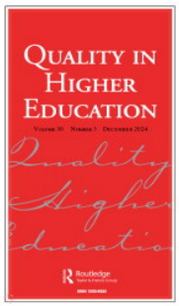

Quality in Higher Education

ISSN: 1353-8322 (Print) 1470-1081 (Online) Journal homepage: www.tandfonline.com/journals/cqhe20

Widespread use of summative online unsupervised remote (SOUR) examinations in UK higher education: ethical and quality assurance implications

Philip M. Newton & Michael J. Draper

To cite this article: Philip M. Newton & Michael J. Draper (15 Jul 2025): Widespread use of summative online unsupervised remote (SOUR) examinations in UK higher education: ethical and quality assurance implications, Quality in Higher Education, DOI: 10.1080/13538322.2025.2521174

To link to this article: https://doi.org/10.1080/13538322.2025.2521174

Published online: 15 Jul 2025.

QUALITY IN HIGHER EDUCATION
https://doi.org/10.1080/13538322.2025.2521174

Routledge
Taylor & Francis Group

Widespread use of summative online unsupervised remote (SOUR) examinations in UK higher education: ethical and quality assurance implications

Philip M. Newton and Michael J. Draper

Swansea University, Singleton Park Campus, Swansea, Wales, UK

ABSTRACT

The use of summative online unsupervised remote (SOUR) examinations is associated with high levels of cheating, which increased further during the COVID lockdown. New generative AI chatbots have added further concern, since they can answer exam questions to a very high level. However, it is currently unclear how common these unsupervised examinations are. A Freedom of Information request methodology was used to determine whether universities in the United Kingdom used SOUR examinations in the 23/24 academic year. Extensive use of these examinations was found with minimal plans to reduce them for future years. Policies for the security of SOUR examinations often put students in the paradoxical position of being required to work under 'examination conditions' remotely, but with no attempt by the university to administer those conditions. This situation raises questions over the validity of SOUR examinations as an assessment format and the quality assurance of degree awards that include these assessments.

KEYWORDS

ChatGPT; academic integrity; validity; quality assurance; cheating

Introduction

In many countries, higher education operates on an outcomes-based approach wherein the conferring of a degree award requires the learner to demonstrate achievement of those outcomes. Quality assurance regulations therefore require higher education providers to provide reasonable evidence-based guarantees about the security and integrity of their assessments, since assessment results are the basis for certifying that a learner has achieved those outcomes. In the UK, the Quality Assurance Agency (QAA) developed the

CONTACT Philip M. Newton p.newton@swansea.ac.uk Swansea University, Singleton Park Campus, Swansea SA2 8PP, Wales, UK.

Quality Code for Higher Education, in which their expectation for standards and practices in relation to assessment stated:

> Assessment….ultimately, it determines whether each student has achieved their course's learning outcomes and allows the awarding body to ensure that appropriate standards are being applied rigorously (QAA, 2018).

Cheating undermines the validity of assessment; if a student has achieved a grade via cheating, then the assessment was not a valid measure of their learning, and so any degree awarded on the basis of these invalid assessments is itself compromised (Dawson *et al.*, 2024). It is therefore recognised by the QAA and similar quality assurance bodies in other countries that assessment should be designed to minimise opportunities for students to commit academic misconduct, specifically that students do not obtain credit or awards through any form of unacceptable academic practice relating to assessment (including plagiarism, cheating in examinations, contract cheating, collusion and impersonation) (QAA, 2018).

Summative online unsupervised remote (SOUR) examinations were a forced necessity of the rapid unplanned transition to online assessment during COVID-19 lockdowns. Examinations that had previously been supervised (i.e., invigilated or proctored) and in-person, on-campus, had to be taken online, remotely, normally at home and with limited time to implement remote supervision methods (Raman *et al.*, 2021). This rapid transition to online assessment was associated with a large increase in the numbers of students who admitted to cheating in online examinations, with students reporting that they were cheating in large part simply because the opportunity was now there for them to do so (Newton & Essex, 2023). Lockdowns have passed, and universities are able to move back to in-person assessment, but there have been calls to retain online unsupervised examinations (Chan & Ahn, 2023) despite their validity being undermined by this widespread cheating (Newton, 2023b).

