAS is a starting point for educators to consider how to reform their assessments, and is not a prescriptive tool.

A further criticism of the original AIAS was the use of traffic light colours, ranging from red through three shades of yellow/amber to green. It has been noted that this use of colours is problematic, with red potentially implying a ‘stop’ or ban approach, and green having a positive implication such as success (Kılınç, 2024). There are also potential connotations of red as “bad” and green as “good” which were unintended: we do not believe that any type of assessment is inherently better or worse than any other, only that they should vary dependent on context. In the updated AIAS, the decision was made to move away from this system and provide a more neutral palette of colours, first to avoid such implications, and second to ensure legibility and accessibility. We also note that there are legitimate defences for the use of a red-amber-green ‘traffic light’ system, such as for ease of clarity with younger students (Cotterell, 2024) and in situations where supervised and more prescriptive assessments may be possible, such as in the comparatively smaller class sizes and more often face to face settings of K-12 education versus the large and more commonly online or hybrid higher education cohort. As with all AIAS resources, we encourage educators to adapt and make their own decisions based on the context, expertise, and needs of the learners.

Pratschke (2024) identified that the AIAS is beneficial for discussing GenAI use, but that eventually, full transformation of assessment will be required to cope with the changes posed by these technologies. At the time of development, the original AIAS was intended to be used to restructure assessment in depth; however, because of the urgent need to accommodate new technologies, it also gained traction as a way to augment existing practices.

JUTLP
Journal of University Teaching & Learning Practice

A critique of scale-based approaches to GenAI integration comes from Liu and Bridgeman's (2023) two-lane approach to GenAI assessment developed at the University of Sydney. They proposed that assessments should be divided between secured 'lane 1' assessments that assure learning outcomes, and 'lane 2' assessments that embrace human-AI collaboration. They argue there is no "*viable middle ground*" between these lanes and that any unsecured assessment must assume GenAI use will occur and cannot be controlled. While we acknowledge this reality, the AIAS serves a different purpose. Rather than policing use, the AIAS clarifies the role GenAI should play in a task and points educators to the redesign decisions that flow from that choice. In this sense, the AIAS is complementary to the two-lane model since they speak to different aspects of assessment design (AIAS) and assessment security (two-lane). A comparison of the AIAS and two-lane model is shown in Table 1.

Table 1

The AIAS in comparison to the two-lane model

AI Assessment Scale	Two-Lane
Level 1 – No AI	Lane 1 – Secure Assessments
Level 2 – AI Planning	Lane 2 – Open Assessments
Level 3 – AI Collaboration	
Level 4 – Full AI	
Level 5 – AI Exploration	

A related critique is that the AIAS is a discursive response to GenAI use. Corbin, Dawson and Liu (2025) distinguish discursive changes, which rely on rules students can ignore, from structural changes, which alter task mechanics, and caution against the 'enforcement illusion' that results from labels without redesign. We agree that labels alone are insufficient, and we have always positioned the AIAS as both a communication device and a scaffold for redesign.

In our original paper, we note that 'this progressive approach requires faculty to consider how assessments may need to be restructured' (Perkins, Furze, et al., 2024, p. 6), and that the AIAS is to be used to 'help educators consider how their assessments might need to be adjusted' (p2). At the same time, the discursive versus structural distinction is beneficial: we note that the AIAS must be part of a broader strategy of assessment reform, rather than merely adopted without any change to what is being assessed and how. Evidence from pilot studies exploring implementations of the AIAS supports this reading of the AIAS as a structural change, not only to assessment tasks, but also in the broader institutional context as assessment mechanics and policies were adjusted (Furze et al., 2024).

JUTLP
Journal of University Teaching & Learning Practice

Evolution of the AIAS

Key revisions and rationale

Drawing from critiques, implementation experiences, and considerations of assessment validity and transparency, we made several significant refinements to the second iteration of the AIAS. Most visibly, we replaced the traffic light colour scheme used in the original AIAS, adjusted the levels of the AIAS as shown in Figure 1, and introduced an additional circular representation of the framework, shown in Figure 3.

Figure 3
The Updated AIAS: Circular Representation

This design emphasises that no level is inherently better than others, aligning with our focus on appropriate rather than restrictive GenAI use based on specific learning outcomes and contexts. The framework was also updated to reflect the rapid evolution of GenAI technologies, particularly acknowledging the increasing prevalence of multimodal content generation across text, images, audio, and video.

The revisions reflect a deeper understanding of assessment validity and transparency in the GenAI era. Dawson et al. (2024) argue that validity in assessment must represent a student's true

level of knowledge and skills, thus providing assurance of learning. Following this principle, we removed the previous Level 3 requirement for students to include an appendix showing their original work, acknowledging that the current detection methods cannot reliably verify such submissions. This aligns with Bearman et al.'s (2024) emphasis on evaluative judgement, focusing on students' ability to assess and identify quality work, rather than attempting to control technology use. The original version of the AIAS was recognised by the Australian Tertiary Education Quality and Standards Agency (TEQSA) (Lodge, 2024) as a potential tool to help delineate the appropriate uses of GenAI in assessment tasks, and was commended in the Prof Tracey Bretag Prize for Academic Integrity 2024 (Studiosity, 2024) This early validation of the framework's value in promoting transparency informed our continued emphasis on clear communication in this revised version.

