



BUCKINGHAMSHIRE NEW UNIVERSITY

EST. 1891

Downloaded from: <https://bnu.repository.guildhe.ac.uk/>

This document is made available with permission. Copyright remains with the author(s).

Use of items from Buckinghamshire New University's institutional repository must comply with the usage guidelines.

This work is made available under a **Creative Commons Attribution-NonCommercial 4.0 International Licence (CC BY-NC 4.0)**: <https://creativecommons.org/licenses/by-nc/4.0/>

Please note that you must also do the following;

- cite the authors, title and full bibliographic details of the item
- include a hyperlink/URL to the repository record in the citation
- keep all files required for use of the item together

You may not

- sell any part of an item
- refer to any part of an item without citation
- use or adapt the work in a way that misrepresents the author or impugns their reputation
- remove or alter the copyright statement on an item

If you need further guidance contact the Library and Learning Services
at Library@bnu.ac.uk

Bridging the trust gap: A mixed-methods study of United Kingdom educators' perceptions of generative artificial intelligence in higher education assessment

Abstract

Situated against the backdrop of an evolving regulatory landscape, this study investigates the perceptions of United Kingdom university educators regarding the use of generative artificial intelligence in summative assessment. Through a mixed-methods survey of 330 academic staff, the research identifies a significant gap between high levels of digital literacy and low evaluative trust. While respondents report high familiarity with AI tools, confidence in the ability of artificial intelligence to provide fair or nuanced marks remains low. Qualitative analysis of open text questions reveals a nuance crisis, where educators perceive these systems as incapable of comprehending complex student submissions. The findings reveal significant gender disparities in technical proficiency, however, a consistent pattern of low evaluative trust persists across teaching experience levels, with no statistically significant differences detected based on professional seniority, suggesting a potential need to consider alternative pedagogical frameworks, such as performative and synchronous assessments, to safeguard academic integrity.

Keywords: Artificial intelligence; Generative AI; Academic assessment; Higher education; Teacher perceptions; UK universities; Technology acceptance.

Introduction

The integration of artificial intelligence into higher education represents a transformative shift in pedagogy where technologies have evolved from novelty experiments to fundamental forces redefining knowledge assessment and feedback delivery (Qu et al., 2025). While the promise of workload reduction is significant, the academic community remains caught in a paradoxical shift that balances the efficiency of digital automation against the necessity of preserving human-centred values and academic integrity (Dubey & Das, 2023; Hooda et al., 2022).

This research is situated at a pivotal juncture in mid-2025, capturing educator sentiment during a period of intense national anticipation regarding generative AI regulation in higher education. While earlier discourse focused primarily on rudimentary scoring engines (Alsalem, 2024), the rapid advancement of sophisticated generative systems has outpaced institutional policy, creating an urgent need for empirical data on teacher attitudes (Kofinas et al., 2025). Consequently, this study captures a baseline of academic perceptions immediately preceding the formalisation of sector-wide standards, providing a necessary empirical foundation against which subsequent regulatory frameworks, such as the UK's Office of Qualifications and Examinations Regulation (Ofqual, 2026) *Principles of AI use in marking* working paper, can be evaluated post-hoc.

Within the United Kingdom, a profound tension has emerged regarding the professional trust educators place in these automated systems. Recent literature suggests that the adoption of such tools is met with deep-seated anxiety and a plea for circumspection among academic staff (Maistry & Singh, 2025). However, it remains unclear whether high levels of technical proficiency among educators actually resolve evaluative scepticism or if they instead lead to a more critical and apprehensive assessment of the technology's inherent limitations (Nazaretsky et al., 2022).

The conceptual grounding for this transition is provided by the unified theory of acceptance and use of technology which suggests that adoption is contingent upon performance expectancy, effort expectancy, and professional trust (Maistry & Singh, 2025). In the current landscape, institutional endorsements of generative tools often overlook the psychological tensions introduced by the artificial intelligence 'ghostwriter effect', where the responsibility for authorship is increasingly externalised to the machine (Qu et al., 2025). This effect contributes to a phenomenon of conceptual non-disclosure, where students may obscure their reliance on generative tools to maintain an image of intellectual authenticity, thereby creating a profound disconnect between the marker and the marked (Qu et al., 2025). For educators, this creates an environment of professional uncertainty where the technology is perceived to lack the capacity for genuine comprehension. This nuance crisis suggests a potential breakdown in the evaluative relationship, where automated systems may fail to appreciate the complex argumentation and individual learning trajectories that define higher education assessment (Jonäll, 2024; Stashevskaja, 2024).

Beyond individual pedagogical concerns, the broader problem encompasses systemic anxieties regarding staff precarity and the potential dehumanisation of the expert relationship (Jonäll, 2024; Ragolane et al., 2025; Stashevskaja, 2024; Van Den Berg & Papadopoulos, 2025). The adoption of automated marking is increasingly viewed through the lens of institutional commodification, raising significant questions about worker rights, the erosion of academic standards, and the value of the university degree itself (Maistry & Singh, 2025).

Despite the proliferation of localised institutional guidelines and mounting sector pressure for a unified national strategy, a significant research lacuna remains regarding how this lack of formal standardisation impacts the professional judgement and ethical orientations of United Kingdom academic staff. This study addresses this gap by investigating the following research questions.