The quality assurance risks from SOUR examinations are well documented and include cheating by plagiarism, collusion and commissioning. There is evidence of students taking advantage of online cheating services when sitting examinations taken under remote conditions (QAA, 2022a). For example, a 2021 study found that students were posting their online exam questions on sites like Chegg, in order to receive rapid answers from third parties, with Chegg acting as an intermediary for this transaction (Lancaster & Cotarlan, 2021). The Australian Tertiary Education Quality and Standards Agency (TEQSA) has initiated legal proceedings against Chegg under national laws designed to prevent quality assurance from being undermined by this exam contract cheating (TEQSA, 2024), clearly indicating the threat to quality assurance from online exam cheating services.

These issues predate the development of novel generative artificial intelligence (GenAI) tools such as ChatGPT. An abundance of evidence shows that GenAI tools can achieve high scores in traditional online examinations and that the performance of these tools is dramatically enhanced when using versions that utilise more powerful underlying large language models (LLMs). For example, early versions of ChatGPT made available to the public in November 2022 were based on the underlying LLM GPT-3 or GPT 3.5. A review of 53 research papers showed that these versions of ChatGPT scored an average of 54% in examinations based on multiple-choice questions, whereas an updated version running GPT-4, released in May 2023, scored an average of 75% (Newton & Xiromeriti, 2024), and a later update, running GPT-4o, scored 94% on the United Kingdom medical licencing assessment (Newton *et al.*, 2025). The improved performance is partly due to the ability of these latest models to search the internet in real-time, process images and undertake calculations through automated integration with other tools such as Python. To test the potential threat from ChatGPT to the validity of online examinations, Scarfe and colleagues undertook what they termed a quality assurance exercise, where they used ChatGPT to generate responses to exam questions, including free-text essay responses, and spiked them into a set of responses generated by students. Markers were not told about the ChatGPT responses and thus blindly marked all exam papers as if they were answers generated by students. The ChatGPT responses scored higher marks on average than those generated by students and were almost never flagged as problematic (Scarfe *et al.*, 2024).

Quality assurance agencies have attempted to address this challenge through rapid guidance to providers. The QAA acknowledged in February 2024 that 'the advent of Gen AI in HE is on the cusp of dramatically reshaping the landscape of assessment and feedback in particular', but the sector has struggled to identify clear policy guidance; for example, the QAA highlighted that 'submissions for assessment that consist only of substantially unmodified output from Generative Artificial Intelligence software may be considered not to be plagiarised using traditional definitions. However, this is very poor academic practice, as it does not represent the student's own work' (QAA, 2023). Student use of GenAI tools for misconduct would not seem to fit traditional definitions of plagiarism (since the content output is original) or contract cheating (since no third-party human being is involved in directly generating the content) and so some providers have sought to introduce policies that require authorised and declared use of generative AI in relation to submissions for assessment. However, it seems unlikely that an assessment will be a meaningful certification of learning if a student uses GenAI to generate it and then simply acknowledges that fact.

One option to address this challenge is to use technology to monitor student behaviour in examinations through online invigilation (proctoring).

There is some evidence that online invigilation reduces cheating (Newton & Essex, 2023; Bierstaker *et al.*, 2024), but it is also, currently, associated with a poor student experience with concerns about privacy, efficacy, cost and more (Marano *et al.*, 2024). Another option is to monitor student responses generated in the exam for their originality and determine whether they have been generated with GenAI using 'AI-detection software'. These approaches are only worth considering for examinations that are based on free text responses and would not prevent cheating behaviours such as collusion or accessing other unauthorised sources. Even then the effectiveness of GenAI detection tools is modest and easily reduced (Perkins *et al.*, 2024), plus the lack of any way of independently verifying the output creates a risk of falsely accusing students of misconduct (Gorichanaz, 2023).

Almost all UK university students are already regularly using GenAI tools in their assessments (Freeman, 2025), but it is unclear whether they are using it to cheat. Media stories report that 'student AI cheating cases have soared at UK universities' (Grove, 2024a) and that academics 'despair as ChatGPT-written essays swamp marking season' (Grove, 2024b), but there are currently few empirical data on this issue, perhaps in part due to uncertainty over what 'cheating' actually means. A UK study simply asked undergraduates whether they had 'cheated using ChatGPT' and found that 22% of them said they had (Newton, 2025), with similar results found in Vietnam (Nguyen & Goto, 2024).