The potential for GenAI tools to be ethically used for text generation is now recognised at earlier levels of the scale, reflecting the reality that co-writing with GenAI has become a common practice and may support learning when appropriately scaffolded (Dhillon et al., 2024; Nguyen et al., 2024; Wiboolyasarin et al., 2024). In Level 5, we have strengthened the emphasis on co-design between educators and students, recognising that, as these technologies evolve, the most innovative applications often emerge from the collaborative exploration of GenAI tools in educational contexts. These modifications maintain the AIAS's core purpose as a framework for assessment redesign and to support transparent dialogue between students and educators, while adapting to technological developments and emerging educational practices.

To support the understanding of the various levels of AIAS, we provide examples of assessments which may be suitable at each level through a series of "vignettes" which are composites of the authors' experiences and feedback from educators around the world who have used the AIAS. Vignettes can provide real-life insights, rather than mere suggestions and can help educators contextualise and frame others' experiences through their own expertise (Sampson & Johannessen, 2020).

Level 1: No AI

This level represents a traditional and controlled assessment environment in which GenAI tools are strictly prohibited, and the prohibition is enforced by removing access to digital technologies. To complete a task, students must rely entirely on their existing knowledge, skills, and understanding of the topic. This approach ensures that the assessment directly measures students' unaided capabilities and their core competencies. It is particularly suitable for evaluating foundational knowledge and the ability to articulate ideas without technological assistance and for the application of core knowledge in authentic or performative contexts, such as workplace scenarios, practical skill demonstrations, or viva voce tasks. In programs with external certification or licensure requirements, Level 1 should be used when unaided performance is expected for safety, professional responsibility, or regulatory assurance, for example, in medicine, engineering, or accounting. These assessments should be conducted in controlled environments to protect

their validity and provide assurance of learning. We also recommend staging such No AI assessments at key points in a program to maintain overall assessment validity.

The revision of Level 1 represents a significant shift from the original conceptualisation, moving beyond the simple prohibition of GenAI tools to a more nuanced understanding of when and why a GenAI-free assessment is appropriate. This refinement emerged from practical implementation experiences and feedback from educators implementing the AIAS, which revealed that merely declaring an assessment 'GenAI-free' without environmental controls was increasingly untenable in contemporary educational settings. Therefore, a key modification in the revised framework is the explicit emphasis on controlled environments, reflecting Dawson et al.'s (2024) argument that assessment validity must take precedence over traditional notions of academic integrity. In practice, this means that Level 1 assessments should occur in supervised settings where the absence of AI can be assured, rather than relying on honour systems or detection tools for take-home work. This presents challenges for online and distance learning environments, where truly AI-free conditions are difficult to guarantee without physical supervision. Some learning outcomes may be better served by allowing strategic GenAI use, thereby making other levels of the framework more appropriate.

A critical consideration in Level 1 implementation is the balance between security and accessibility. While these assessments preclude AI use, they must accommodate students who require assistive technologies. This requires a careful distinction between AI tools that can compromise assessment validity and assistive technologies that enable equitable participation. The framework acknowledges that blanket bans on technology can create unnecessary barriers for students with disabilities and that assessment design should consider accessibility from the outset rather than as an afterthought. As we progress into an increasingly 'postplagiarism' world (Eaton, 2023), the ability of educational institutions to truly create an entirely controlled assessment environment may soon disappear entirely. This requires a major reconsideration of the relationship between assessment security and validity.

Example of Level 1 Assessment

In a writing course, a faculty member determines that a particular rhetorical skill cannot be validly assessed if students have access to GenAI. The students' demonstration of these skills, such as control of sentence structure, grammar, and rhetorical devices, would be obfuscated by software such as Grammarly, ChatGPT, or Microsoft Word with the Copilot function. As such, the faculty determines that a 'No AI' assessment is required. The students complete a written exercise by hand under supervision. For students requiring assistive technologies, word processor technologies (such as an e-ink tablet with a keyboard) or a scribe are provided.

Level 2: AI-Assisted Planning

The revision of Level 2 represents a shift in the recognition of how GenAI can support the planning and ideation phases of assessments. While the original framework emphasised basic

brainstorming capabilities, the revised Level 2 acknowledges GenAI's broader role in supporting preliminary work while focusing on students' active development of ideas. Simply declaring that AI can be used 'only for planning' in traditional assessments is insufficient; tasks must be explicitly designed to evaluate students' ability to develop, refine, critique, and expand ideas generated through AI assistance.

A key consideration in Level 2 implementation is how GenAI can support rather than restrict learning processes. Rather than attempting to control GenAI use through unreliable detection methods as we discuss earlier, Level 2 emphasises supporting students in developing their planning and ideation skills with GenAI as a collaborative tool. This requires carefully structured assessment components that distinguish between GenAI-supported planning phases and independent development, while acknowledging that a strict delineation between these phases may not be technically feasible. Assessment design should acknowledge that while we cannot technically restrict GenAI use to the planning stages, we can create tasks that meaningfully engage students in demonstrating their development of ideas.