RQ1: What are the perceptions of United Kingdom university educators regarding the use of generative artificial intelligence for summative marking?

RQ2: To what extent does an educator's self-perceived technical proficiency influence their evaluative trust in automated grading systems?

RQ3: What thematic tensions and ethical anxieties, such as the nuance crisis or staff precarity, drive the rejection of wholly automated assessment models in the academy?

Theoretical Framework

The conceptual architecture of this study is grounded in the Unified Theory of Acceptance and Use of Technology (UTAUT), which posits that, performance expectancy (the belief that a tool improves job performance) and effort expectancy (the ease of use) are the primary drivers of technology adoption (Maistry & Singh, 2025). However, in the context of generative artificial intelligence, traditional UTAUT constructs must be extended to account for the unique psychological and ethical dimensions of professional trust.

Professional trust is defined here as an educator's willingness to delegate evaluative authority to an automated system in high-stakes assessment contexts. This definition builds upon foundational models of organisational trust, most notably Mayer, Davis, and Schoorman's (1995) integrative framework, which posits that trust is a willingness to be vulnerable based on perceived ability, benevolence, and integrity. In a high-stakes educational environment, this vulnerability manifests as ceding qualitative marking authority to an algorithmic system. When applied specifically to digital tools, McKnight et al.'s (2011) framework of technology-specific trust outlines that an individual's willingness to engage in post-adoption exploration is predicted by explicit experiential beliefs regarding the software's functional attributes. Rather than treating technical reliability as a proxy for capability, the authors (McKnight et al., 2011) disaggregate specific trust into three distinct, co-varying dimensions: functionality (the belief that the system has the feature sets and capacity to execute required tasks), reliability (the expectation of consistent operation free of glitches or failures), and helpfulness. They explicitly note that while a technology's 'functionality' is conceptually similar to interpersonal 'competence', it must be evaluated strictly in terms of amoral, non-volitional feature sets. However, while Mayer et al. (1995) assumes that perceived ability satisfies task-specific needs and that integrity mitigates systemic vulnerability, it inherently does not account for the unique cognitive and ethical boundaries introduced by generative artificial intelligence.

By explicitly integrating these concepts, this research expands the UTAUT model (Khlaif et al., 2024; Ragolane et al., 2025; Van Den Berg & Papadopoulos, 2025) by identifying evaluative nuance and ethical accountability as critical mediators that can disrupt the relationship between technical familiarity (effort expectancy) and adoption (behavioural intention). Within the United Kingdom higher education landscape, the hybrid imperative arises from a theoretical tension: while institutional 'facilitating conditions' encourage AI adoption for efficiency, the individual educator's 'performance expectancy' is fundamentally capped by the perceived inability of machines to replicate the expert relationship or manage the nuance crisis of human-centric feedback. By adopting this extended framework, the study evaluates whether technical self-efficacy is sufficient to overcome the principled ethical resistance inherent in professional academic identity.

This extended framework directly builds upon, yet substantively differs from, emergent models of trust in educational artificial intelligence, such as those pioneered by Nazaretsky et al. (2022). The authors focus primarily on K-12 environments, framing teacher trust as a dynamic psychological attitude that can be positively modified through professional development programs (PDPs) that leverage procedural transparency to mitigate cognitive

heuristics like algorithm aversion and confirmation bias (Nazaretsky et al., 2022). While our framework acknowledges these cognitive barriers, it departs from an interventionist approach by integrating trust structurally within a modified technology acceptance architecture tailored to higher education. Rather than treating scepticisms as operational misunderstandings correctable via training, our extended model posits that 'evaluative nuance' and 'ethical accountability' serve as entrenched, macro-level mediators that actively cap 'performance expectancy'. Thus, the novelty of this extension lies in its capacity to psychometrically demonstrate that high 'effort expectancy' (technical literacy) cannot drive 'behavioural intention' if the technological implementation violates the expert-student relationship or creates institutional vulnerability. We shift the focus from psychological debiasing to structural identity barriers within the academy.

Methods

This study adopted a convergent mixed-methods design (Pachori et al., 2026) to investigate the perceptions of United Kingdom educators regarding the use of generative artificial intelligence in summative assessment. By collecting quantitative survey data and qualitative open-ended responses simultaneously through a single instrument, the research seeks to achieve methodological triangulation, ensuring findings are statistically robust while reflecting the complexity of individual professional experiences (Lee, 2023; Xiao et al., 2019). This approach is essential for capturing the complex reality of technology acceptance in higher education, as quantitative data alone are often insufficient to fully grasp the underlying factors in intricate areas like artificial intelligence (Stashevskaya, 2024; Van Den Berg & Papadopoulos, 2025). The research was conducted using the Jisc Online Surveys platform, which is widely utilised in the United Kingdom higher education sector for its compliance with data protection and accessibility standards. Ethical approval was obtained from Northumbria University Ethics Committee and all participants provided informed consent before beginning the survey, adhering to the ethical guidelines of the British Educational Research Association [BERA] (2024).

The final sample consists of 330 valid responses from academic staff members collected between 12 May and 30 September 2025 (see Table 1). Participants were recruited through a multi-channel strategy involving convenience and snowball sampling, a practical approach for capturing the views of a wide demographic of busy academic professionals (Ragolane et al., 2025; Yan, Sha, et al., 2024). For the purpose of comparative analysis, participants were categorised by their teaching experience into three distinct groups: junior educators with 0-5 years of experience, experienced educators with 6-19 years, and senior educators with 20 or more years of teaching practice.