Thus, the potential quality assurance implication of the continued use of SOUR examinations is to take a situation where cheating was already very high, but now students have access to a tool (GenAI) that can help them achieve a very high grade without doing any original work themselves, in a way that is difficult to meaningfully detect and which does not require them to collude with any others in doing so. This would seem to negate the basic validity of SOUR examinations as assessments and so undermine the credibility of any degree awarded where these examinations were used to certify learning. Such a circumstance would undermine the basic function of higher education and further contribute to 'grade inflation', which has been a growing concern in UK higher education. A 2022 analysis of degree classifications over time, reviewing and comparing degree classifications between 2010-11 and 2020-21, found that the percentage of students achieving the highest grade had more than doubled in 10 years (Office for Students, 2022).

However, it is currently unclear to what extent universities are using SOUR examinations and what measures are being taken to ensure the security and integrity of those examinations. Here, then, the United Kingdom Freedom of Information (FOI) request process was used, whereby UK universities were asked whether they were using summative online unproctored remote examinations. Universities were also asked whether they plan to carry on using these examinations and whether they use any monitoring systems, such as online proctoring, for the supervision of these examinations. Finally,

universities were asked to share with us their policies for maintaining the security and integrity of these assessments.

Methodology

Epistemological underpinnings

Pragmatism was used as a research paradigm. Pragmatism is an approach that prioritises asking research questions whose answers will be practically useful, particularly for evidence-based policy making (Newton *et al.*, 2020). The choice of methodology reflects the best way of answering the question, rather than any other ontological perspective (James, 1907).

Procedure

A list of FOI request contacts was obtained for 133 UK universities from the 'FOI Directory' under a Creative Commons Licence (Burgess, 2024). On 11 June 2024, all these universities were written to with the questions below.

1. In 23/24, approximately how many summative examinations were both remote and online, i.e., no invigilators were physically present?
2. In 23/24, did the university use any remote invigilating or lockdown browser systems for any of these summative online remote examinations, e.g., ProctorU, Examinationsoft, Respondus, etc?
3. If the university did use a remote invigilating/lockdown system, approximately what percentage of summative remote online examinations were monitored in this way?
4. Is the university going to change the number of summative remote online examinations that are invigilated (or not) in 24/25? If so, please estimate the change and direction (e.g., X% fewer online summative unproctored remote examinations).
5. Could you please provide links to, or copies of, any university policies or guidance that were designed to ensure the security and integrity of online unproctored remote examinations in 23/24?

Universities were not given any definition of an examination: the current QAA glossary does not contain a definition of an examination (QAA, 2022b), while dictionary definitions are very broad; for example, the Collins English Dictionary defines an examination as 'An examination is a formal test that you take to show your knowledge or ability in a particular subject, or to obtain a qualification' (Collins English Dictionary, 2024), which could refer to most assessment formats in higher education. Thus an assumed understanding had to be relied on. However, some universities responded that they did not operate formal examinations online and unsupervised but that they ran

similar summative assessments using this format, for example, 'online tests' or 'timed assessments'. For the purposes of this research, these pseudo-examinations were included in the analysis since they are equally vulnerable to misconduct as described in the introduction and thus a threat to assessment validity. Where a university responded that they did not use *any* summative online remote examinations, then the relevant university website was searched for items that might reasonably be considered pseudo-examinations but that a university might not consider a formal examination. Specifically, the following terms were searched for: 'online examination', 'online exam', 'digital examination', 'digital exam', 'take-home', 'Blackboard test', 'Canvas test', 'Moodle test', 'timed assessment' and 'timed test'. For these universities, the academic misconduct policy and any assessment regulations were manually reviewed, as well as (where found) any regulatory or staff development pages that described the types of assessments in use at the university. Where these searches suggested that a university was indeed using remote online unsupervised pseudo-examinations, then the university was written to with the relevant evidence and asked to reconsider. If a university initially responded that they *were* using online remote examinations, then no further searching was undertaken.

Policy analysis: 'Forbidden Fruit' clause and generative AI

The supplied policies were analysed to determine whether it was clear to students what the expectations were of them regarding summative online unsupervised remote examinations. To test this, a 'forbidden fruit' clause was specifically searched for, which was defined as a regulation that banned students from using methods that could be used to easily obtain a better mark (see below) but where the university did not attempt to prevent them from doing so, essentially requiring students to work under some form of exam conditions during online unsupervised examinations. Specifically banning one or both of:

1. Communication/working with other people (including students, friends/family, etc)
2. Use of external sources, for example, the internet, GenAI, books

Thus an open-book exam would still be considered as being conducted under 'forbidden fruit' conditions if it banned students from contacting other people. Where universities reported that they were using SOUR examinations and supplied a policy in response to Q5, the supplied policy was, or links from the policy were, for regulations that put students in a 'forbidden fruit' position. This judgement was based solely on the policy supplied in response to Q5 or any policies linked to from the supplied policy. If an

answer could not be determined from these materials, then it was recorded as 'unclear', for example, if the policy did not state what was expected of students. The policies were also searched for any mention of GenAI tools like ChatGPT specifically in relation to online examinations. If the policy covered these, then this was recorded as yes/no, and the specific regulation was noted. If the supplied policy addressed the use of these tools elsewhere, for example, in coursework writing, then this was recorded as 'no' here.

