Designing and Evaluating Template-Based AI Use Scales for University Assessment: A Contextual Lane Approach

Stella Peng, Winn Wing-Yiu Chow, Mike Zhuang, Mingxin Zhang, Arzoo Atiq, Muqing Guo, and Tingru Cui
School of Computing and Information Systems, University of Melbourne
stella.peng@unimelb.edu.au

ABSTRACT

CONTEXT

The rapid rise of generative AI (GenAI) tools presents major challenges to academic integrity and effective assessment in higher education. As proficiency with GenAI tools increasingly becomes a vital skill for the future workforce, many higher education institutions now allow their conditional use in assessments. However, instructors often struggle to determine appropriate boundaries for AI use and to communicate expectations. This ambiguity can leave students uncertain about the appropriate AI use in their assessments, increasing the risk of unintentional misuse. While previous studies have proposed AI use scales to address this concern, there is still a lack of appropriate scales tailored for different assessment formats and empirical research on their effectiveness.

PURPOSE

To address the research gaps above, this study aims to develop AI use scales for different assessment formats and evaluate their effectiveness in communicating AI usage expectations to students. The research questions are: (1) What are the key design principles of AI use scales for assessments? and (2) How effective are the scales in communicating the AI use requirements?

METHODOLOGY

Adopting a design-based research approach, this pilot study was conducted in a postgraduate Information Systems subject with diverse assessments (project reports, oral presentations, video creation). Customizable, template-based AI use scales were developed and iteratively refined based on established pedagogical frameworks. A mixed-methods evaluation was employed, gathering data from instructor reflections, a post-semester student survey, and semi-structured interviews with students.

ACTUAL OR ANTICIPATED OUTCOMES

The study found that a “one-size-fits-all” approach to AI use scales is likely ineffective. Through an iterative development process, the research derived nine key design principles for creating adaptable AI use scales tailored to various assessment formats. Instructors found the templates easy to adapt, which enabled them to more effectively communicate their expectations. Preliminary findings from student interviews indicated that the scales provided clearer guidelines, reduced uncertainty, and enhanced their AI literacy.

CONCLUSIONS

This study provides evidence that a flexible, “contextual lane approach” is more effective for communicating AI requirements than generic scales. By offering customizable, template-based scales, instructors can better align AI use with specific learning outcomes and assessment types. The findings lay the groundwork for a more nuanced and context-sensitive integration of GenAI in higher education assessment.

KEYWORDS: GenAI, AI use scale, assessment, design-based research, template-based

Introduction

According to a recent survey conducted by the Digital Education Council (2025), generative artificial intelligence (GenAI) tools have become increasingly prominent in higher education, with 86% of educators worldwide expecting AI to play a significant role in teaching and learning. The rapid evolution of GenAI is transforming students' learning environments and influencing pedagogical practices. Tools such as ChatGPT, Claude, and GitHub Copilot are now widely accessible, enabling students to complete a wide range of assessment tasks, such as writing, programming, problem-solving, and creative work, with minimal cognitive effort or direct engagement. While higher education institutions acknowledge the opportunities GenAI offers to innovate and enhance personalized learning and academic development, they also recognize the significant challenges it poses, particularly concerning academic integrity and assessment practices (Balalle & Pannilage, 2025). In response, many institutions are revising academic integrity policies and guidelines, placing greater emphasis on clear principles and expectations regarding the ethical use of AI in assessments, as well as on secure and authentic, skill-based evaluation methods (Nikolic et al., 2024).