[Insert Table 1 about here]

The survey instrument was informed by UTAUT, where specific items could be mapped to established theoretical constructs (Maistry & Singh, 2025). 'Performance expectancy' was measured through items regarding the accuracy of machine marking and the fairness of automated grades, specifically assessing confidence in the ability of current systems to accurately mark end-of-module assignments. 'Effort expectancy' and self-efficacy were captured through measures of digital literacy and tool familiarity, specifically asking respondents to rate their confidence in their own digital skills and their familiarity with generative artificial intelligence tools like ChatGPT, Claude, and Gemini. 'Facilitating conditions' were examined through questions regarding institutional data security and the privacy of uploading assignments to university-approved systems.

Each item utilised a five-point scale. These variables were operationalised as independent single-item measures rather than multi-item psychometric scales. This design choice was driven primarily by severe survey length constraints and the necessity to minimise respondent burden and survey fatigue within a highly surveyed professional population, thereby safeguarding completion rates across a broad national sample ($N = 330$). Furthermore, the use of single-item indicators for distinct exploratory constructs (such as immediate evaluative trust or self-perceived literacy) follows established precedents in recent higher education and educational technology literature examining rapid institutional shifts, where concise instruments are critical to capturing timely baseline attitudes (e.g., Khlaif et al. (2024); Yan, Echeverria, et al. (2024)). Consequently, because these are treated as single-item indicators rather than multi-item composite scales, traditional internal consistency metrics such as Cronbach's alpha statistics are mathematically inappropriate (Yan, Echeverria, et al., 2024).

Quantitative data analysis was performed using SPSS version 29.0.2.0. Descriptive statistics including means and standard deviations were generated to establish the baseline for digital literacy and evaluative trust. To determine whether significant differences existed between demographic groups, non-parametric tests were employed due to the ordinal nature of the Likert-scale data. Disparities between gender (two independent groups) were analysed using the Mann-Whitney U test. For teaching experience (three independent groups), the Kruskal-Wallis H test was utilised. These tests were selected to establish whether technical self-efficacy and evaluative trust significantly varied across career stages or gender identities.

The qualitative component of the study utilised a Gioia-inspired thematic analysis to ensure procedural transparency and conceptual rigour (Gioia et al., 2013). Crucially, no participant responses were excluded from the final qualitative dataset, as all 330 respondents actively provided valid written answers to the open-ended fields, meaning uninformative strings (such as "N/A" or "I don't know") were entirely absent from the corpus. The open-ended questions targeted core qualitative richness across several focused fields, prompting educators to elaborate on: institutional training and policy guidance (Q14 & Q15); the classroom visibility of student AI use (Q17); the moral and ethical boundaries of AI grading (Q24); the practical benefits and disruptions of automated marking (Q26 & Q27); the evolutionary future of AI in assessment (Q28); expectations of student transparency (Q29); and an open field for additional professional reflections (Q30).

The usable qualitative responses exhibited immense textual richness and structural variety, ranging from succinct, focused bullet points to deeply reflective, passionate, paragraph-length essays. This structural depth was particularly pronounced in discussions surrounding ethical risks and automated evaluation, where educators provided detailed, nuanced critiques regarding the environmental footprints of large language models, data privacy violations, algorithmic biases against neurodivergent or international students, and the perceived futility of an educational model where both production and grading are externalised to machines. Conversely, when weighing perceived benefits, such as increased marking speed and cohort scalability, or addressing student transparency (Q29), the text was characterised by highly contextualised reasoning and an emphatic, almost unanimous demand for "complete transparency" and "100% disclosure".

In sum, while the closed-ended items generated scalable quantitative metrics, the complete qualitative dataset successfully captured an extensive, high-quality dialogue from all 330 educators. The analysis of this rich material was a collaborative effort involving all five authors who participated in a systematic, inductive coding process. Raw participant responses were first coded into first-order concepts that remained close to the original language of the educators. These were subsequently grouped into second-order themes, such as 'The Nuance Crisis' and the 'Loss of the Expert Relationship,' to ensure the nuances of educator experience were accurately captured. Finally, these themes were distilled into aggregate dimensions including the 'dehumanisation of education' and the 'erosion of trust and accountability'. To ensure the reliability and validity of the findings, the coding process underwent a formal verification stage where authors cross-checked categories to reach a consensus, thereby establishing high inter-rater reliability for the emergent themes.

Results and Discussion

This section presents an integrated analysis of the empirical findings from 330 United Kingdom university educators, synthesising quantitative statistical trends and qualitative thematic depth through the lens of our extended UTAUT framework. By merging data presentation with conceptual interpretation, we directly examine how evaluative nuance and ethical accountability act as critical mediators that disrupt the pathway between technical familiarity (effort expectancy) and institutional adoption (behavioural intention).

General Perceptions of Generative AI for Summative Marking (RQ1)

The Hybrid Imperative and the Preservation of the Expert Relationship (RQ1)

The empirical data reveal a dominant hybrid imperative among respondents, with 62.4% advocating for a hybrid model where AI assists human markers, while 36.4% reject automation entirely in favour of wholly human-marked processes (see Figure 1). Only a marginal minority (1.2%) support fully automated grading.