Assessments designed at this level of the AIAS may permit or prohibit GenAI content in submissions. The distinguishing factor is that the assessment itself is a planning task: either pre-work that feeds a later assessment or an activity that requires students to use GenAI to generate a broad set of ideas and then refine, critique, and develop them. Tasks should be redesigned so that the planning is visible and assessable, for example, through prompt rationales, planning notes, or storyboard artefacts aligned to the learning outcomes. Simply tagging an unchanged assignment as *'Level 2'* will be ignored and does not improve validity; the design of the task does.

Example of Level 2 Assessment

In a Media or Film course, an interim assessment is developed in which students create a comprehensive storyboard and plan for a documentary to be produced as a final project. The assessment task is focused on the storyboard and contributes 25% of the total final grade for the unit, with the remaining 75% coming from the production of the documentary. For the storyboard, the course leader decides that GenAI use at Level 2 is acceptable because it is possible to make a valid judgement of the students' abilities to plan and design a shot sequence, whether they use GenAI or not. Furthermore, this task can be supported by a variety of multimodal GenAI tools, including research (e.g. Perplexity, ChatGPT), sketch-to-image generation (e.g. Firefly, ControlNet), and image-to-video generation to create moving storyboards (e.g. Runway, Kling, Sora). Students are explicitly taught how to use these tools and are assessed on their understanding of storyboarding conventions, shot sequencing, pace, and other elements. Faculty members may also decide to assess the use of GenAI in this process.

Level 3: AI-Assisted Task Completion

The revision of Level 3 represents a significant shift from the original conception, moving beyond simple editing to acknowledge the reality of GenAI-assisted drafting and composition. While the

original framework attempted to restrict GenAI use to surface-level improvements of students' original work, the revised Level 3 recognises that attempting to enforce such limitations is neither practical nor pedagogically sound. Instead of focusing on authorship verification, Level 3 emphasises the development of critical evaluation skills and the maintenance of student voice in GenAI-assisted writing. Although technological solutions exist that purport to show how students develop their work, such as document tracking or version control systems, they are fallible and can be circumvented by determined students. More importantly, the emphasis on 'proving' original authorship can create inequitable conditions, particularly disadvantaging students without access to more advanced (and paid-for) GenAI models or technical knowledge of how to effectively integrate GenAI assistance while maintaining apparent independence.

Supporting students at this level must consider how to overcome the *illusion of finality* - the tendency to accept AI-generated text as complete and authoritative. Students must be supported in understanding the limitations of GenAI systems and recognising that their own knowledge, critical thinking skills, and subject expertise are essential for producing high-quality academic work. This involves developing critical AI literacy (CAIL) skills (Roe et al., 2025), identifying potential errors or biases, and effectively integrating AI-generated content with their own insights and understanding to ensure that their voice is retained. The formative use of GenAI detection tools might also support this process, not as a means of catching misconduct but as a way of helping students understand how their writing style may be influenced by GenAI and how to retain their distinctive voice.

Example of Level 3 Assessment

In a science faculty, educators determine that students can effectively use GenAI as part of the write-up of an earlier practical experiment. Prior to the assessment task, students collect data and conducted practical work. AI is used in the write-up of this data in numerous ways. These include the use of AI-assisted editing tools (e.g. Grammarly and Microsoft Copilot in Word), the use of GenAI for data analytics and assistance with understanding data (e.g. Microsoft Copilot in Excel, Claude, or ChatGPT's data interpretation functions), or generating short exemplars in the required style that students use to guide their own drafting. During instructional time, either face-to-face, online, or via asynchronous modules, the instructor teaches students how to use the relevant applications in this context, for example, demonstrating how MS Copilot in Excel can be used with Python for data analysis.

Level 4: Full AI

Level 4 represents a significant shift in the integration of GenAI into assessments, focusing on the strategic deployment of GenAI tools to achieve specified learning outcomes. While this level previously represented the greatest degree of GenAI integration in the original scale, its refinement acknowledges that as GenAI capabilities expand, particularly with multimodal tools, students need opportunities to develop a sophisticated understanding of when and how to

leverage these technologies effectively. Level 4 supports students' ability to direct GenAI tools while demonstrating critical thinking and subject knowledge. Unlike lower levels, where GenAI use is more constrained, Level 4 assessments evaluate how effectively students leverage GenAI to solve problems and demonstrate understanding. This aligns with Dawson et al.'s (2024) emphasis on assessment validity, where the focus shifts from controlling GenAI use to ensuring that students demonstrate genuine learning through their strategic deployment of GenAI tools.