Results

123/133 universities responded to the FOI request within four months, although 3 of these were refusals to supply the data based on cost or privacy. Of the 10 that did not respond, 6 sent an initial acknowledgement but no further response, while 4 did not reply at all. The results are presented by question, although some universities provided responses to some but not all questions; hence, the number responding is different for each. A summary of all responses is given in table form in supplementary data S1, while the raw text of the full set of responses is available upon request.

Question 1. *In 23/24, approximately how many summative examinations were both remote and online, i.e., no invigilators were physically present?*

120 universities provided some sort of response. One answered 'none' to Q1 but did not respond to a further query about the use of pseudo-examinations. Of the 119 remaining, 93 (78.2%) used online remote examinations in some form, while 26 (21.8%) reported using none.

Question 2. *In 23/24, did the university use any remote invigilating or lockdown browser systems for any of these summative online remote examinations, e.g., ProctorU, Examinationsoft, Respondus, etc?*

Of the 93 who used online remote examinations, two refused to answer all further questions based on cost or commercial sensitivity, while another responded that the data for this question were not available. For the remaining 90 universities, 67 (74.4%) did not use any proctoring, while the remaining 23 (25.6%) used some form of proctoring or lockdown browser for some (though not necessarily all) examinations.

Question 3. *If the university did use a remote invigilating/lockdown system, approximately what percentage of summative remote online examinations were monitored in this way?*

A further university refused this request based on cost. Of the remaining 89, nine (10.1%) reported that they proctored all their online remote summative examinations, while 61 (68.5%) proctored none of them. Sixty-eight universities reported the number of unproctored online examinations used, which totalled 16,744 with an average of 246 examinations per university and ranging from 3 to 1482.

Question 4. *Is the university going to change the number of summative remote online examinations that are invigilated (or not) in 24/25? If so, please estimate the change and direction (e.g., X% fewer online summative unproctored remote examinations).*

Only 67 universities gave data in response to this question, of which 47 (70.1%) said they had no plans to reduce the number. Thirteen (19.4%) indicated that there were plans to reduce the number, of which two (3%) planned to eliminate them completely. Seven universities responded that nothing had yet been decided. No universities responded that they were planning to increase the number of summative remote online examinations.

Question 5. *Could you please provide links to, or copies of, any university policies or guidance that were designed to ensure the security and integrity of online unproctored remote examinations in 23/24?*

If a university did not use online remote examinations, or they used them but they were all supervised, then this was recorded as 'N/A'. If they used these examinations but responded that they did not have a specific policy or that 'no information was held', then this was recorded as 'None' for a policy. If the university supplied a policy that was erroneous (e.g., behind a secure login), then this was recorded as 'None Provided'.

Of the universities that used unproctored online remote examinations, 60 provided a policy, 15 did not have a policy and 2 provided a response but not a policy (e.g., responded that the answer was 'N/A' even though they offered online remote unproctored examinations (this was recorded as 'None')).

Of the 60 supplied policies, 40 (66.7%) did not mention GenAI, 17 (28.3%) did and 3 (5%) were unclear. In these unclear cases, the policy did not specifically refer to SOUR examinations (e.g., it was about how to set up the examinations or was a link to the entirety of the academic regulations). A forbidden fruit clause was found in 39 (65%) of policies, 19 (31.7%) were unclear, while 2 did not contain such a clause. Of these two, one was for a policy where all examinations were open book, although the policy did not mention GenAI. The other required that students cite GenAI in the same way that private correspondence is cited.

Implications

These findings suggest that the use of SOUR examinations was widespread in UK higher education in the 23/24 academic year, with limited consideration of GenAI. SOUR examinations show little validity as a form of assessment. Cheating in online examinations was already very high before GenAI (Newton & Essex, 2023). New GenAI tools make it even easier to cheat on these examinations, and this cheating would allow students to achieve a high grade with limited means by which this could be detected (Newton &

Xiromeriti, 2024; Scarfe *et al.*, 2024). Even if a SOUR exam was open-book and the use of GenAI was explicitly permitted, then the performance of GenAI tools means that it is unlikely that the exam would be a valid certification of the learning of a student. In any case, only one university appeared to use SOUR examinations in this way.