[Insert Figure 1 about here]

When interpreted through our extended UTAUT framework, this overwhelming resistance to full automation demonstrates that institutional 'facilitating conditions' (such as the provision of AI tools) fail to translate into behavioural intention because 'performance expectancy' is fundamentally capped by a desire to preserve the expert relationship. The qualitative data illuminate this tension. Educators frame assessment not as a mechanical process of rubric verification, but as a core component of professional academic identity. As Participant 301 noted, "*students expect their work to be marked by a lecturer expert in the field, which is something an AI is not.*"

This rejection of wholly automated models reflects a deeper anxiety regarding the 'dehumanisation of education' (Table 3), where assessment is viewed as an interactive dialogue. Outsourcing this boundary to algorithmic systems transforms the educational contract into a transactional exchange, validating Maistry and Singh's (2025) assertion that faculty perspectives remain deeply anchored in human-centric pedagogical values despite institutional pressures to modernise.

This defensive posture toward core evaluative tasks is heavily reinforced by contemporary qualitative evidence from European and North African contexts. Chaieb et al. (2026, p. 11) introduce the construct of "pedagogical role integrity" within their Job Characteristics Model, proving that educators systematically shield high-stakes, student-centred tasks (such as grading, deep dialogue, and mentorship) from automation. Much like our participants who view marking as an interactive dialogue, international faculty frame core pedagogical boundaries not merely as functional workflows, but as profound moral and professional accountabilities. Consequently, when automated tools encroach upon these identity-associated functions, educators experience a "deep misfit" (Chaieb et al., 2026, p. 3) with their underlying pedagogical philosophies, driving a conscious decision to restrict GenAI strictly to low-stakes, peripheral administrative assistance.

The Digital Literacy Scepticism Paradox: Validating the Extended UTAUT Model (RQ2)

The quantitative analysis exposes a profound trust gap: high technical proficiency (effort expectancy) does not correlate with professional confidence in evaluative fairness (performance expectancy). While respondents reported high self-perceived digital literacy (Mean = 4.12, SD = .938) and tool familiarity (Mean = 3.78, SD = 1.067), confidence that these systems can provide a fair mark dropped precipitously to a mean of 2.28 (see Table 2).

This finding provides strong empirical validation for our proposed extension of the UTAUT model. In classical UTAUT applications, high ease of use (effort expectancy) directly promotes adoption. Here, however, the pathway is severely disrupted by the mediating variables of evaluative nuance and ethical accountability.

Interestingly, the trust gap remains highly prominent across all career stages. While Mann-Whitney U tests indicate that male respondents reported statistically higher digital literacy confidence ($p = .003$) and tool familiarity ($p = .009$) than female respondents, Kruskal-Wallis H tests revealed no statistically significant differences across teaching experience levels regarding evaluative trust ($p = .073$; see Table 2). Although this marginal p -value approaches the threshold of conventional significance, no definitive statistical divergence was detected, suggesting a generally consistent pattern of evaluative scepticism across career stages. This shared professional baseline challenges traditional institutional strategies that focus purely on technical upskilling. Because the magnitude of the gender divergence regarding tool familiarity is small ($r = .149$), yet the drop-off in evaluative trust is catastrophic and uniform across all demographics (Mean = 2.28), it provides definitive empirical proof that technical familiarity (effort expectancy) possesses negligible leverage over professional acceptance.

The low effect sizes on technical variables demonstrate that incremental increases in software exposure do not dilute ethical anxiety. Consequently, providing basic operational training will fail to bridge the trust gap unless universities directly confront the underlying 'nuance crisis' and the structural opacity of generative evaluation frameworks.

This breakdown of the classical adoption pathway directly exposes the systemic blind spots of linear technology acceptance frameworks. Traditional models like the TAM and standard UTAUT over-prioritise technical parameters such as ease of use and perceived functional utility. However, as both our quantitative effect sizes and recent task-based adoption research show, ease of use can occasionally act as an intellectual deterrent for academics if it is perceived to flatten pedagogical complexity or foster superficial student engagement. Chaieb et al. (2026) similarly document that technical familiarity does not yield professional acceptance in evaluative spaces. Instead, educators prioritise alignment with their epistemological commitments and moral values over sheer technical efficiency. The breakdown is therefore systemic: when professional trust is compromised by ethical ambiguities, operational capability is rendered secondary.

[Insert Table 2 about here]

Structural and Ethical Anxieties: The Nuance Crisis and Staff Precarity (RQ3)

To fully close the loop with our theoretical framing, the three aggregate dimensions derived from the Gioia thematic analysis, The Dehumanisation of Education, Erosion of Trust and Accountability, and Staff Precarity (Table 3), must be understood as explicit conceptual barriers to performance expectancy.