A key consideration at this level is the equity of access to GenAI tools. While providing specific tools through shared institutional accounts may help address some inequities, educators must carefully design assessments that remain valid regardless of students' access to advanced GenAI models. This might involve providing institutional GenAI tools or designing tasks that can be completed effectively using freely available resources. Although this may help address equity concerns, we acknowledge that it remains an imperfect solution. Nevertheless, transparency about permitted or required tools helps create clearer expectations while potentially reducing access barriers. Level 4 opens up opportunities for multimodal assessment designs that reflect the expanding GenAI capabilities. Beyond text generation, students might engage with GenAI-assisted video creation, synthetic media production, or combinations of multiple modes to achieve outcomes that would be greater than those the student could realistically create, given time restrictions. Assessments at this level might also ask students to engage in scenario-based exercises or role-plays which involve GenAI participants, use Agentic AI tools, or use or even develop their own CustomGPTs or equivalent alternatives to work through problems in structured ways.

Example of Level 4 Assessment

The teachers of a third-year university computer science course determine that students would be best served by learning and having access to the most cutting-edge GenAI-assisted coding tools, many of which are already deployed in the software design industry (for example, GitHub Copilot, Claude Code, or OpenAI's Codex platform). The institution provides students access to these technologies during the course so that they can develop skills in using them prior to assessments taking place. Students may complete this assessment using any GenAI-assisted technology available to them, but they are explicitly taught to use the software that the institution provides access to. Students are assessed on both their abilities to design solutions to the problem in the task and their use of GenAI systems.

Level 5: AI Exploration

Level 5 represents a forward-looking vision of assessment design, centred not just on incorporating GenAI tools but on using them to reimagine what academic tasks can entail. Unlike earlier levels, which focus on controlling, channelling, or strategically deploying GenAI, this level encourages the co-creation and development of entirely new methods, media, and outputs that challenge conventional disciplinary boundaries. It builds upon the foundational understanding of

GenAI use outlined in the lower levels of the framework, particularly the emphasis on authenticity, critical engagement, and strategic tool selection found in Level 4. However, it extends these principles to future-oriented scenarios that may not yet be fully feasible with current technology.

At this stage, students are not only using GenAI tools to enhance their work; they are expected to conceptualise and implement new GenAI applications, artefacts, or tools that go beyond existing templates and straightforward multimodal generation. This might mean creating world models or integrating advanced synthetic media production into an assessment task, which transcends the more direct and instrumental uses of GenAI at Level 4. While Level 4 acknowledges the arrival of advanced multimodal systems, Level 5 imagines a space where these tools become the raw materials from which students and educators collaboratively sculpt novel approaches to enquiry, whether it involves simulating complex social phenomena, generating bespoke datasets that do not currently exist, or building adaptive systems that respond dynamically to user input and changing conditions. The role of the educator also changes, moving from an authority that defines the parameters of the task to a collaborator who shares ownership of the learning process. This co-creation aspect, suggested in earlier treatments of GenAI integration but not fully realised, aligns with the need for flexibility, adaptability, and openness to new forms of assessment in a world where digital and human cognition are intertwined.

The implementation of Level 5 assessments requires careful consideration of equity issues, as highlighted in our Level 4 discussion. However, at this level, the focus shifts from providing equal access to tools to ensuring equal opportunities for innovation and experimentation in the classroom or assessment. This might involve the institutional provision of advanced GenAI capabilities through self-hosted GenAI models or API access or the development of custom tools specific to disciplines or research areas. As Bearman et al. (2024) suggest, evaluative judgement becomes crucial here: students must not only use GenAI effectively but also critically assess its contributions to their innovative work.

This level may be most relevant to advanced undergraduate projects, postgraduate coursework, doctoral research, or cutting-edge independent endeavours at the secondary level, such as the extended projects of the International Baccalaureate (IB). This challenges the assumption that GenAI must be contained or limited to specific tasks. Instead, it posits that the measure of learning lies in how well students can harness, adapt, and extend GenAI capabilities for new intellectual, aesthetic, or practical ends. Such assessments might respond to the evolving state of AI, acknowledging that as brain-computer interfaces (BCIs) and other technologies develop (Eaton, 2023), traditional assessment controls or even conventional understanding of ‘authentic’ tasks may become obsolete. Thus, Level 5 is deliberately futuristic, anticipating that new toolkits, disciplinary challenges, and ways of knowing will emerge and that assessment must evolve to ensure that students are equipped to navigate this uncertain terrain.

JUTLP
Journal of University Teaching & Learning Practice

Example of Level 5 Assessment

In a creative arts specialist institution, educators and students are exploring the limits of multimodal generative AI, including image recognition and segmentation, to understand how AI can be used to support contemporary dance education. Some students work on a system that can provide real-time feedback on performers' movements, posture, and position, whereas others explore the use of machine learning to create procedurally generated lighting and effects which map to a dancer's movement. The educator invites external speakers who work in both dance and GenAI to help the students refine their ideas and engage in the process of co-creation and assessment co-design alongside students and educators.

