

Moving Face-to-Face Exams to Student-Centric Online TRAs: Design, Evidence and Partnership in a Second Year Immunology Module

Ana P. Costa-Pereira¹ , Manuela Mura² , Robert E. White³ 

¹Department of Surgery & Cancer, Imperial College London, UK, ²Faculty Education Office, Imperial College London, UK, ³Department of Infectious Diseases, Faculty of Medicine, Imperial College London, UK.

How to cite: Costa-Pereira, A.P.; Mura, M.; White, R.E. (2026). Moving Face-to-Face Exams to Student-Centric Online TRAs: Design, Evidence and Partnership in a Second Year Immunology Module. In: 12th International Higher Education Advances (HEAd'26). Valencia, 15-18 June 2026. <https://doi.org/10.4995/HEAd26.2026.21737>

Abstract

During the pandemic, the Immunology & Inflammation (IMI) module transitioned its 3-hour closed-book exam to an online time-restricted assessment (TRA) without altering structure or timing. This paper presents a local evaluation of that transition, informed by anonymised student perspectives gathered through student-led focus group discussions conducted as part of routine module review and enhancement. Feedback addressed assessment design, usability, cognitive load, integrity, and inclusivity. Students identified challenges related to navigation and time pressure, while also valuing the use of authentic questions that rewarded conceptual understanding. As part of this evaluative process, students proposed and co-developed practical refinements, including improved integration of figures within questions, access to practice environments, and adding accessibility features. These reflections suggest that student-centred TRAs can maintain academic standards and enhance learning when design, layout, timing, and accessibility are iteratively optimised through collaborative educational evaluation.

Keywords: Online time-restricted assessment | higher-order thinking | authentic pedagogy | academic integrity | student-staff partnership | UDL

1. Introduction

The COVID-19 pandemic forced universities to reconfigure assessment at speed, with the UK Quality Code's emphasis on maintaining standards, consulting students, and evidencing decisions providing a crucial anchor during rapid change (QAA, 2020). In this context, an Immunology & Inflammation (IMI) second year undergraduate module, moved its face-to-face

exam to a time-restricted online assessment (TRA) that preserved the original blueprint and cognitive demands. The module's pedagogy blends bespoke online eModules with face-to-face active learning sessions, prioritising conceptual understanding over rote learning. The summative exam was therefore designed to probe understanding rather than recall, notably through interpretation (as opposed to data-handling) of primary literature figures.

Sector discussions during and after COVID have stressed that emergency pivots should evolve into evidence-informed, student-centred redesigns, favouring authentic tasks and explicit attention to equity and accessibility. Authentic assessment frameworks call for tasks that mirror disciplinary practice, require judgement/ creativity, and which allow rehearsal and feedback, rather than one-shot recall under artificial constraints (Zhan et al. 2025). At the same time, the online exam integrity debate has evolved: while some studies find more cheating in unproctored contexts (Fendler et al. 2024), others show unsupervised online exams can validly rank-order learning comparable to in-person proctored tests, especially when the assessment is well-designed (JISC 2021, Chan & Ahn 2023).

We thus sought student feedback and their perceptions to support our reflection on the pragmatic approach we took, focussing on how it served students during COVID and whether the findings proposed examples of good practice beyond this specific assessment, which might also be helpful to other disciplines and contexts. This paper presents (1) the design logic for our TRA, (2) the implementation choices that privileged integrity, inclusivity, and student experience, and crucially (3) feedback from student focus-group discussions on usability, cognitive load, and perceived fairness, using those insights to propose design heuristics for student-centric TRAs. The resulting approach aligns with literature on students-as-partners in assessment (Bovill 2019) and universal design for learning (UDL)-informed blueprints for accessibility (CAST).