The first driver, the 'nuance crisis', stems from the perceived inability of deep neural networks to comprehend complex, non-linear student submissions. This structural limitation is best understood by categorising academic responsibilities into routine and non-routine tasks. While generative systems present a high functional fit for routine cognitive tasks such as

citation formatting, syllabus structuring, or basic rubric verification, they exhibit a severe, unresolvable misfit when applied to non-routine interactive or interpersonal tasks. As noted by international literature educators, algorithmic systems inherently fail to comprehend metaphor, sarcasm (Awidi, 2024), or complex semantic nuance (Stashevskaya, 2024), attributes that are natural to the human expert relationship. Educators reject fully automated marking because evaluating non-routine analytical submissions requires contextual acumen, human intuition, and emotional intelligence that fundamentally exceed the boundaries of predictive text generation. This aligns closely with the UK Office of Qualifications and Examinations Regulation's (Ofqual, 2026) recent working paper, which notes that generative systems lack genuine semantic understanding. Educators fear an evaluative vacuum where "the results can be judged by a probabilistic monkey" (P254), leading to a 'farce' where both essay production and marking are externalised to machines (Qu et al., 2025).

Secondly, the erosion of accountability is tied directly to data ethics and the 'black box' problem (P118). Educators refuse to cede evaluative authority when algorithmic errors cannot be transparently justified to students, exposing a major conflict with institutional data privacy expectations (Mean = 2.48, See Table 2).

Finally, the dimension of 'Staff Precarity' highlights a crucial systemic divergence from state regulatory perspectives. While the Ofqual (2026, p. 63) framework depersonalises this risk by framing it as a threat to "total assessment expertise," our qualitative data reveal that educators experience this as an existential threat to worker rights and professional status. Senior educators, in particular, frame their resistance as a principled defence against institutional commodification and the de-skilling of the academic workforce (e.g., P143, P157). This confirms that technology acceptance in higher education cannot be divorced from the broader sociopolitical anxieties of staff precarity and labour uncoupling.

[Insert Table 3 about here]

Discussion

The findings of this study provide a critical assessment of the human-AI marking paradox within United Kingdom higher education in 2025. While educators demonstrate high levels of digital literacy, their scepticism regarding the ability of artificial intelligence to deliver nuanced and fair marks suggests that familiarity does not inherently translate into professional trust. This discussion situates these results within the broader context of evolving authorship norms and assessment design.

A central finding of this research is the widespread rejection of fully automated marking, which aligns with recent experimental evidence suggesting that while generative systems can produce authentic-looking assessments, human markers often perceive a perfection in the output that triggers suspicion. This nuance crisis echoes the concerns of Kofinas et al. (2025) regarding obscured authorship and the altered nature of knowing in the digital age. The educator scepticism observed here contrasts with student perspectives where generative tools

are increasingly framed as personal cognitive aids rather than external authors. This suggests a growing pedagogical gap where students are externalising their cognitive processes through conceptual non-disclosure while educators remain anchored in traditional evaluative frameworks (Qu et al., 2025).

A comparative analysis of this study's findings and the regulatory framework outlined in the January 2026 Ofqual paper, *Principles of AI use in marking*, reveals significant thematic alignment regarding the limitations and risks of automated assessment. Both the empirical data from United Kingdom higher education and the regulatory working paper identify the 'black box' nature of current deep neural networks as a primary barrier to transparency and acknowledge that generative systems lack the capacity for true semantic understanding or human-like judgment. This common ground leads both this study and the Ofqual paper to reject purely automated marking in favour of hybrid systems that keep human experts in charge of quality.

However, a key area of divergence lies in the framing of professional resistance. While the qualitative findings of this research identify 'Staff Precarity' and the personal threat of job displacement as primary drivers of academic scepticism, the Ofqual report frames this issue through the broader systemic lens of safeguarding "total assessment expertise" (Ofqual, 2026, p. 63) within the education apparatus to prevent operational vulnerabilities. Furthermore, while this study documents a universal trust gap that transcends professional seniority, the Ofqual (2026, p. 27) paper situates these concerns within the broader "social contract" of qualifications and the functional necessity for transparency to uphold public confidence in the value of degrees.

Our findings demonstrate that, while gender is associated with differences in technical self-efficacy (effort expectancy), the trust gap regarding evaluative fairness remains a universal professional constant across the sample. This validates the study's extended UTAUT model, confirming that high effort expectancy is insufficient to drive behavioural intention when professional trust is compromised by ethical anxieties. The qualitative mapping in Table 3 indicates that educator scepticism is a principled pedagogical stance based on the preservation of the expert relationship, which acts as a conceptual barrier to performance expectancy. Consequently, institutional strategies focusing solely on technical training (effort expectancy) will likely fail to bridge the trust gap unless they address the deeper 'evaluative nuance' and accountability concerns that currently cap the perceived utility of generative AI.

Although this study is situated within the unique regulatory landscape of United Kingdom higher education, the underlying tensions identified here transcend national borders and hold clear relevance for the international academic community. The 'nuance crisis' and anxieties regarding staff precarity are not localised anomalies. Rather, they represent systemic, global challenges as universities worldwide grapple with rapid technological disruption alongside the ongoing marketisation of higher education. When viewed through our extended UTAUT lens, the friction between institutional operational targets (facilitating conditions) and individual academic identity remains a universal constant. In countries with highly globalised higher education frameworks, such as Australia, Canada, and the United States (Sullivan et

al., 2023), educators face identical challenges regarding obscured student authorship, data privacy, and the preservation of the human expert relationship. Consequently, our findings offer a valuable comparative baseline for international institutions, demonstrating that technical training initiatives will likely meet structured resistance globally unless policy-makers directly confront the cross-cultural ethical anxieties surrounding evaluative fairness and professional deskilling.