Given that proficient use of GenAI to supplement human tasks is expected to become a vital skill for the future workforce, many university instructors are beginning to integrate GenAI into their teaching practices and allow students to use it in assessments. However, unconditional use of GenAI in assessments is problematic, as it can undermine the development of essential skills and compromise the validity of assessment outcomes (Curtis, 2025). Consequently, a key challenge lies in defining appropriate boundaries for AI use and clearly communicating the expectations to students to ensure that intended learning outcomes are demonstrated. The appropriate extent of AI tool use often varies depending on the discipline, level of study (e.g., undergraduate and postgraduate), nature of the subject, and the type and purpose of the assessment. However, current institutional or faculty-level guidelines provide high-level AI principles and frameworks that lack the specificity needed to support discipline- and assignment-specific requirements meaningfully (Perkins et al., 2024). As a result, instructors often struggle to adapt the guidelines to the specific assessment types within their subjects, failing to establish clear boundaries for AI use and to communicate these expectations to students. These issues can lead to miscommunication, inconsistencies across subjects, and increased student uncertainty and anxiety about the responsible use of GenAI.

In addition, without a clear, pedagogically informed approach to AI use that supports instructors in subject assessment, there is a risk of either over-restricting potentially valuable learning tools or inadvertently encouraging unintentional misuse. While previous studies have suggested that AI use scales (i.e., frameworks that define acceptable levels of AI use) can effectively communicate AI use expectations in assessments, the scales are often developed at the institutional or faculty level and may not adequately address the diverse needs of specific assessment types or disciplinary and subject contexts. This contrasts with grading rubrics, which are adaptable to clarify expectations for subject assessments. Inspired by the flexibility and clarity offered by rubrics, this study addresses these gaps by designing and developing customizable template-based AI use scales tailored to specific assessments, guided by the following research questions (RQs):

RQ1: What are the key design principles for developing AI use scales that effectively communicate expectations for responsible AI use across different subject assessments?

RQ2: How effective are these AI use scales in helping students understand AI use expectations and in supporting academic integrity?

In this study, through an iterative process of development and evaluation conducted in a postgraduate Information Systems capstone subject at an Australian university, key design principles were developed and refined, and initial insights were gained into the scale's effectiveness. The findings contribute to existing knowledge on GenAI integration in higher education by identifying key design principles for assessment-specific AI use scales. The proposed customizable scale offers a flexible framework that bridges the gap between high-level institutional policies and the specific needs of individual disciplines.

Literature Review

GenAI in Higher Education

Generative AI tools bring new technological capabilities that can potentially match human intelligence. In higher education, these tools can generate natural language text, programming code, images, and other forms of content that support the completion of many academic tasks (Casal & Kessler, 2023). They are reshaping how students access information, engage in learning, complete assessments, and receive feedback (Lee & Moore, 2024). However, since AI-generated content is designed to match human quality and can be difficult even for experts to distinguish, these tools blur the line between original student work and AI-generated work, posing critical challenges for authenticating student submissions and maintaining academic integrity. As students gain widespread access to increasingly powerful GenAI tools, higher education institutions face growing pressure to rethink student skill development, curriculum design, and assessment practices, with particular concern for preserving academic integrity and the credibility of university education (Balalle & Pannilage, 2025).

The rise of GenAI has introduced new complexities to the longstanding issue of academic integrity. Before the AI era, tools like Turnitin and iThenticate aided in detecting academic misconduct by comparing student submissions against extensive databases of existing student work, academic publications, and public internet content. However, the emergence of GenAI has rendered the traditional plagiarism detection tools less reliable (Perkins & Roe, 2024). In response, higher education institutions adopt a twofold strategy: detection and prevention.

Several studies have explored the detection of AI misuse. For example, Khalil & Er (2023) demonstrated that ChatGPT-generated content can easily bypass plagiarism detection tools such as Turnitin and iThenticate, supporting Casal and Kessler's (2023) finding that even linguists struggle to distinguish AI-generated from human-written texts. Although new AI-detection tools such as GPTZero, Crossplag, and Copyleaks have emerged, they remain unreliable and inconsistent (Chaka, 2024). Moreover, some studies have raised concerns about detection tools' biases against non-native English speakers (Perkins & Roe, 2024). As a result, detection remains a problematic strategy, prompting a growing focus on the prevention of AI misuse.