2. Methods

2.1. Teaching context

IMI is a second-year elective in a biomedical sciences programme in a UK university, focused on STEM education and taken by approximately 70 to 120 students. The module is taught using a blended approach with ten online eModules for self-paced factual/theoretical groundwork, each followed by a corresponding ten face-to-face sessions where students work in small groups to contextually apply their learning, and two end-of-term sessions to consolidate learning. The teaching philosophy prioritises conceptual and mechanistic understanding over exhaustive memorisation of immunological markers or cytokine nomenclature.

The original examination included a three-hour, in-person, closed-book assessment comprising three parts: Part A, 20 single best answer (SBA) questions (25% of the total mark; ~30 minutes); Part B, 15 assertion-reason questions (ARQs) (30%; ~40 minutes), and Part C, two multi-part

data interpretation questions drawing on figures adapted from published research papers, re-labelled and edited for clarity (45%, ~110 minutes). All sections were composed and scrutinised by all three module academics, Head of Assessment & Feedback, and External Examiners.

2.2. Online Transfer to a Time-Restricted Assessment (TRA)

With the onset of the COVID pandemic in 2020, all teaching and assessment was delivered using a VLE. We replicated the exam blueprint and duration in the VLE to ensure construct continuity and emphasis on assessing understanding and reasoning over response speed. We de-identified primary-literature figures in Part C by renaming molecules and organisms to reduce direct searchability. We reviewed SBA/ARQ items to confirm they demanded comprehension rather than fact recall. Although this iteration of the exam was open-book, students were advised to treat it as closed-book, reminded of integrity expectations, and warned that searching would likely detract from time better spent reasoning.

2.3. Student Partnership: Focus Group Design

To evaluate students' experience with the online TRA, a focus group was conducted following completion of the assessment. The focus group allowed for participants' brainstorming and the provision of valuable qualitative and contextual feedback for a student-centred improvement cycle. All students enrolled on IMI ($n = 98$) were invited to participate via the institutional VLE and cohort mailing lists, seven students volunteered and were gathered for a single focus group session. The student-led semi-structured discussion was guided to systematically explore five predefined domains: (1) usability and layout (including figure navigation, scrolling, and side-by-side viewing), (2) perceived cognitive load and time pressure, (3) authenticity and alignment with learning outcomes, (4) perceptions of assessment integrity and associated norms, and (5) inclusivity and accessibility (e.g. screen-reader compatibility, font scaling, the positioning/partitioning of data and figures). Anonymous feedback from the discussion was collected using Mentimeter; students also provided a list of recommendations. The focus group contributions were analysed using reflexive thematic analysis to identify codes and construct themes in relation to the evaluation question. The construction of the themes was integrated with the Academics' insights about grade distribution before and after the implementation of the online TRA, and their perspective about the advantages and disadvantages of using the online platform.

3. Evaluation

Overall, grade distributions were similar to the previous academic year, though in this iteration two students failed the exam. Performance tracked a prior oral assessment and ongoing face-to-face engagement during class time; students who were consistently active in conceptual sessions were more likely to do well in the TRA. Item analysis suggested that ARQs were not trivially

searchable. SBA questions were more amenable to cheating (particularly where less synthesis was required) but weaker students still missed seemingly straightforward questions requiring understanding rather than recall. We found no evidence of serious misconduct, though VLE exams at the time could not be scanned systematically for plagiarism; it is conceivable that some students might have colluded via messaging apps, as COVID lockdowns and social distancing were then in place.

Our observations are consistent with emerging evidence that well-designed, unproctored online exams can provide meaningful assessments of learning that is comparable to proctored face-to-face exams (Chan & Ahn 2023); this emergency compromise may thus be more widely applicable to disciplines beyond immunology and biomedical science.

3.1. Student Focus Groups Themes

Students were asked to focus on exam format, length, content and presentation; their feedback and perceptions led to the construction of additional, more refined, themes.