Limitations

While the convergent mixed-methods design provides a robust, dual-perspective exploration of educator attitudes, several methodological limitations warrant explicit reflection to enhance transparency and qualify the scope of the findings.

Firstly, the use of convenience and snowball sampling through a multi-channel recruitment strategy introduces a potential self-selection bias. Educators with pre-existing, strong opinions regarding generative artificial intelligence may have been more motivated to participate, meaning the descriptive results may not perfectly mirror the wider United Kingdom higher education sector. Furthermore, certain demographic segments, specifically those identifying as 'Other' ($n = 3$) or 'Prefer not to say' ($n = 14$), were of insufficient sample size to be included in comparative non-parametric testing, meaning their unique perspectives could not be fully disaggregated quantitatively.

Secondly, the study's geographical scope is confined entirely to a single national framework, the United Kingdom higher education sector. Because higher education funding, regulatory approaches (such as the Ofqual framework), and institutional policies vary significantly across borders, these findings may lack immediate generalisability to alternative national or international educational contexts.

Thirdly, the data rely entirely on self-reported measures, which are inherently vulnerable to social desirability and recall biases. In high-stakes environments where academic integrity and staff precarity are sensitive topics, respondents may over-report their digital literacy skills or subtly align their qualitative answers with prevailing professional norms. Similarly, the use of single-item measures to reduce survey fatigue precludes the calculation of internal consistency metrics like Cronbach's alpha and may capture less latent nuance than multi-item psychometric scales.

Finally, this study utilises a cross-sectional design, capturing educator perceptions at a single, specific point in time between May and September 2025. Because data collection was synchronous, it is impossible to infer definitive causality regarding changes in attitudes over time, particularly as generative technologies and institutional policies continue to evolve rapidly. Future research would benefit from employing longitudinal designs to track shifting perceptions, purposive international stratification to cross-examine national differences, and multi-item validated scales to further scrutinise the psychometric paths of the extended UTAUT model.

Conclusion

This study has examined the complex intersection of generative artificial intelligence and academic assessment within United Kingdom higher education during a period of rapid technological transition. The findings reveal a significant trust gap where educators demonstrate high digital literacy and familiarity with AI tools yet remain deeply sceptical of the ability of the technology to replicate the nuanced qualitative judgement essential for fair and effective assessment. This tension is further complicated by clear demographic divides where gender influences technical self-efficacy, but evaluative scepticism appears to exhibit a consistent pattern across both junior and senior career stages. The significant appetite for hybrid models (62.4%) suggests a pragmatic pathway where AI provides efficiency for low-stakes, structured administrative tasks, such as rubric verification, while high-stakes summative assignments remains protected by human intervention. Our findings suggest that this distinction could be important for institutional quality assurance and assessment validity.

However, current adoption is hindered by a 'nuance crisis' and low confidence in institutional data security. These are structural and ethical anxieties that require robust policy reconciliation to uphold the professional social contract of degree-awarding bodies. The results indicate that generative artificial intelligence is not merely a tool for automation but a catalyst for a fundamental re-evaluation of what constitutes knowing and authoring in the digital age (Ragolane et al., 2025; Twabu & Nakene-Mginqi, 2024). These findings indicate that addressing the 'trust gap' necessitates more than just a gender equity approach to technical proficiency, it highlights a potential widening of the digital divide through unequal access to training. Institutional priority should be placed on uniform policy adoption for upskilling staff to ensure that technological transitions do not further entrench existing professional disparities. Moving forward, institutions must move away from generic guidelines in favour of frameworks that acknowledge these diverse demographic and professional realities (Ragolane et al., 2025; Stashevskaya, 2024).

Furthermore, while not directly evaluated by the empirical data of this study, the documented capacity for these tools to successfully mimic authentic written assessments leads us to infer that a broader pedagogical paradigm shift may be required. Future institutional strategies should critically consider moving toward performative and synchronous evaluations as a means to verify true student learning, though further research is necessary to evaluate the practical effectiveness and scalability of such assessment frameworks.