A growing body of literature suggests that proactive prevention is more effective than relying solely on detection to address GenAI misuse. Preventive approaches include updating academic integrity policies to clearly define AI-related misconduct (Perkins & Roe, 2024); redesigning assessments to emphasize process over product and promote transparency in AI use (Corbin et al., 2025); offering educational training to enhance student AI literacy (Walter, 2024); and developing AI use guidelines and scales to clearly communicate expectations (Perkins et al., 2024). Across these initiatives, the combination of clearly articulated expectations and an educative, values-based approach has emerged as critical to promoting responsible AI use. While guidelines and policies established at institutional and faculty levels provide a general framework for responsible GenAI use in assessments, how AI use expectations can be clearly and effectively specified and communicated within individual assessments at the subject level remains underexplored.

Common Frameworks for AI Use in Assessment

As higher education institutions navigate the complexities of GenAI, two prominent high-level frameworks have emerged as commonly referenced guides for AI use in assessment: the two-lane approach and the AI Assessment Scale (AIAS).

The "two-lane" approach (also known as "all-or-none") was proposed by Liu and Bridgeman (2024; 2023) at the University of Sydney, which aims to achieve a "balance between assurance and human-AI collaboration" to prepare students for future AI-integrated work environments. In "Lane 1," GenAI use is strictly prohibited for "secured" assessments, such as invigilated exams, whereas "Lane 2" permits unrestricted GenAI use in "unsecured" or "open" assessments. Although this approach shows a clear delineation of expectations, its binary nature and oversimplification lack the nuance needed to accommodate diverse assessment types and learning objectives (Curtis, 2025). Therefore, some argue that a "multiple lane" approach may be more appropriate, allowing for varying levels of conditional AI use.

The AI Assessment Scale (AIAS) embraces a "multiple lane" approach and defines five different levels of AI use in assessment, ranging from no AI to full AI (Perkins et al., 2024). This spectrum allows greater flexibility and nuance in integrating AI into assessments. Instructors can more effectively align the permitted use of AI with their subject's intended learning outcomes, while also promoting transparency and academic integrity. However, despite its increased granularity, the scale remains somewhat limited in scope, given the diversity of disciplines, subject nature, and assessment types. For example, the AI-assisted editing level is not relevant for oral, practical, or performance-based assessments, where written outputs are minimal or absent. In essence, a "one-size-fits-all" AI use scale often falls short of accommodating the pedagogical needs and assessment realities of diverse educational contexts. Effective implementation, therefore, requires subject-level adaptation, with customizable applications tailored to specific assessments and supported by discipline-specific guidance. This underscores the need for further development of a more flexible and context-sensitive AI use scale, which may be termed a "contextual lane approach," that enables meaningful integration of GenAI across varied academic settings.

Methodology

This study aims to gain actionable knowledge on guiding students' responsive AI use in HE. Therefore, we adopted an educational design-based research (DBR) approach (Bakker, 2018) to develop and evaluate a "contextual lane approach" AI use scale that can be customized for various assessment types within a subject. The approach is well-suited to the study's goals, as it follows an iterative cycle of designing an educational intervention (i.e., AI use scales), implementing it in authentic classroom contexts, and systematically evaluating its effectiveness. Through this process, both the practical tool and its underlying design principles are refined based on real-world application and feedback. A postgraduate Information Systems capstone subject at an Australian university, comprising 69 students and two instructors, was selected as the testbed for implementation. The subject's assessments included project reports, oral presentations, and video creation, thereby providing a diverse context for empirical evaluation. The study consisted of two parts: the development of customizable template-based AI use scales and empirical evaluation. The study was conducted under research ethics protocol 29248, secured by the Faculty's XXX (anonymised for blind review).