An additional challenge associated with the use of these assessments is the 'forbidden fruit' position that students are placed in when the policy for the security and integrity of SOUR examinations requires that students should work under examination conditions, but these conditions are not enforced. Any student assessed under these conditions is placed in a no-win situation. Either they follow the rules and do not cheat, but they know that a majority of their peers will cheat (Newton & Essex, 2023), and so if they do not cheat, then they will get lower grades than their peers, which is seen as basically unfair to those who do not cheat (Lee & Fanguy, 2022), and may lead to them being disadvantaged, for example, in the job market. Or they cheat and thus have to carry that emotional baggage with them, along with the perceived risk of being caught and the anxiety it creates. This unethical 'no-win' situation is created knowingly and deliberately by universities, for reasons that are not explicitly stated but seem likely due to cost and perhaps a perception that online exams are more flexible and inclusive. This situation seems likely to further undermine the validity of SOUR examinations as assessments and to have knock-on consequences. For example, it may indicate to students that the institution does not take academic integrity seriously, which can diminish their respect for the academic regulations and the institution itself, particularly where it is obvious that other students can easily break these unenforced rules and so obtain a higher mark for themselves.

A 2024 report estimated one in three UK students were using GenAI weekly as part of their studies, with higher use amongst those from higher socioeconomic backgrounds and by international students (Freeman, 2024). The same organisation conducted a similar study in 2025 and found that this figure had increased dramatically, with 88% of UK students reporting using GenAI in their assessments (Freeman, 2025). Studies in the US and UK have found that students use these tools for academic writing tasks in a way that blurs the already fuzzy line between cheating and acceptable academic practice, for example, to 'get started' or to 'generate drafts' (Golding *et al.*, 2024; Newton, 2025).

Reservations

These results are highly likely to be an underestimation of the use of SOUR examinations in UK higher education. The FOI request was sent at the end of Semester 2 for most UK universities, meaning that any Semester 3 and re-sit examinations would not have been taken. Many universities responded that they could only use the FOI response to collate examinations that were

centrally administered (i.e., not run by departments), and some attempted to exclude pseudo-examinations based on the supposed amount of time taken to complete the assessment. Some of these have been identified in this research, but almost certainly not all of them. The findings also do not allow any insight into the overall assessment profile of students at the responding institutions; for example, the validity of a degree that is assessed entirely by SOUR exams will be compromised more than a degree for which these exams form a small component. This point is itself a limitation of the methodology: the use of the FOI process was made necessary since there is no obvious national data repository that would allow for the evaluation of the assessments used across the UK HE sector. This information would be useful to other stakeholders, including students and employers, and so it seems reasonable to propose that the QA process collect and publish such data. Finally, the policy analysis focused specifically on the research questions and did not consider the wider approaches taken to secure assessment integrity.

Recommendations

It is important to consider that it is not any/all online examinations that are necessarily problematic, but it is the specific combination of all four elements of the SOUR approach that results in an assessment format that appears to lack basic validity. Just addressing just one of the four SOUR elements could result in an assessment with some positive features. For example, online unsupervised and remote examinations might still be useful if they are formative rather than summative. Keeping the summative element might be possible if the examinations are in-person or invigilated. Some of the literature published in the immediate aftermath of the COVID-19 lockdowns was positive and enthusiastic about the opportunities offered by online assessment. For example, the UK Office for Students published a report that proposed many benefits of the COVID-enforced switch to online assessment, including increased inclusivity, offering flexibility and accessibility to students who might face barriers with traditional in-person examinations, including students in different time zones or students with caring obligations (Barber *et al.*, 2021). Students who use assistive technologies like screen readers or voice dictation software could complete assessments in a familiar environment with their own tools. Online assessment might also allow students to take their examinations in a place, and possibly at a time, that suits them best, accommodating those with varying schedules, caregiving responsibilities, those who may be in different time zones. It may also be cheaper for students to take examinations at home, and the familiar environment might reduce test anxiety for some students. However, the evidence base for these proposed benefits is mixed and is complicated by a potential 'digital divide', which presents as inequity in access to reliable internet connections and/or

the necessary technology for proctoring, along with the technical skills to navigate online assessments effectively, etc. These issues negatively affect the quality assurance and student/staff experience of online examinations (Hosseini *et al.*, 2021; Vellanki *et al.*, 2023) and will likely result in disproportionate disadvantage for students and education providers in lower socioeconomic situations and rural areas. Some studies show that students with caring responsibilities are less positive towards online examinations (Aristeidou *et al.*, 2024). Even with the positive approach presented in their report, Barber *et al.* noted that less than a third of students want to be assessed solely online (Barber *et al.*, 2021). It thus might be most appropriate to recommend that any benefits of online exams be considered as part of a wider portfolio of assessment methods, combining multiple approaches.