References

- Alsalem, M. S. (2024). EFL teachers' perceptions of the use of an AI grading tool (CoGrader) in English writing assessment at Saudi universities: an Activity Theory Perspective. *Cogent Education*, 11(1), 2430865. <https://doi.org/10.1080/2331186X.2024.2430865>
- Awidi, I. T. (2024). Comparing expert tutor evaluation of reflective essays with marking by generative artificial intelligence (AI) tool. *Computers and Education: Artificial Intelligence*, 6, 100226. <https://doi.org/10.1016/j.caeai.2024.100226>
- British Educational Research Association [BERA]. (2024). *Ethical Guidelines for Educational Research*. www.bera.ac.uk/publication/ethicalguidelines-for-educational-research-2024
- Chaieb, M., Cuel, R., & Bouzaabia, R. (2026). Reconceptualizing of GenAI adoption in higher education: A task-based perspective. *The International Journal of Management Education*, 24(2), 101365. <https://doi.org/10.1016/j.ijme.2026.101365>
- Dubey, M. R. K., & Das, S. K. (2023). An impact of artificial intelligence on marking research: An analysis. *Dogo Rangsang Research Journal* 13(6), 18-25.
- Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. *Organizational Research Methods*, 16(1), 15-31. <https://doi.org/10.1177/1094428112452151>
- Hooda, M., Rana, C., Dahiya, O., Rizwan, A., & Hossain, M. S. (2022). Artificial intelligence for assessment and feedback to enhance student success in higher education. *Mathematical Problems in Engineering*, 2022(1), 5215722. <https://doi.org/10.1155/2022/5215722>
- Jonäll, K. (2024). *Artificial intelligence in academic grading: A mixed-methods study* [MSc, University of Gothenburg].
- Khlaif, Z. N., Ayyoub, A., Hamamra, B., Bensalem, E., Mitwally, M. A., Ayyoub, A., Hattab, M. K., & Shadid, F. (2024). University teachers' views on the adoption and integration of generative AI tools for student assessment in higher education. *Education Sciences*, 14(10), 1090. <https://doi.org/10.3390/educsci14101090>
- Kofinas, A. K., Tsay, C. H. H., & Pike, D. (2025). The impact of generative AI on academic integrity of authentic assessments within a higher education context. *British Journal of Educational Technology*, 56, 2522–2549. <https://doi.org/10.1111/bjet.13585>
- Lee, A. V. Y. (2023). Supporting students' generation of feedback in large-scale online course with artificial intelligence-enabled evaluation. *Studies in Educational Evaluation*, 77, 101250. <https://doi.org/10.1016/j.stueduc.2023.101250>
- Maistry, S. M., & Singh, U. G. (2025). Faculty perspectives on the role of ChatGPT-4.0 in higher education assessments. *Journal of Education (University of KwaZulu-Natal)*(98), 86-102.
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An Integrative Model of Organizational Trust. *The Academy of Management Review*, 20(3), 709-734. <https://doi.org/10.2307/258792>
- McKnight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures: An investigation of its components and measures. *ACM transactions on management information systems*, 2(2), 1-25. <https://doi.org/10.1145/1985347.1985353>
- Nazaretsky, T., Ariely, M., Cukurova, M., & Alexandron, G. (2022). Teachers' trust in AI-powered educational technology and a professional development program to improve it. *British Journal of Educational Technology*, 53(4), 914-931. <https://doi.org/10.1111/bjet.13232>

- Ofqual. (2026). *Principles of AI use in marking*. Retrieved January 2026 from <https://www.gov.uk/government/publications/principles-of-ai-use-in-marking>
- Pachori, R., Devarakonda, R. V., & Kumar, R. (2026). From crisis to transformation: Evaluating the implementation of DIKSHA in India's EdTech landscape. *British Journal of Educational Technology*. <https://doi.org/10.1111/bjet.70042>
- Qu, Y., Loo, H. E., & Wang, J. (2025). Generative artificial intelligence in higher education: Emotional tensions and ethical declaration. *British Journal of Educational Technology*, 0(0), 1-20. <https://doi.org/10.1111/bjet.70029>
- Ragolane, M., Patel, S., & Phiri, J. (2025). Navigating Artificial Intelligence in Assessment: Academic Professionals' Perceptions of AI Grading in South Africa. *Technium Education and Humanities*, 11, 156-172.
- Stashevskaja, E. (2024). *How do teachers envision AI grading for open-ended questions in universities?* University of Twente].
- Sullivan, M., Kelly, A., & McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning & Teaching*, 6(1), 31-40. <https://doi.org/10.37074/jalt.2023.6.1.17>
- Twabu, K., & Nakene-Mginqi, M. (2024). Developing a design thinking artificial intelligence driven auto-marking/grading system for assessments to reduce the workload of lecturers at a higher learning institution in South Africa. *Frontiers in Education*, 9, 1512569. <https://doi.org/10.3389/feduc.2024.1512569>
- Van Den Berg, S., & Papadopoulos, P. M. (2025). Summative assessment with Artificial Intelligence: Qualitative analysis and comparison of technology acceptance in student and teacher populations. *Innovations in Education and Teaching International*, 62(5), 1529-1544. <https://doi.org/10.1080/14703297.2024.2436613>
- Xiao, C., Qiu, H., & Cheng, S. M. (2019). Challenges and opportunities for effective assessments within a quality assurance framework for MOOCs. *Journal of Hospitality, Leisure, Sport & Tourism Education*, 24, 1-16. <https://doi.org/10.1016/j.jhlste.2018.10.005>
- Yan, L., Echeverria, V., Jin, Y., Fernandez-Nieto, G., Zhao, L., Li, X., Alfredo, R., Swiecki, Z., Gašević, D., & Martinez-Maldonado, R. (2024). Evidence-based multimodal learning analytics for feedback and reflection in collaborative learning. *British Journal of Educational Technology*, 55(5), 1900-1925. <https://doi.org/10.1111/bjet.13498>
- Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., & Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, 55(1), 90-112. <https://doi.org/10.1111/bjet.13370>

Tables

Table 1: Participant Demographics (N = 330)

Demographic Category	Sub-Category	Frequency (n)	Percentage (%)
Age	21–24	1	0.3%
	25–30	4	1.2%
	31–39	68	20.6%
	40–49	111	33.6%
	50–59	104	31.5%
	60+	42	12.7%
Geographic Location	England	280	84.8%
	Scotland	16	4.8%
	Wales	24	7.3%
	Northern Ireland	10	3.0%
Gender	Male	142	43.0%
	Female	171	51.8%
	Other	3	0.9%
	Prefer not to say	14	4.2%
Experience Teaching	Junior Educator	53	16.1%
	Experienced Educator	140	42.4%
	Senior Educator	137	41.5%

Note: Items measured on a 1 (low) to 5 (high) scale

[PRODUCTION NOTE TO TYPESETTER: Due to its substantial width, please display Table 2 in Landscape Orientation or utilising a reduced font size to ensure optimal legibility and prevent clipping in the final PDF compilation.]