Part 1: Customizable Template-based AI Use Scale Development. The first part addressed RQ1. Building on the AIAS framework (Perkins et al., 2024), we developed general design principles and customizable AI use scale templates through an iterative and collaborative development process. The development was grounded in Bloom's Taxonomy to align AI use with cognitive skill levels, the Cynefin framework to assess task complexity, and the TPACK framework to integrate technology with pedagogy and content. Initial templates were reviewed by the two instructors of the chosen subject. Their feedback on clarity, adaptability, and alignment with learning outcomes informed refinements. Using the templates, an AI use scale tailored to the subject was developed. After introducing the scale to students, the instructors collected preliminary student feedback and discussed it with the research team to further refine the design. The final scales and principles were adjusted to enhance practical utility and ensure flexibility across different assessment types to promote clear communication of responsible AI use expectations.

Part 2: AI Use Scale Empirical Evaluation (Pilot). The second part focused on RQ2. A mixed-methods approach was adopted. Data collection involved three main sources. First, reflections from the instructors (from the authors) were gathered to evaluate the templates' adaptability and communication effectiveness. Second, a post-semester survey, administered as part of a broader study, gathered student perceptions of the AI use scale, which was introduced to them as the *GenAI Checklist*, along with basic demographic data (n = 52; Female: Male = 63%:37%; all were international students). Student perceptions were measured using a 5-level Likert scale: (1) "The 'GenAI Checklist' in this subject clearly communicated the requirements, including declarations, for using AI tools in assessment tasks.” (2) “The ‘GenAI Checklist’ clearly communicated the different levels of permitted AI use across different assessment tasks.” and (3) “The ‘GenAI Checklist’ made me confident in using AI appropriately in assessment tasks.” Third, four semi-structured interviews with purposively sampled students provided qualitative insights into student experiences and perspectives.

Findings

A customizable, template-based AI use scale is the practical outcome of the design-based research process aimed at making AI use in assessments transparent, fair, and pedagogically sound. An example template designed for report-based assessments in Information Systems subjects is presented in Figure 1.

AI Assessment Use Levels	Instructions to Students	Examples	AI Acknowledgement
AI-Assisted Proofing	Basic spell-checking and grammar-checking with no AI-based rewording or AI content generation; translation of individual words only.	YES - The student uses Grammarly (Free version) for spelling and grammar checks but avoids AI-powered rewording or clarity suggestions. YES - The student uses Microsoft Word’s built-in spell-check and grammar-check tools to correct errors, but avoids any features such as “Rewrite Suggestions” that would alter the sentence structure or content. NO - Instead of just checking for spelling and grammar, the student accepts Grammarly’s AI-generated rewording suggestions that restructure sentences significantly.	Students MUST acknowledge the use of AI by adding a declaration at the end of their submission.
AI-Assisted Idea Generation	Brainstorming ideas; searching for topics or information relevant to the task.	YES - The student asks ChatGPT, “What are some key topics I should consider when discussing solar energy incentives?” and ChatGPT lists relevant subtopics. NO - The student asks ChatGPT, “Can you write the introduction to a report on renewable energy policies?” NO - The student uploads a detailed business problem statement (e.g., sales decline, operational inefficiencies, or market expansion challenge) and then asks, “Can you provide a solution to this business problem?”	
AI-Assisted Drafting of Task Elements with Human Evaluation	Applying theories, frameworks, or techniques to analyse some specific elements of a task; content analysis (summarizing text or visualizing and interpreting data) for only certain parts of a task; using AI as a tool to assist with specific components of a complex task, but not the entire task; the output by AI must be evaluated by students.	YES - Student asks ChatGPT: “Can you help me analyse the market position of Company X using Porter’s Five Forces framework? Here’s some information about the company and its competitors,” then evaluates the provided analysis, and combines other findings to make final recommendations by themselves. NO - Student uploads a complex business case study to ChatGPT and asks, “Can you analyse this case study and provide a detailed report with recommendations?”	Students MUST (a) cite and reference AI output that is either paraphrased or directly quoted in their submission, and (b) acknowledge the use of AI by adding a declaration at the end of their submission. University’s instructions on acknowledging AI tools and technologies: *(URL)*
AI for Learning (Not assessment-related)	Information browsing (excluding assessment-specific questions); explaining concepts with examples or simplified details; revision quizzes, flashcards, other memorization help; AI asking you open-ended questions (“Socratic learning”); emotional support (motivation, hype, habit tracking, etc.)	YES - A student asks ChatGPT, “Can you help me understand the concept of Porter's five forces with some examples?” NO - A student asks ChatGPT, “Can you help me answer this assignment question about Porter's five forces?” NO - A student uploads lecture slides (copyrighted materials) to ChatGPT and asks, “Can you help me summarise the Lecture content?”	