3.1.1. Layout, Navigation and Cognitive Load

Students consistently reported friction in Part C (data interpretation): figures appeared sequentially and separately from the question prompts, compelling excessive scrolling and split attention. Many attempted to use the supplementary PDF provided alongside the exam, but VLE's constraints (including no two windows and no second device access) made side-by-side viewing unfeasible, increasing time pressure and cognitive load, consistent with others' work (Pengelley et al. 2025). Students recommended co-locating each figure panel adjacent to its question stem or providing an embedded pannable gallery within each question.

3.1.2. Authenticity and Perceived Fairness

Students valued that the exam rewarded conceptual reasoning and data interpretation over memorisation. Part C, described as "*real science*", mirrored how students practised in the module and what they expected happened within research. This authenticity was linked to fairness by students who found questions challenging: "*Data interpretation [...] start to get difficult, and it is quite different from what we have in the F2F session. But I think 3 hrs is sufficient.*"

3.1.3. Timing versus Task Design

Students appreciated that time allocations mirrored the face-to-face exam and were not overly tight, but several felt that scrolling and figure toggling consumed working time, especially for those on small screens and, presumably, for students with compromised executive skills. Students suggested pre-exam orientation to the interface (e.g., a "practice assessment" sandbox)

and explicit time allocation per section as stress mitigations, echoing sector guidance that mode shifts require orientation and scaffolding to preserve validity (QAA, 2020).

3.1.4. Integrity Perceptions

Participants offered that “*Googling won’t help*” with the exam because questions required integration and interpretation; some admitted attempting quick searches but abandoned them as these felt counterproductive under time limits. Students described an informal norm of fairness - a sense that collaborative cheating would disadvantage peers and dilute the value of their learning. This concurs with research showing that task design (authentic, higher-order) and time-boundedness can reduce opportunistic cheating, and that unsupervised online exams can still yield stable rank-orders (Holden et al. 2021, Chan & Ahn 2023).

3.1.5. Inclusivity and Accessibility

Students with visual strain or small displays found navigation particularly demanding; some asked for adjustable image zoom and high-contrast figure versions. Others favoured being allowed a second device (e.g. tablet for figure PDF) to reduce scrolling. Participants urged more UDL-consistent affordances: flexible display options, clearer instructions, and explicit accessibility statements (CAST, Rao et al. 2014). Although participants’ neurotype was not disclosed, it is reasonable to assume that neurodivergent students may have felt disadvantaged.

3.2. Academics’ reflections

From the academic side, the VLE also presented several constraints: absence of integrated plagiarism detection for exam attempts, clunky and impossible blind double-marking due to examiner randomisation, occasional system instability. The VLE provider continued to enhance accessibility, but varied user experience and accessibility gaps remained.

Marking was pleasingly faster and more consistent (typed scripts; consistent script annotation), moderation was smoother (digital exchange; improved visibility of examiner actions and motives), environmental impact was reduced (no paper), and the risk of losing scripts was nil.

4. Discussion

Computer-assisted examinations had been under consideration for some time, but the onset of COVID-19 accelerated their implementation, leaving insufficient time to explore appropriate proctoring solutions or to design assessments aligned with the new modes of assessment.

4.1. Validity, Integrity and the Case for Authentic TRAs

Our observations and student testimony suggest that student-centric, authentic TRAs can maintain standards, differentiate performance, and support learning, provided the assessment

design (what is being asked) and the interface design (how it is presented) are aligned. The stability of performance distribution and alignment with ongoing engagement echo broader findings that unsupervised online exams, when designed thoughtfully, can meaningfully assess learning rather than simply inflate scores (Chan & Ahn 2023). The integrity literature is nuanced: evidence exists of higher cheating in some unproctored contexts (Fendler et al. 2024), yet task authenticity and question design reduce the advantage of searching or collusion. Time restriction can also impact on the usefulness of cheating; this, however, can greatly disadvantage neurodivergent students. Designing for integrity by focussing on higher order understanding, and not simply defaulting to heavy surveillance, seems to be a reasonable and educationally coherent strategy. Proctoring also raises ethical, privacy, and equity concerns, especially regarding intrusive data capture and algorithmic flagging; many UK discussions now advocate transparent, proportionate approaches and design-led deterrence (Barrio 2022, Heinrich 2025).