Implementation Considerations

Any implementations of the AIAS should result in substantive changes to assessment tasks, not just discursive ones (Corbin et al., 2025; Perkins et al., 2025). The AIAS should drive structural changes to assessment, not security theatre. Simply labelling an unchanged task 'Level 3' or 'Level 4' is a discursive change that students can ignore, resulting in a loss of assessment validity. Instead, educators should start with the learning outcomes, choose the AIAS level that fits their class, unit, or program objectives, and then redesign the specific mechanics of the task: the role GenAI may play, required evidence of process, staged checkpoints, and marking criteria.
For example, a Level 3 task might ask students to use GenAI for ideation and drafting and submit a short prompt log or reflection. A former essay task might shift to a Level 1 assessment in the form of an in-class performance or oral explanation of the method, aligned to the same outcomes. The scale is a design tool for educators, not a script for students; therefore, student-facing instructions can be adjusted or simplified, potentially without naming the scale level itself. We recommend that educators looking to implement the AIAS to engage in real changes to assessment practices use the tools available on aiassessmentscale.com to support an implementation that goes beyond a basic label of a framework.

Conclusion

The revised version of the AIAS represents a significant evolution in our understanding of how to integrate GenAI into educational assessments. Key changes include moving away from the traffic light colour scheme to avoid implying hierarchical values, recognising text generation capabilities at earlier levels of the scale considering evidence regarding both GenAI text detection limitations and potential learning benefits, and adjusting Level 5 to acknowledge the rapid advancement of these technologies. The framework now better reflects the complex reality of assessment in an AI-enabled world while maintaining its core purpose of supporting transparent dialogue between educators and students and the effective redesign of assessments.

However, we must acknowledge that frameworks for GenAI integration in education cannot ignore the broader ethical and environmental impacts of these technologies. The planetary costs of training and running LLMs are substantial (Driessens & Pischetola, 2024), and concerns regarding human rights and values in AI development remain urgent (Eaton, 2024; Rudolph et al., 2024). The challenge of maintaining transparency in AI use is highlighted by research showing students' reluctance to declare AI use, even when permitted (Gonsalves, 2024), suggesting that deeper cultural shifts may be needed in how we view and discuss AI assistance in academic work. This transparency extends to concerns relating to GenAI and research integrity (Bjelobaba et al., 2025) for educators who engage in research activities. Additionally, understanding and addressing the underlying biases in GenAI models is critical as these technologies become increasingly embedded in educational practices. Therefore, it is vital that educational providers explicitly teach students about the ethical issues related to the use of GenAI tools so that they may develop the skills needed to use this technology in a way that demonstrates their CAIL.

While early evidence suggests that the AIAS can help improve student outcomes, reduce academic misconduct, and support educators in effective assessment design (Furze et al., 2024), further empirical validation is crucial. Future research should examine both the framework's effectiveness across different educational contexts and the broader impact of GenAI integration in education (Koh & Doroudi, 2023). This research should include qualitative investigations of students and educators' experiences with the framework, as well as quantitative measures of its impact on learning outcomes and assessment validity. The addition of Level 5 represents our attempt to future-proof the framework, acknowledging Dawson and Bearman's (2020) observation that forward-looking assessment requires some prediction of future developments. Although we cannot perfectly anticipate technological advancements, Level 5 provides flexibility for incorporating emerging AI capabilities while maintaining a focus on human creativity and critical engagement.

We emphasise that the AIAS is not prescriptive; institutions should adapt it to their specific contexts and requirements. However, the fact that GenAI is now an established part of the educational landscape cannot be ignored. Rather than attempting to restrict or control GenAI use through increasingly ineffective technical measures, educators must engage in open dialogue with students about appropriate GenAI integration in assessments and stage 'No AI' assessments at appropriate intervals throughout the program of study to increase overall assessment validity. The AIAS provides a structured way for these conversations to occur while acknowledging both the opportunities and challenges of GenAI in education. Ultimately, the success of any framework for GenAI integration will depend not only on its technical robustness or theoretical grounding but also on how effectively it supports meaningful learning and assessment in an increasingly AI-enabled world. As we continue to gather evidence and refine our approaches, maintaining a focus on assessment validity, the impact of student learning, and ethical considerations is crucial. The AIAS represents one step toward this goal, but much work remains to be done to understand how

best to prepare students for a future in which human and artificial intelligence are increasingly intertwined.

Acknowledgements

The authors disclose that they have no actual or perceived conflicts of interest. The authors disclose that they have not received any funding for this manuscript beyond resourcing for academic time at their respective universities. The authors have used GenAI tools for ideation and in some passages of draft text creation which was then heavily revised, along with editing and revision during the production of the manuscript. The tools used were ChatGPT (o3-mini-high) and Claude 3.5 Sonnet, which were chosen for their ability to provide sophisticated feedback on textual outputs. These tools were selected and used supportively and not to replace core author responsibilities and activities. The authors reviewed, edited, and take responsibility for all outputs of the tools used. The authors confirm that they have met the ethical standards expected as per Purvis & Crawford (2024). We are grateful for the support and guidance provided by Jason MacVaugh during the redevelopment of the revised version of the AIAS. We thank all educators worldwide who have provided comments and critiques on all versions of the scales, as well as those who have spent their time providing translations. The authors list the following CRediT contributions: **Perkins**: Conceptualisation, Investigation, Writing - Original Draft, Writing - Review & Editing, Project administration; **Roe:** Conceptualisation, Investigation, Writing - Original Draft, Writing - Review & Editing; **Furze**: Conceptualisation, Investigation, Writing - Original Draft, Writing - Review & Editing, Visualization.