Use level descriptors

Choose the most relevant AI use level descriptors to design a scale that aligns with your assessment needs

Figure 1: An Example AI Use Scale Template

The scale is fundamentally designed as a template. An instructor would not use the template as-is; instead, they should choose the most relevant AI use level descriptors to design a scale that aligns with assessment needs. The customisation ensures that the AI use scale is pedagogically appropriate, assessment-specific, and meaningfully aligned with the cognitive and disciplinary goals of each task. Each level of use in the scale is accompanied by a dedicated “Instructions to Students” section that clearly defines the scope of permitted activity. This is reinforced by a practical “Examples” section, which offers concrete “YES” and “NO” scenarios to help students distinguish acceptable from unacceptable use. Moreover, a dedicated “AI Acknowledgement” section provides clear instructions that vary depending on the level of AI use, often with reference to institutional citation guidelines. This reinforces ethical academic practice and transparency. In addition, the scale includes an “AI for Learning” section, which offers guidance on how students can use AI responsibly as a learning tool, outside of assessments.

Design Principles

Through an iterative process of development and evaluation, key design principles were derived from the creation and implementation of a customizable, template-based AI use scale (Table 1).

Table 1: Key AI Use Scale Design Principles

Design Principles
DP1: Subject-Specific Tailoring. The scale should be customized for the specific level, discipline, and nature of a subject.
DP2: Assessment-Specific Adaptation. The scale should be further adapted for the unique requirements of each assessment type, such as reports, presentations, or creative works.
DP3: Explicit and Clear Guidance: The scale should provide unambiguous and specific instructions on how GenAI tools can and cannot be used, clearly delineating the boundaries of acceptable use.
DP4: Clear Acknowledgment Guidelines. The scale should include clear instructions on how students should acknowledge and cite their use of GenAI tools, promoting good academic practice.
DP5: Flexible Template-Based Structure. The scale should be designed as a customizable template that instructors find easy to adapt to various subjects and assessments.
DP6: Common Terminology and Language. The scale should employ a consistent and shared vocabulary to ensure uniform understanding and application by students.
DP7: Learning-Oriented Framework. The scale should serve as a pedagogical tool to help students achieve learning outcomes and develop essential AI-related skills for the future workforce.
DP8: AI Literacy Development Support. The scale should be part of a broader strategy that includes additional educational training resources.
DP9: Collaborative Development: The scale should be a living document, agreed upon through discussion and co-creation with instructors and students.

Institution-wide policies are often too broad to offer meaningful guidance at the subject level. ***DP1*** is supported by the Cynefin Framework which provides a lens to assess the complexity of subject matter. Foundational subjects, which focus on delivering simple, well-established knowledge, may be vulnerable to learning erosion if students rely on AI to bypass essential understanding. In contrast, advanced subjects often involve complex or ambiguous problems, where AI can be used strategically to offload routine or procedural tasks. Building on this, ***DP2*** highlights the importance of customization at the level of individual assessments within each subject. This principle is grounded in the concept of constructive alignment (Biggs et al., 2022), a widely accepted pedagogical framework emphasizes the alignment between assessment and learning outcomes.

Our preliminary findings highlighted that ambiguity was a key driver of student uncertainty and unintentional academic misconduct. ***DP3*** underscores the importance of AI use scales providing unambiguous, specific instructions, including concrete “Yes” and “No” examples. By offering clear guidance, instructors establish the “rules of the game,” which shape students’ attitudes toward responsible GenAI use and set a clear norm, as explained by the Theory of Planned Behaviour (TPB). This clarity helps reduce student anxiety and empowers them to make informed decisions, increasing intention to comply with the established boundaries. Meanwhile, it is also important to enable students to properly acknowledge their use of GenAI (***DP4****).* The preliminary findings noted that the scale prompted student questions about acknowledgment, demonstrating that clear guidelines are essential for helping students navigate the ethical complexities of using AI.