4.2. Student Partnerships as a Quality Mechanism

The focus groups were not merely a feedback exercise; they co-produced design actions. This aligns with the concept of students-as-partners; such partnerships can enhance assessment literacy, improve perceived fairness, and democratise quality enhancement (Bovill 2019). Of note, the groups surfaced silent struggles such as screen size, display clarity, cognitive load from scrolling, which would not be visible in grade data alone. Students' recommendations mapped directly to UDL principles: multiple means of representation (high-contrast images, zoomable galleries), simple, clear instructions, and amenable accommodations without diluting standards. These are practical, low-cost steps that increase validity by reducing e.g. screen-navigation costs.

4.3. Sector Guidance and the Future Landscape - What Now in an AI-Enabled World?

The literature increasingly emphasises the importance of moving from fixes to foresight and rethinking assessment for authenticity, adaptability, and trustworthiness, including in an AI-enabled landscape (Ashford-Rowe et al. 2014, Ajjawi et al. 2020, Villaroel et al. 2020; Bearman et al. 2022). Our TRA purposefully prioritised interpretation and integration, capabilities that are less amenable to generic AI generation and more reflective of authentic biomedical practice. This is consistent with calls to assess critical thinking, contextual understanding, and metacognition rather than rehearse what AI can readily do (Clewley & Clewley 2025).

4.4. Final Reflections

Drawing on student feedback and reflective review, this work highlights how relatively small design choices in examination structure can have a significant impact on students' ability to demonstrate understanding under time-limited conditions. Prioritising depth over breadth in data interpretation tasks, sequencing linked questions to support pacing, and positioning open-

ended questions to ensure equitable access to higher-order demonstration all emerged as important considerations. Careful attention to the placement, clarity, and consistency of visual materials seems to reduce unnecessary cognitive load and support sustained engagement.

4.5. Limitations

This study evaluated one module in a defined disciplinary and platform context. While limitations also include restricted plagiarism detection, inability to deal with collusion and small self-selected number of students, convergence across qualitative student accounts, outcome measures, and item-level observations provide some confidence in our reflections. We have attempted to mitigate academic bias by staying true to the student voice and iterative evaluation.

5. Conclusions

Lessons learnt from our evaluation support student-staff partnership as a means of interrogating and refining assessment design in complex STEMB contexts. By highlighting students' lived experiences of a time-limited, online examination, the work went beyond evaluation of student experience, offering more broadly applicable insights into how assessment structure, sequencing, and presentation can shape students' capacity to demonstrate understanding.

A key implication is that assessment design choices have substantive pedagogic consequences. Decisions relating to the number and framing of sub-questions, the sequencing of tasks, and the positioning of figures likely amplifies or alleviates cognitive load, particularly in disciplines characterised by dense conceptual relationships and specialist visual literacies. Designing for depth, conceptual flow, and clarity may support more sustained reasoning and more authentic demonstrations of disciplinary thinking, applicable beyond our module.

It seems that opportunities for higher-order synthesis and extension are most equitable when they are deliberately positioned to be accessible to all students, rather than reserved for the final moments of an examination. This could have broader implications for how assessments differentiate achievement without inadvertently privileging speed or exam endurance over understanding across multiple settings.

Importantly, this work reinforced the role of students as credible partners in assessment design. Their contributions extended beyond identifying surface-level issues to articulating how assessment structures intersect with cognition, engagement, and confidence. Such insights are rarely visible through traditional evaluation mechanisms but proved central to the IMI exam. We suspect that in an era of increasing attention to accessibility, authenticity, and AI-resilient assessment, student-led reviews will be useful approaches to the creation of assessments that remain rigorous while better aligned with how students think, learn, and demonstrate expertise.