JUTLP
Journal of University Teaching & Learning Practice

References

Bearman, M., Tai, J., Dawson, P., Boud, D., & Ajjawi, R. (2024). Developing evaluative judgement for a time of generative artificial intelligence. *Assessment & Evaluation in Higher Education*, *0*(0), 1–13. https://doi.org/10.1080/02602938.2024.2335321

Bewersdorff, A., Zhai, X., Roberts, J., & Nerdel, C. (2023). Myths, mis- and preconceptions of artificial intelligence: A review of the literature. *Computers and Education: Artificial Intelligence*, *4*, 100143. https://doi.org/10.1016/j.caeai.2023.100143

Bjelobaba, S., Waddington, L., Perkins, M., Foltýnek, T., Bhattacharyya, S., & Weber-Wulff, D. (2025). Maintaining research integrity in the age of GenAI: An analysis of ethical challenges and recommendations to researchers. *International Journal for Educational Integrity*, *21*(1), Article 1. https://doi.org/10.1007/s40979-025-00191-w

Corbin, T., Dawson, P., & Liu, D. (2025). Talk is cheap: Why structural assessment changes are needed for a time of GenAI. *Assessment & Evaluation in Higher Education*, *0*(0), 1–11. https://doi.org/10.1080/02602938.2025.2503964

Cotterell, A. (2024, September 3). A defence of the traffic light metaphor. *Adrian Cotterell*. https://adriancotterell.com/2024/09/04/a-defence-of-the-traffic-light-metaphor/

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, *0*(0), 1–12. https://doi.org/10.1080/14703297.2023.2190148

Crawford, J., Cowling, M., & Allen, K.-A. (2023). Leadership is needed for ethical ChatGPT: Character, assessment, and learning using artificial intelligence (AI). *Journal of University Teaching and Learning Practice*, *20*(3). https://doi.org/10.53761/1.20.3.02

Dawson, P., & Bearman, M. (2020). Concluding Comments: Reimagining University Assessment in a Digital World. In M. Bearman, P. Dawson, R. Ajjawi, J. Tai, & D. Boud (Eds), *Re-imagining University Assessment in a Digital World* (pp. 291–296). Springer International Publishing. https://doi.org/10.1007/978-3-030-41956-1_20

Dawson, P., Bearman, M., Dollinger, M., & Boud, D. (2024). Validity matters more than cheating. *Assessment & Evaluation in Higher Education*, *49*(7), 1005–1016. https://doi.org/10.1080/02602938.2024.2386662

Dhillon, P. S., Molaei, S., Li, J., Golub, M., Zheng, S., & Robert, L. P. (2024). Shaping Human-AI Collaboration: Varied Scaffolding Levels in Co-writing with Language Models. *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 1–18. https://doi.org/10.1145/3613904.3642134

Driessens, O., & Pischetola, M. (2024). Danish university policies on generative AI: Problems, assumptions and sustainability blind spots. *MedieKultur: Journal of Media and Communication Research*, *40*(76), Article 76. https://doi.org/10.7146/mk.v40i76.143595

Eaton, S. E. (2023). Postplagiarism: Transdisciplinary ethics and integrity in the age of artificial intelligence and neurotechnology. *International Journal for Educational Integrity*, *19*(1), 23. https://doi.org/10.1007/s40979-023-00144-1

Eaton, S. E. (2024). Future-proofing integrity in the age of artificial intelligence and neurotechnology: Prioritizing human rights, dignity, and equity. *International Journal for Educational Integrity*, *20*(1), Article 1. https://doi.org/10.1007/s40979-024-00175-2

Elkhatat, A. M., Elsaid, K., & Almeer, S. (2023). Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text. *International Journal for Educational Integrity*, *19*(1), 17. https://doi.org/10.1007/s40979-023-00140-5

European Commission. (2018). *A definition of AI: Main capabilities and scientific disciplines*. European Commission High-Level Expert Group on Artificial Intelligence. https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf

Farazouli, A., Cerratto-Pargman, T., Bolander-Laksov, K., & McGrath, C. (2023). Hello GPT! Goodbye home examination? An exploratory study of AI chatbots impact on university teachers' assessment practices. *Assessment & Evaluation in Higher Education*, *0*(0), 1–13. https://doi.org/10.1080/02602938.2023.2241676

Furze, L. (2024, December 9). *AI Assessment Scale (AIAS) Translations from Around the World*. Leon Furze. https://leonfurze.com/2024/12/09/ai-assessment-scale-aias-translations-from-around-the-world/

Furze, L., Perkins, M., Roe, J., & MacVaugh, J. (2024). The AI Assessment Scale (AIAS) in action: A pilot implementation of GenAI-supported assessment. *Australasian Journal of Educational Technology*. https://doi.org/10.14742/ajet.9434

Gonsalves, C. (2024). Addressing student non-compliance in AI use declarations: Implications for academic integrity and assessment in higher education. *Assessment & Evaluation in Higher Education*, *0*(0), 1–15. https://doi.org/10.1080/02602938.2024.2415654