DP5 advocates for the development of adaptable AI use scale templates, which allow instructors to tailor the specific rules to individual assessments while maintaining a consistent format and shared categories of AI use. Preliminary findings suggest that this approach lowers the barrier to adoption, thereby promoting broader use across the institution. Consistency in structure is further reinforced by ***DP6****.* The wide adoption of AI use scales fosters a shared language that students and staff can easily understand. This consistency is critical for reducing ambiguity and minimising the risk of misinterpretation, as students do not need to learn new terms in each subject.

The goal of permitting AI use in assessments should not be limited to preventing academic misconduct; rather, it should guide students in developing the essential 21st-century competency of working productively and ethically with AI. ***DP7*** repositions the AI use scale from a purely restrictive mechanism to a pedagogical tool that actively supports and scaffolds student learning. Crucially, the scale can also articulate how AI-informed learning may be fostered, such as prompt engineering, critical evaluation of GenAI outputs, and meta-learning. In this way, the scale functions as a tool for assessment for learning, supporting students' growth as adaptive, AI-literate learners prepared for an evolving workforce.

DP8 argues that the AI use scale should be embedded within a broader educational strategy designed to foster comprehensive AI literacy among students. This involves not only clearly defining categories of AI use but also providing the support necessary for students to engage with these categories effectively and ethically.

Finally, ***DP9*** suggests that the AI use scale should function as a "living document" that is agreed upon through discussion and co-creation between instructors and students. This collaborative approach positions students as active stakeholders in shaping the integration of GenAI into learning and assessment. As frequent and often sophisticated users of GenAI tools, students possess valuable, first-hand knowledge of the technology's real-world capabilities and limitations. Leveraging this expertise through a participatory development process helps ensure that the scale remains both practical and responsive to the actual scenarios students encounter.

Empirical Evaluations

Reflections from the two instructors showed that the scale was a valuable tool for translating broad institutional policies into practice. They agreed that the template was easy to adapt for different assessment types (project reports, oral presentations, and video creation), which facilitated clearer communication of expectations to students. This clarity helped foster more open and productive conversations with students about AI use. Instructor 1, the more experienced instructor responsible for adapting the scale template, viewed the scale as essential for complying with institutional requirements while providing detailed GenAI use guidance. However, the instructors also identified areas for improvement. They observed that initial student engagement was low, suggesting that some students might not have been intrinsically motivated to consult it beyond basic compliance. Moreover, Instructor 2 noted that students occasionally discovered novel uses of AI that were not covered in the scale and were unknown to the instructors, which highlights the importance of collaborative development (DP9).

Survey data revealed that the AI use scale was successful in building student confidence and clarifying expectations. Specifically, (1) *Overwhelming Clarity and Confidence:* A significant majority of students, over 70%, agreed or strongly agreed that the AI use scale clearly communicated how to use and acknowledge AI tools in their assessments. This positive response extended to their confidence in using AI appropriately after consulting the scale. (2) *Effective Communication of AI Use Levels:* The scale was successful in detailing the different permitted levels of AI use across various assessment tasks. Over 75% of students (30.77% strongly agreeing and 44.23% agreeing) confirmed the scale's clarity in this regard. (3) *Low Levels of Disagreement:* For all questions regarding the scale's clarity and its effect on student confidence, fewer than 6% of respondents expressed disagreement. (4) *Strong Positive Consensus:* The median response for all three survey items was "Agree," further demonstrating that the scale was generally perceived by students as clear, helpful, and supportive.