Thus, higher education is in a position where there is widespread use of an assessment format that lacks basic validity and for which even peripheral benefits are questionable. What might be done about this? Some universities have rapidly switched to in-person examination as a result of concerns about AI, but this has angered some students, partly due to a perceived lack of warning (Scott & Bonar, 2024). One constructive approach is to ensure that staff are trained and supported to develop 'assessment literacy'. There have been many decades of repeated government reports that recommend that staff in the UK higher education sector be formally trained how to teach and assess (Kernohan, 2017), and yet even external examiners report that they lack assessment literacy (Medland, 2019), and the teacher training programmes used by universities have focused on the scholarship of learning and teaching rather than these practical skills (Foxe *et al.*, 2017). Textbooks aimed at UK academic staff do not address academic misconduct (Ransome & Newton, 2018), and the apparent lack of assessment literacy by staff at UK universities seems likely to undermine any enhancement-led approaches to incorporating GenAI into (and out of) assessment (Newton, 2023a).

Assessment redesign is not sufficient, even if 'authentic assessment' (assessments that are more reflective of the ways in which students will actually use the knowledge they learn) is employed (Ellis *et al.*, 2020). However, authentic assessment might provide a disincentive to cheating as well as improving the student assessment experience by focusing on what students can do, rather than what they can write about, and by signposting the relevance of the learning to future careers. Authentic assessments for many careers might incorporate practical examinations or oral assessments and thus may be more resilient to cheating.

In addition to assessment redesign, it seems reasonable to propose that the validity concerns of SOUR examinations should lead to a regulatory requirement for them to be abandoned. The QAA makes judgments about providers, with a positive judgement meaning the provider has robust arrangements for securing academic standards, managing academic quality

and enhancing the student experience. Negative judgements in those areas or areas for improvement and/or recommendations are subject to a formal follow-up by QAA, which involves the production and monitoring of an action plan or, more usually, a follow-up report. These reports are made publicly available in some jurisdictions, but it was not possible to identify any in which institutions were required to address concerns over academic integrity or misconduct. This included 77 'Enhancement led Institutional Reviews' in Scotland and Wales, within which there were 16 follow-up reports.

Banned or not, the definitions of (un)acceptable academic practices used by quality assurance agencies will almost certainly need to be quickly redefined, particularly since most universities indicated that they plan to continue using SOUR examinations. Indeed, a basic definition of an examination could assist providers in addressing some of these challenges.

In summary, this study has found that in UK higher education in 2023/24, there is widespread use of an assessment format (summative, online, unsupervised remote examinations) that does not offer a reasonable guarantee as an assessment of student learning. This appears to raise challenging questions about the quality assurance of degree awards taht include these assessments and the need for urgent review of quality assurance of higher education in the era of GenAI. It is therefore recommended that quality assurance agencies explicitly require that the invalid SOUR approach to exams be abandoned and that this issue be considered when making quality judgements about higher education providers.

Acknowledgements

We are grateful to those University administrators who replied to our FOI requests, especially those who addressed our follow-up queries.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Philip M. Newton http://orcid.org/0000-0002-5272-7979
Michael J. Draper http://orcid.org/0000-0003-1272-8122

References

Aristeidou, M., Cross, S., Rossade, K.D., Wood, C., Rees, T. & Paci, P., 2024, 'Online exams in higher education: Exploring distance learning students' acceptance and satisfaction', *Journal of Computer Assisted Learning*, 40(1), pp. 342–59.

Barber, M., Bird, L., Fleming, J., Titterington-Giles, E., Edwards, E. & Leyland, C., 2021, *Gravity Assist: Propelling higher education towards a brighter future – Office for Students* (Bristol, Office for Students).

Bierstaker, J., Brink, W.D., Khatoon, S. & Thorne, L., 2024, 'Academic fraud and remote evaluation of accounting students: An application of the fraud triangle', *Journal of Business Ethics*, 195(2), pp. 425–47.