Gorichanaz, T. (2023). Accused: How students respond to allegations of using ChatGPT on assessments. *Learning: Research and Practice*, *9*(2), 183–196. https://doi.org/10.1080/23735082.2023.2254787

Head, A., & Willis, S. (2024). Assessing law students in a GenAI world to create knowledgeable future lawyers. *International Journal of the Legal Profession*, *0*(0), 1–18. https://doi.org/10.1080/09695958.2024.2379785

Hsiao, Y.-P., Klijn, N., & Chiu, M.-S. (2023). Developing a framework to re-design writing assignment assessment for the era of Large Language Models. *Learning: Research and Practice*, *9*(2), 148–158. https://doi.org/10.1080/23735082.2023.2257234

Kılınç, S. (2024). Comprehensive AI assessment framework: Enhancing educational evaluation with ethical AI integration. *Journal of Educational Technology and Online Learning*, *7*(4-ICETOL 2024 Special Issue), 521–540. https://doi.org/10.31681/jetol.1492695

Koh, E., & Doroudi, S. (2023). Learning, teaching, and assessment with generative artificial intelligence: Towards a plateau of productivity. *Learning: Research and Practice*, *9*(2), 109–116. https://doi.org/10.1080/23735082.2023.2264086

Liu, D., & Bridgeman, A. (2023, July 12). What to do about assessments if we can't out-design or out-run AI? *What to Do about Assessments If We Can't out-Design or out-Run AI?* https://doi.org/10.53761/1.20.3.02

Lodge, J. M., Howard, S., Bearman, M., & Dawson, P. (2023). *Assessment reform for the age of artificial intelligence*. Tertiary Education Quality and Standards Agency. https://www.teqsa.gov.au/guides-resources/resources/corporate-publications/assessment-reform-age-artificial-intelligence

Lorenz, P., Perset, K., & Berryhill, J. (2023). *Initial policy considerations for generative artificial intelligence* (OECD Artificial Intelligence Papers No. 1; OECD Artificial Intelligence Papers, Vol. 1). OECD Publishing. https://doi.org/10.1787/fae2d1e6-en

Luccioni, S., Jernite, Y., & Strubell, E. (2024). Power Hungry Processing: Watts Driving the Cost of AI Deployment? *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, 85–99. https://doi.org/10.1145/3630106.3658542

Luo (Jess), J. (2024a). A critical review of GenAI policies in higher education assessment: A call to reconsider the "originality" of students' work. *Assessment & Evaluation in Higher Education*, *0*(0), 1–14. https://doi.org/10.1080/02602938.2024.2309963

Luo (Jess), J. (2024b). How does GenAI affect trust in teacher-student relationships? Insights from students' assessment experiences. *Teaching in Higher Education*, *0*(0), 1–16. https://doi.org/10.1080/13562517.2024.2341005

Mao, J., Chen, B., & Liu, J. C. (2024). Generative Artificial Intelligence in Education and Its Implications for Assessment. *TechTrends*, *68*(1), 58–66. https://doi.org/10.1007/s11528-023-00911-4

Newton, P. M., Summers, C. J., Zaheer, U., Xiromeriti, M., Stokes, J. R., Bhangu, J. S., Roome, E. G., Roberts-Phillips, A., Mazaheri-Asadi, D., Jones, C. D., Hughes, S., Gilbert, D., Jones, E., Essex, K., Ellis, E. C., Davey, R., Cox, A. A., & Bassett, J. A. (2024). *Can ChatGPT-4o really pass medical science exams? A pragmatic analysis using novel questions* (p. 2024.06.29.24309595). medRxiv. https://doi.org/10.1101/2024.06.29.24309595

Nguyen, A., Hong, Y., Dang, B., & Huang, X. (2024). Human-AI collaboration patterns in AI-assisted academic writing. *Studies in Higher Education*, *49*(5), 847–864. https://doi.org/10.1080/03075079.2024.2323593

OECD. (2024). *AI risks and incidents*. Artificial Intelligence Promises Tremendous Benefits but Also Carries Real Risks. Some of These Risks Are Already Materialising into Harms to People and Societies: Bias and Discrimination, Polarisation of Opinions, Privacy Infringements, and Security and Safety Issues. Trustworthy AI Calls for Governments Worldwide to Develop Interoperable Risk-Based Approaches to AI Governance and a Rigorous Understanding of AI Incidents and Hazards. https://www.oecd.org/en/topics/ai-risks-and-incidents.html

OpenAI. (2024). *Introducing OpenAI o1 | OpenAI*. https://openai.com/index/introducing-openai-o1-preview/

Pearce, J., & Chiavaroli, N. (2023). Rethinking assessment in response to generative artificial intelligence. *Medical Education*, *57*(10), 889–891. https://doi.org/10.1111/medu.15092

Perkins, M. (2023). Academic Integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching & Learning Practice*, *20*(2). https://doi.org/10.53761/1.20.02.07

Perkins, M., Furze, L., Roe, J., & 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*(06), Article 06. https://doi.org/10.53761/q3azde36