Preliminary semi-structured interviews were purposely conducted with four students who represented a spectrum of AI adoption before taking the subject: a conservative user who avoided AI tools (**S1**); two strategic users who worked with AI tools only when explicitly permitted, demonstrating a moderate level of AI proficiency **(S2 & 3)**; and a proactive user who had deeply integrated AI into their daily life and studies, indicating a relatively high level of AI usage (**S4**).

Four key themes emerged from their experience with the scale. *(1) Clear Expectations.* All four students acknowledged that the AI use scale provided "clearer guidelines" by categorizing different types of AI uses, mapping them to specific assessment tasks, and outlining specific examples. It signalled a "very supportive" and "open-minded" stance by the teaching team. This clarity reduced uncertainty, particularly for S1-3, who reported feeling "more confident" to explore new AI uses within the defined boundaries. This aligns with the principles of providing subject- and assessment-specific adaptation with explicit guidance (DP1-3, 5). *(2) AI Literacy Enhancement.* The scale functioned as a pedagogical tool, particularly for the students with less prior AI use experience. S1-3 reported that the concrete examples and discussions helped them learn new ways to use AI and explore new AI functions, while also making them more aware of its limitations, such as the potential for "generating false information". This finding supports the scale's role as a learning-oriented framework that builds AI literacy and the importance of discussion between instructors and students (DPs 7-8, 9). *(3) Effective Communication*. Although S4 was already familiar with many of the AI uses outlined in the scale, they found the scale provided a common language that facilitated effective communication and coordination of AI use with their group members (DP6). *(4) Promoted Responsible AI Use.* By the end of the semester, all four students felt confident that they had used AI appropriately and properly acknowledged it as required, demonstrating that the scale's clear acknowledgment protocols were effective (DP4). S3&4 also noted a limitation of this preventive approach, as some students may still engage in misconduct without effective detection mechanisms in place. Nevertheless, all students would recommend the scale to other subjects.

Discussion

A key finding of this study is that generic, "one-size-fits-all" AI use policies are insufficient for the diverse landscape of higher education assessment. In contrast to rigid models like the binary "two-lane" approach or the generalized AI Assessment Scale (AIAS), this research proposes a "contextual lane approach" that empowers instructors to tailor AI use guidelines to the specific needs of their subject areas and individual assessment types. The effectiveness of this structure was supported by instructor reflections. The study highlights that ambiguity around acceptable AI use is a major contributor to student anxiety and unintentional academic misconduct. To address this, the proposed scale integrates explicit guidance, practical examples, and consistent terminology, establishing unambiguous expectations. This clarity is instrumental in shaping students' attitudes and intentions toward responsible GenAI use. Empirical findings from student interviews and surveys confirm that the scale provided clearer guidelines that made them feel more confident to explore GenAI within the defined boundaries. However, isolated adoption is insufficient. To make a broader impact on student concerns, a coordinated and university-wide adoption is needed. Crucially, the research advocates for treating these scales as "living documents," developed and refined through ongoing co-creation with students to both enhance their buy-in and leverage their practical expertise with GenAI tools. This collaborative model helps cultivate a shared culture of academic integrity grounded in mutual understanding and transparency. Ultimately, the scale is positioned not merely as a compliance tool but as a pedagogical framework. It supports "assessment for learning" by helping students build AI literacy needed to work productively with AI, skills that are increasingly essential in the contemporary workforce. Rather than focusing solely on preventing misuse, the scale encourages students to engage with AI in ways that deepen their learning and prepare them for real-world applications.

Conclusion

This study's proposed customizable, template-based AI use scale offers significant contributions to the ongoing discourse surrounding the integration of GenAI in higher education assessment. The findings challenge the efficacy of one-size-fits-all solutions and advocate for a more nuanced, pedagogically grounded "contextual lane approach" that aligns AI use with specific disciplinary and assessment contexts. Given that this study only conducted a pilot empirical evaluation, future research could involve more subjects and assessment types across multiple disciplines to validate and refine the proposed design principles and scale templates. In particular, further investigation is needed to explore the effectiveness of the last design principle, collaborative development, to gain insights into how co-creation helps enhance motivation, relevance, and responsible AI use.