Table 2: Non-parametric analysis of technical proficiency and evaluative trust by gender and teaching experience (Mann-Whitney U and Kruskal-Wallis H tests)

	N	Mean	Standard Deviation	Gender	Gender N	Gender Mean	Mean Rank	Sum of Ranks	Mann-Whitney U	Z	u-test (p)	Mann-Whitney U Effect Sizes (r)	Experience	Experience N	Kruskal-Wallis H	Kruskal-Wallis p
How confident are you in your own digital literacy?	330	4.12	.938	Male	142	4.25	172.77	24534.00	9901.0	-3.010	.003	.170	Junior Educator	53	3.305	.192
				Female	171	3.99	143.90	24607.00					Experienced Educator	140		
				Other/Prefer not to say	17								Senior Educator	137		
How familiar are you with AI tools like ChatGPT, Claude, Gemini, or Copilot?	330	3.78	1.067	Male	142	3.94	171.14	24301.00	10133.5	-2.627	.009	.149	Junior Educator	53	2.409	.300
				Female	171	3.66	145.26	24839.50					Experienced Educator	140		
				Other/Prefer not to say	17								Senior Educator	137		
How confident are you that an AI system will give a fair mark for essays or report style work?	330	2.28	1.187	Male	142	2.42	167.39	23769.00	10666.0	-1.918	0.055	.108	Junior Educator	53	5.236	.073
				Female	171	2.19	148.37	25372.00					Experienced Educator	140		
				Other/Prefer not to say	17								Senior Educator	137		

Do you think an AI system will be more or less biased than a human marker?	330	3.05	.961	Male	142	3.13	160.37	22772.50	11662.5	-.641	.521	.036	Junior Educator	53	.820	.664
				Female	171	3.04	154.20	26368.50					Experienced Educator	140		
				Other/Prefer not to say	17								Senior Educator	137		
How confident are you about the privacy and security of your data if assignments are being uploaded to a university-approved AI system?	330	2.48	1.36	Male	142	2.68	167.17	23738.00	10697.0	-1.869	0.062	.106	Junior Educator	53	1.093	.579
				Female	171	2.39	148.56	25403.00					Experienced Educator	140		
				Other/Prefer not to say	17								Senior Educator	137		

Table 3: Gioia data structure of educator resistance to generative artificial intelligence in summative assessment

First-Order Concepts (Representative Quotes)	Second-Order Themes	Aggregate Dimensions
“Another issues is that it can't account for the characteristics of students, e.g., do they have a disability? How does the work compare to their other submissions?” (P1) “Concerns over students for whom English is not their first language and for those with neurodiversity and learning needs” (P5) "students pay to learn from our expertise. If AI is being used for assessing, what are they paying for?" “Students pay to learn from our expertise. If AI is being used for assessing, what are they paying for?” (P112) “students expect their work to be marked by a lecturer expert in the field, which is something an AI is not” (P301)	Loss of the Expert Relationship	The Dehumanisation of Education
“That it will mark the text based on statistically predictable patterns of words and can't demonstrate any common sense or real understanding of anything” (1) “that students use AI to produce their work and we use it to mark it seems pointless waste of time for everyone” (P153) “it is a race to the bottom- AI to set assessments, AI to write assessments, AI to mark assessments- humans get more stupid and we end up like the fat lumps in chairs at the start of WALL-E” (P249)	The Nuance Crisis	
“GDPR and data protection issues, data mining and metadata usage by tech firms as well as environmental issues” (P298) “You are placing student work into a LLM, and therefore using the students work to contribute to the dataset for the LLM. This requires student informed consent.” (P307)	Data and Training Ethics	Erosion of Trust and Accountability
“Another concern is the lack of transparency-AI decisions can feel like a black box, making it hard to explain or justify a grade” (P118) “no accountability; 'the machine' will not explain feedback, nor be available to discuss marks and feedback” (P120)	Transparency and the 'Black Box'	

“...concerns that staff hours would be reduced” (P5) “Academic job losses, risk to university as institution” (P95) “Leaning more into AI uncouples labour from the person (the teacher) and so it pays to deskill them, for the tech companies and the more sinister VCs out there” (P143)	Professional Displacement	Staff Precarity and De-skilling
• "I am a better academic writer and thinker because I mark work... Staff will become deskilled" “I am a better academic writer and thinker *because* I mark work. I find it strange that as a way to "save time" (for example) AI use may be suggested to mark work. (This isn't me saying that I enjoy marking, because I don't, but it's a duty that we have to shoulder” (P157) “Staff will become deskilled at marking, and disconnected to the progress their students are making” (P188) “I am also worried that human educators may risk degrading their marking abilities if they outsource it too much to AI” (P244)	Loss of Pedagogical Skills	

Figures

Figure 1: Preferred Marking Models among UK Educators

