

Assessment & Evaluation in Higher Education

ISSN: 0260-2938 (Print) 1469-297X (Online) Journal homepage: www.tandfonline.com/journals/caeh20

Beyond open book versus closed book: a taxonomy of restrictions in online examinations

Phillip Dawson, Kelli Nicola-Richmond & Helen Partridge

To cite this article: Phillip Dawson, Kelli Nicola-Richmond & Helen Partridge (2024) Beyond open book versus closed book: a taxonomy of restrictions in online examinations, Assessment & Evaluation in Higher Education, 49:2, 262-274, DOI: 10.1080/02602938.2023.2209298

To link to this article: https://doi.org/10.1080/02602938.2023.2209298

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

Published online: 09 May 2023.

Submit your article to this journal

Article views: 8842

View related articles

View Crossmark data

Citing articles: 17 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=caeh20

ASSESSMENT & EVALUATION IN HIGHER EDUCATION
2024, VOL. 49, NO. 2, 262–274
https://doi.org/10.1080/02602938.2023.2209298

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Beyond open book versus closed book: a taxonomy of restrictions in online examinations

Phillip Dawson, Kelli Nicola-Richmond and Helen Partridge

Deakin University, Geelong, Victoria, Australia

ABSTRACT

Educators set restrictions in examinations to enable them to assess learning outcomes under particular conditions. The open book versus closed book binary is an example of the sorts of restrictions examiners have traditionally set. In the late 2000s this was expanded to a trinary to include open web examinations. However, the current technology environment, particularly for online examinations, makes this trinary not particularly useful. The web now includes generative artificial intelligence tools, and contract cheating sites, both of which are capable of completing examination questions within the examination period. Closed book, open book and open web no longer offers enough clarity or specificity when communicating examination restrictions. This article proposes a new taxonomy of restrictions for examinations, with a particular focus on online examinations. The taxonomy consists of three dimensions: information, people and tools. The paper explores the possible settings for those dimensions. Five criteria are provided to help examination designers in selecting restrictions: the learning outcomes being assessed; the feasibility of restrictions; consequential validity; authenticity; and values. Taken together, this taxonomy and the criteria provide ways of thinking about restrictions in examinations that can prompt educators towards examination designs that are more valid and robust against cheating.

KEYWORDS

Online examinations; open book examinations; cheating; academic integrity; remote proctoring

Introduction

One of the consequences of the COVID-19 pandemic was a rapid shift towards 'emergency remote teaching' (Hodges et al. 2020), and with it a shift towards what could be termed 'emergency remote assessment'. In the new world of physical distancing requirements, it became impossible to conduct many traditional forms of summative assessment, most notably the in-person invigilated examination. While some educators who previously used summative examinations moved away from them entirely, many moved to online examinations during the pandemic (Selwyn et al. 2023). In some contexts, this shift to online examinations appears likely to persist.

Examinations have existed for a very long time, dating back nearly 2,000 years (O'Sullivan and Cheng 2022). Passionate arguments have been made to abandon summative examinations

CONTACT Phillip Dawson p.dawson@deakin.edu.au
© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

from a variety of perspectives. Examinations have long been criticised by the assessment for learning movement, as they are seen to hurt learning through directing students towards poorer study habits; this has been conceptualised as a consequential validity problem in the form of negative effects on students beyond the immediate act of assessment (Sambell, McDowell, and Brown 1997). But examinations persist, for a variety of reasons. Higher education tends to trust traditional approaches more than new ones (Carless 2009), and examinations are arguably the most traditional of assessment traditions. Examinations are often trusted particularly due to their perceived anti-cheating benefits, which unfortunately may not be as great as thought (Harper, Bretag, and Rundle 2021). More broadly, testing is used in education and psychology due to its benefits for reliability, validity and fairness (American Educational Research Association, American Psychological Association and National Council on Measurement in Education 2014). There is also substantial evidence for the 'testing effect', which refers to benefits for learning due to being tested, as well as nuanced work into the learning benefits of examination preparation for students (Sotola and Crede 2021). And as with any decision in assessment, workload considerations likely play a part in the continued use of examinations (Bearman et al. 2017), possibly including that examinations are exempt from time-consuming feedback processes (Scoles, Huxham, and McArthur 2013). It appears that, while not without their critics, examinations are here to stay.

Examinations have traditionally been conducted in person, under the watchful eye of a teacher or a dedicated invigilator/proctor, and sometimes in a special venue. The main function of this surveillance is to attempt to ensure that examination conditions are upheld, and that cheating does not happen. But the sort of online examinations that have become dominant during the pandemic are different to face-to-face examinations. The physically-distanced examination of the COVID period has been largely conducted at home, on student-owned computers. Some examinations have been invigilated remotely through specialist remote proctoring software that locks down student computers, some through adaptations of existing online conferencing tools such as Zoom, and some examinations have been completely unsupervised. Such a range of changes brings with it a need to reconsider choices made in examination design and administration that may have previously been taken for granted.