Burgess, M., 2024, 'FOI List of Emails for UK Universities', FOI Directory. Available at http://www.foi.directory/uk-universities-list-foi-emails/ (accessed 11 June 2024).

Chan, J.C.K. & Ahn, D., 2023, 'Unproctored online exams provide meaningful assessment of student learning', *Proceedings of the National Academy of Sciences*, 120(31), p. e2302020120.

Collins English Dictionary, 2024, 'EXAMINATION definition and meaning'. Available at https://www.collinsdictionary.com/dictionary/english/examination (accessed 2 December 2024).

Dawson, P., Bearman, M., Dollinger, M. & Boud, D., 2024, 'Validity matters more than cheating', *Assessment & Evaluation in Higher Education*, 49(7), pp. 1005–16.

Ellis, C., Van Haeringen, K., Harper, R., Bretag, T., Zucker, I., McBride, S., Rozenberg, P., Newton, P. & Saddiqui, S., 2020, 'Does authentic assessment assure academic integrity? Evidence from contract cheating data', *Higher Education Research & Development*, 39(3), pp. 454–69.

Foxe, J.P., Frake-Mistak, M. & Popovic, C., 2017, 'The instructional skills workshop: A missed opportunity in the UK?', *Innovations in Education and Teaching International*, 54(2), pp. 135–42.

Freeman, J., 2024, 'New HEPI Policy Note finds more than half of students have used generative AI for help on assessments – but only 5% likely to be using AI to cheat'. Available at https://www.hepi.ac.uk/2024/02/01/new-hepi-policy-note-finds-more-than-half-of-students-have-used-generative-ai-for-help-on-assessments-but-only-5-likely-to-be-using-ai-to-cheat/ (accessed 14 September 2024).

Freeman, J., 2025, *Student Generative AI Survey 2025*. Available at https://www.hepi.ac.uk/2025/02/26/student-generative-ai-survey-2025/ (accessed 11 March 2025).

Golding, J.M., Lippert, A., Neuschatz, J.S., Salomon, I. & Burke, K., 2024, 'Generative AI and college students: Use and perceptions', *Teaching of Psychology*, 2024, p. e00986283241280350.

Gorichanaz, T., 2023, 'Accused: How students respond to allegations of using ChatGPT on assessments', *Learning: Research and Practice*, 9(2), pp. 183–96.

Grove, J., 2024a, 'Student AI cheating cases soar at UK universities', *Times Higher Education* (*THE*). Available at: https://www.timeshighereducation.com/news/student-ai-cheating-cases-soar-uk-universities (accessed 12 November 2024).

Grove, J., 2024b. 'Academics despair as ChatGPT-written essays swamp marking season', *Times Higher Education*, 1 November 2024.

Hosseini, M.M., Egodawatte, G. & Ruzgar, N.S., 2021, 'Online assessment in a business department during COVID-19: Challenges and practices', *The International Journal of Management Education*, 19(3), pp. 100556.

James, W. 1907. *Pragmatism. A new name for some old ways of thinking* (New York, NY, Dover Publications Inc).

Kernohan, D., 2017, 'The history of efforts to improve university teaching', Wonkhe. Available at https://wonkhe.com/blogs/analysis-history-of-efforts-to-improve-university-teaching/ (accessed 1 May 2025).

Lancaster, T. & Cotarlan, C., 2021, 'Contract cheating by STEM students through a file sharing website: A Covid-19 pandemic perspective', *International Journal for Educational Integrity*, 17(1), p. 3. Available at (accessed 12 November 2024).

Lee, K. & Fanguy, M., 2022, 'Online exam proctoring technologies: Educational innovation or deterioration?', *British Journal of Educational Technology*, 53(3), pp. 475–90.

Marano, E., Newton, P.M., Birch, Z., Croombs, M., Gilbert, C. & Draper, M.J., 2024, 'What is the student experience of remote proctoring? A pragmatic scoping review', *Higher Education Quarterly*, 78(3), pp. 1031–47.

Medland, E., 2019, '"I'm an assessment illiterate": Towards a shared discourse of assessment literacy for external examiners', *Assessment & Evaluation in Higher Education*, 44(4), pp. 565–80.

Newton, P. and Xiromeriti, M., 2024, 'ChatGPT performance on multiple choice question examinations in higher education. A pragmatic scoping review', *Assessment & Evaluation in Higher Education*, 49(6), pp. 781–98.