Perkins, M., Roe, J., & Furze, L. (2025). How (not) to use the AI Assessment Scale. *Journal of Applied Learning and Teaching*, *8*(2). https://doi.org/10.37074/jalt.2025.8.2.15

Perkins, M., Roe, J., Vu, B. H., Postma, D., Hickerson, D., McGaughran, J., & Khuat, H. Q. (2024). Simple techniques to bypass GenAI text detectors: Implications for inclusive education. *International Journal of Educational Technology in Higher Education*, *21*(1), 53. https://doi.org/10.1186/s41239-024-00487-w

Pratschke, B. M. (2024). Assessing Learning. In B. M. Pratschke (Ed.), *Generative AI and Education: Digital Pedagogies, Teaching Innovation and Learning Design* (pp. 91–108). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-67991-9_6

Rasul, T., Nair, S., Kalendra, D., Balaji, M. S., Santini, F. de O., Ladeira, W. J., Rather, R. A., Yasin, N., Rodriguez, R. V., Kokkalis, P., Murad, M. W., & Hossain, M. U. (2024). Enhancing academic integrity among students in GenAI Era:A holistic framework. *The International Journal of Management Education*, *22*(3), 101041. https://doi.org/10.1016/j.ijme.2024.101041

Roe, J., Furze, L., & Perkins, M. (2025). Reflecting Reality, Amplifying Bias? Using Metaphors to Teach Critical AI Literacy. *Journal of Interactive Media in Education*, 1–15. https://doi.org/10.5334/jime.961

Roe, J., Perkins, M., & Ruelle, D. (2024). Is GenAI the Future of Feedback? Understanding Student and Staff Perspectives on AI in Assessment. *Intelligent Technologies in Education*. https://doi.org/10.70770/rzzz6y35

Roe, J., Perkins, M., & Tregubova, Y. (2024). *The EAP-AIAS: Adapting the AI Assessment Scale for English for Academic Purposes* (No. arXiv:2408.01075). arXiv. https://doi.org/10.48550/arXiv.2408.01075

Rudolph, J., Ismail, F. M. M., & Popenici, S. (2024). Higher Education's Generative Artificial Intelligence Paradox: The Meaning of Chatbot Mania. *Journal of University Teaching and Learning Practice*, *21*(06), Article 06. https://doi.org/10.53761/54fs5e77

Sadasivan, V. S., Kumar, A., Balasubramanian, S., Wang, W., & Feizi, S. (2023). *Can AI-Generated Text be Reliably Detected?* (No. arXiv:2303.11156). arXiv. https://doi.org/10.48550/arXiv.2303.11156

Shahriar, S., Lund, B. D., Mannuru, N. R., Arshad, M. A., Hayawi, K., Bevara, R. V. K., Mannuru, A., & Batool, L. (2024). Putting GPT-4o to the Sword: A Comprehensive Evaluation of

Language, Vision, Speech, and Multimodal Proficiency. *Applied Sciences*, *14*(17), Article 17. https://doi.org/10.3390/app14177782

Smolansky, A., Cram, A., Raduescu, C., Zeivots, S., Huber, E., & Kizilcec, R. F. (2023). Educator and Student Perspectives on the Impact of Generative AI on Assessments in Higher Education. *Proceedings of the Tenth ACM Conference on Learning @ Scale*, 378–382. https://doi.org/10.1145/3573051.3596191

Steel, A. (2024). *2 lanes or 6 lanes? It depends on what you are driving: Use of AI in Assessment*. https://www.education.unsw.edu.au/news-events/news/two-six-lanes-ai-assessment

Studiosity. (2024). *The Prof Tracey Bretag Prize for Academic Integrity*. https://www.studiosity.com/traceybretagprize

Thanh, B. N., Vo, D. T. H., Nhat, M. N., Pham, T. T. T., Trung, H. T., & 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), Article 5. https://doi.org/10.14742/ajet.8902

Thompson, K., Corrin, L., & Lodge, J. M. (2023). AI in tertiary education: Progress on research and practice. *Australasian Journal of Educational Technology*, *39*(5), Article 5. https://doi.org/10.14742/ajet.9251

Vygotsky, L. S. (1978). *Mind in Society: Development of Higher Psychological Processes*. Harvard University Press. https://doi.org/10.2307/j.ctvjf9vz4

Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., Šigut, P., & Waddington, L. (2023). Testing of detection tools for AI-generated text. *International Journal for Educational Integrity*, *19*(1), Article 1. https://doi.org/10.1007/s40979-023-00146-z

Wiboolyasarin, W., Wiboolyasarin, K., Suwanwihok, K., Jinowat, N., & Muenjanchoey, R. (2024). Synergizing collaborative writing and AI feedback: An investigation into enhancing L2 writing proficiency in wiki-based environments. *Computers and Education: Artificial Intelligence*, *6*, 100228. https://doi.org/10.1016/j.caeai.2024.100228

Xia, Q., Weng, X., Ouyang, F., Lin, T. J., & Chiu, T. K. F. (2024). A scoping review on how generative artificial intelligence transforms assessment in higher education. *International Journal of Educational Technology in Higher Education*, *21*(1), 40. https://doi.org/10.1186/s41239-024-00468-z