References

- Ashford-Rowe, K., Herrington, J. and Brown, C. (2014) 'Establishing the critical elements that determine authentic assessment', *Assessment & Evaluation in Higher Education*, 39(2), 205-222. doi: 10.1080/02602938.2013.819566.
- Ajjawi, R., Tai, J., Huu Nghia, T.L., Boud, D., Johnson, L. and Patrick, C.-J. (2020) Aligning assessment with the needs of work-integrated learning: the challenges of authentic assessment in a complex context. *Assessment & Evaluation in Higher Education*, 45(2), 304-316. doi: 10.1080/02602938.2019.1639613.
- Barrio, F. (2022) Legal and pedagogical issues with online exam proctoring. *European Journal of Law and Technology*, 13(1). <https://ejlt.org/index.php/ejlt/article/view/886/1054>.
- Bearman, M., Nieminen, J. H. and Ajjawi, R. (2023) 'Designing assessment in a digital world: an organising framework', *Assessment & Evaluation in Higher Education*, 48(3), 291-304. doi: 10.1080/02602938.2022.2069674.
- Bovill, C. (2019) Student-staff partnerships in learning and teaching: An overview of current practice and discourse. *Journal of Geography in Higher Education*, 43(4), 385-398.
- CAST. (n.d.). *UDL on campus: Universal Design for Learning in higher education*. Retrieved December 21, 2025, from <https://udloncampus.cast.org>
- Chan, J. C. K. and Ahn, D. (2023) Unproctored online exams provide meaningful assessment of student learning. *Proceedings of the National Academy of Sciences USA*, 120(31), e2302020120, <https://doi.org/10.1073/pnas.2302020120>
- Clewley, C. and Clewley, L. (2025) *Could generative AI improve the REF?* Times Higher Education, 5 June.
- Fendler, R., Beard, D. and Godbey, J. (2024) A Robust Examination of Cheating on Unproctored Online Exams, *Electronic Journal of e-Learning*, 22(5), 26-38, <https://doi.org/10.34190/ejel.22.5.3173>
- Heinrich, E. (2025) A Systematic-Narrative Review of Online Proctoring Systems and a Case for Open Standards. *Open Praxis*, 17(3), 485-499. Available at: <https://doi.org/10.55982/openpraxis.17.3.836>.
- Holden, O. L., Norris, M. E., & Kuhlmeier, V. A. (2021) Academic integrity in online assessment: A research review. *Frontiers in Education*, 6, 639814. <https://doi.org/10.3389/educ.2021.639814>
- JISC. (2021). *Rethinking assessment in a digital age*. Retrieved December 21, 2025, from <https://www.jisc.ac.uk>
- Pengelly, J., Whipp, P.R. and Malpique, A. (2025) A testing load: A review of cognitive load in computer- and paper-based learning and assessment. *Technology, Pedagogy and Education*, 34(1), 1-17.
- QAA. (2020) *COVID-19: Thematic guidance - Securing academic standards and supporting student achievement*. Retrieved December 21, 2025, from <https://www.qaa.ac.uk>
- Rao, K., Ok, M.W. and Bryant, B.R. (2014) A review of research on universal design educational models. *Remedial and Special Education*, 35(3), 153-166. <https://doi.org/10.1177/0741932513518980>

Villarroel, V., Boud, D., Bloxham, S., Bruna, D. and Bruna, C. (2020) Using principles of authentic assessment to redesign written examinations and tests. *Innovations in Education and Teaching International*, 57(1), 38-49. doi: 10.1080/14703297.2018.1564882.

Zhan, Y., Boud, D. and Du, Z. (2025) Designing for authentic assessment: A scoping review. *Higher Education*. Adv. online publication. <https://doi.org/10.1007/s10734-025-01588-9>.

Article in Press