Newton, P.M., 2023a, 'Making higher education assessment literate', *Wonkhe*. Available at: https://wonkhe.com/blogs/making-higher-education-assessment-literate/ (accessed 1 May 2025).

Newton, P.M., 2023b, 'The validity of unproctored online exams is undermined by cheating', *Proceedings of the National Academy of Sciences*, 120(41), p. e2312978120. Available at: (accessed 12 November 2024).

Newton, P.M., 2025. 'How vulnerable are UK universities to cheating with new GenAI tools? A pragmatic risk analysis', OSF. Available at: (accessed 12 November 2024).

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, D. & Hughes, S., 2025, 'Can ChatGPT-4o Really Pass Medical Science Exams? A Pragmatic Analysis Using Novel Questions', *Medical Science Educator*.

Newton, P.M., Da Silva, A. & Berry, S., 2020, 'The case for pragmatic evidence-based higher education: A useful way forward?', *Frontiers in Education*, 5, 157.

Newton, P.M. & Essex, K., 2023, 'How common is cheating in online exams and did it increase during the COVID-19 pandemic? A systematic review', *Journal of Academic Ethics*, 22, pp. 323–343.

Nguyen, H.M. & Goto, D., 2024, 'Unmasking academic cheating behavior in the artificial intelligence era: Evidence from Vietnamese undergraduates', *Education and Information Technologies*, 29, pp. 15999–16025.

Office for Students, 2022, *Analysis of degree classifications over time: Changes in graduate attainment from 2010-11 to 2020-21 – Office for Students* Office for Students. Available at https://www.officeforstudents.org.uk/media/7054/ofs-202222.pdf (accessed 12 November 2024).

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), p. 53 (accessed: 03 June 2025).

The Quality Assurance Agency for Higher Education (QAA), 2018, *UK Quality Code, Advice and Guidance: Assessment* (Gloucester, QAA). Available at: https://www.qaa.ac.uk/docs/qaa/quality-code/advice-and-guidance-assessment.pdf?sfvrsn=ca29c181_5 (accessed: 03 June 2025).

QAA, 2022a, *Contracting to Cheat in Higher Education* (3rd edition). (Gloucester, QAA). Available at: https://www.qaa.ac.uk/docs/qaa/guidance/contracting-to-cheat-in-higher-education-third-edition.pdf (accessed 03 June 2025).

QAA, 2022b, *Glossary* (Gloucester, QAA). Available at: https://www.qaa.ac.uk/glossary (accessed: 03 June 2025).

QAA, 2023, *Maintaining quality and standards in the ChatGPT era: QAA advice on the opportunities and challenges posed by Generative Artificial Intelligence* (Gloucester, QAA).

Available at: https://www.qaa.ac.uk/docs/qaa/members/maintaining-quality-and-standards-in-the-chatgpt-era.pdf?sfvrsn=2408aa81_10 (accessed: 03 June 2025).

Raman, R., Vachharajani, H. & Nedungadi, P., 2021, 'Adoption of online proctored examinations by university students during COVID-19: Innovation diffusion study', *Education and Information Technologies*, 26(6), pp. 7339–58.

Ransome, J. & Newton, P.M., 2018, 'Are we educating educators about academic integrity? A study of UK higher education textbooks', *Assessment & Evaluation in Higher Education*, 43(1), pp. 126–37.

Scarfe, P., Watcham, K., Clarke, A. & Roesch, E., 2024, 'A real-world test of artificial intelligence infiltration of a university examinations system: A "Turing Test" case study', *PLOS one*, 19(6), p. e0305354.

Scott, K. & Bonar, M., 2024, 'Glasgow University students' anger over reintroduction of in-person exams', *BBC News*, 28 February 2024. Available at https://www.bbc.com/news/uk-scotland-glasgow-west-68380264 (accessed 27 August 2024).

The Tertiary Education Quality and Standards Agency (TEQSA), 2024, 'TEQSA commences legal proceedings against Chegg'. Available at https://www.teqsa.gov.au/about-us/news-and-events/latest-news/teqsa-commences-legal-proceedings-against-chegg (accessed 7 November 2024).

Vellanki, S.S., Mond, S. & Khan, Z.K., 2023, 'Promoting academic integrity in remote/online assessment – EFL Teachers' Perspectives', *TESL-EJ*, 26(4), p. 7. https://eric.ed.gov/?id=EJ1383953 (accessed, 22 December 2024).