References

Bakker, A. (2018). *Design research in education: A practical guide for early career researchers*. Routledge. https://doi.org/10.4324/9780203701010

Balalle, H., & Pannilage, S. (2025). Reassessing academic integrity in the age of AI: A systematic literature review on AI and academic integrity. *Social Sciences & Humanities Open*, *11*, 101299. https://doi.org/10.1016/j.ssaho.2025.101299

Biggs, J., Tang, C., & Kennedy, G. (2022). *Teaching for Quality Learning at University* (5th edition). Open University Press.

Bridgeman, A., & Liu, D. (2024, July 2). Frequently asked questions about the two-lane approach to assessment in the age of AI. *Teaching@Sydney*. https://educational-innovation.sydney.edu.au/teaching@sydney/frequently-asked-questions-about-the-two-lane-approach-to-assessment-in-the-age-of-ai/

Casal, J. E., & Kessler, M. (2023). Can linguists distinguish between ChatGPT/AI and human writing?: A study of research ethics and academic publishing. *Research Methods in Applied Linguistics*, *2*(3), 100068. https://doi.org/10.1016/j.rmal.2023.100068

Chaka, C. (2024). Reviewing the performance of AI detection tools in differentiating between AI-generated and human-written texts: A literature and integrative hybrid review. *Journal of Applied Learning and Teaching*, *7*(1), Article 1. https://doi.org/10.37074/jalt.2024.7.1.14

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*, 1–11. https://doi.org/10.1080/02602938.2025.2503964

Curtis, G. J. (2025). The two-lane road to hell is paved with good intentions: Why an all-or-none approach to generative AI, integrity, and assessment is insupportable. *Higher Education Research & Development*, 1–8. https://doi.org/10.1080/07294360.2025.2476516

Digital Education Council. (2025). *AI Meets Academia: What Faculty Think*. Digital Education Council. https://www.digitaleducationcouncil.com/post/digital-education-council-global-ai-faculty-survey

Khalil, M., & Er, E. (2023). Will ChatGPT get you caught? Rethinking of plagiarism detection. In P. Zaphiris & A. Ioannou (Eds.), *Learning and Collaboration Technologies* (pp. 475–487). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-34411-4_32

Lee, S. S., & Moore, R. L. (2024). Harnessing Generative AI (GenAI) for Automated Feedback in Higher Education: A Systematic Review. *Online Learning*, *28*(3), Article 3. https://doi.org/10.24059/olj.v28i3.4593

Liu, D., & Bridgeman, A. (2023, July 12). *What to do about assessments if we can't out-design or out-run AI? – Teaching@sydney*. https://educational-innovation.sydney.edu.au/teaching@sydney/what-to-do-about-assessments-if-we-cant-out-design-or-out-run-ai/

Nikolic, S., Wentworth, I., Sheridan, L., Moss, S., Duursma, E., Jones, R. A., Ros, M., & Middleton, R. (2024). A systematic literature review of attitudes, intentions and behaviours of teaching academics pertaining to AI and generative AI (GenAI) in higher education: An analysis of GenAI adoption using the UTAUT framework. *Australasian Journal of Educational Technology*, *40*(6), Article 6. https://doi.org/10.14742/ajet.9643

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*(6), Article 06. https://doi.org/10.53761/q3azde36

Perkins, M., & Roe, J. (2024). Decoding academic integrity policies: A corpus linguistics investigation of AI and other technological threats. *Higher Education Policy*, *37*(3), 633–653. https://doi.org/10.1057/s41307-023-00323-2

TEQSA. (2022, October 13). *What is academic integrity? | tertiary education quality and standards agency*. Tertiary Education Quality and Standards Agency. https://www.teqsa.gov.au/students/understanding-academic-integrity/what-academic-integrity

Walter, Y. (2024). Embracing the future of artificial intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education. *International Journal of Educational Technology in Higher Education*, *21*(1), 15. https://doi.org/10.1186/s41239-024-00448-3

Copyright statement