This paper focuses on a particular 'assessment design decision' (Bearman et al. 2016) and how it may have changed in its nature and importance during the shift to at-home online examinations: the choice of how 'open' an online examination is. This has traditionally been thought of as a binary of closed book versus open book, with some examinations occupying a middle ground between the two, but what do those terms mean in the context of an online examination, where instead of just a book students have access to technologies that do not just store information, but provide access to tools and people?

Previous work has discussed the impact of computer-based examinations on the open versus closed book dilemma, most notably Williams and Wong (2009) who proposed a shift to closed-book, open-book and open-book-open-web classifications, and documented an institution-wide shift towards the latter. However, this body of work is largely from the pre-2010s when the technology context was markedly different. With the rise of remote proctoring and lockdown browsers, examination designers face new choices around openness, such as the choice to allow students to access notes, websites, algebra solvers, artificial intelligence writing tools and social media (Dawson 2021). Technology powerfully shapes assessment design decisions (Bennett et al. 2017), and the technology landscape has changed so much so quickly that it is time to revisit types of openness in online examinations and their consequences for assessment. The simple trinary of closed book, open book and open book open web is no longer sufficient. This paper provides a new taxonomy for setting restrictions in summative online examinations, and a set of criteria to be considered when making decisions around those restrictions.

Assessment security in online examinations

Examination restrictions are closely connected with concerns about cheating and validity. Examiners set restrictions against, for example, the use of notes, and then they need to enforce such restrictions through branding the use of notes as forbidden and taking measures against the use of notes. This is done for reasons of validity: when students do not complete a task under the conditions the examiner has designed the task for, judgements about student performance become less valid (Cizek and Wollack 2017; Dawson 2021). This means a sophisticated conversation about cheating is necessary to discuss examination restrictions in a similarly sophisticated way.

A key distinction to be drawn when discussing approaches to address cheating in examinations is between the positive idea of academic integrity and the more adversarial notion of assessment security (Dawson 2021). Examinations are predominantly an assessment security approach as they rely on strategies like surveillance to address cheating. In contrast, academic integrity approaches focus on developing students so they have particular values and capabilities that are associated with ethical educational practice. Academic integrity might be of benefit to students sitting an examination, but the examination remains predominantly an exercise in assessment security. Decisions about the restrictions imposed on students sitting an examination need to be viewed through an assessment security lens.

There are two components to assessment security: control of circumstances and authentication (Dawson 2021), and both are relevant to restrictions in online examinations. Control of circumstances refers to approaches taken to assure what could be called 'examination conditions', whatever they may be in a particular examination. In a traditional closed-book face-to-face examination the techniques used for control of circumstances include the distribution of examination papers at a specific time, inspection of the items students have at their tables, and listening for unauthorised communications. Authentication refers to approaches taken to ensure that the student is who they say they are, and that they have produced their own work. In traditional closed-book face-to-face examinations, authentication approaches include checking student identification cards, and disallowing communication devices that would allow a third party to tell the student what to write.

Assessment security enforces any restrictions in assessment, so it is valuable to take stock of the assessment security approaches available to online examination designers. These include: remote proctoring by humans or artificial intelligence, which may be done *via* a recording or in real time; lockdown approaches to restrict access to computer functions; identity checks, which may be against stored biometrics such as a voiceprint or photograph, or against a student-provided ID card; and the use of randomisation of examination questions. There is very limited evidence as to the effectiveness of any of these approaches, especially in terms of their robustness against circumvention by an online examination taker (Dawson 2021). This is a notable absence, and a hindrance for discussion around restrictions to be imposed in examinations; however, it is not unique to the online context, as there is evidence that face-to-face examinations are likely more vulnerable to cheating than commonly thought (Harper, Bretag, and Rundle 2021).

Perhaps counterintuitively, each restriction imposed on students in assessment increases the range of possible cheating approaches that can be taken, making the task potentially less secure. For example: if students are forbidden from using a calculator, that restriction must be enforced; if students are not allowed to talk with their peers, then this needs to be enforced; and if a task is set where students are prohibited from accessing particular information, they need to be stopped from accessing that information. This corresponds to the cybersecurity concept of a system's 'attack surface': 'the set of ways in which the system can be attacked' (Manadhata 2013, 1). Tasks with many restrictions have a larger attack surface, as each restriction brings with it a set of ways that it can be circumvented.

Knowingly or unknowingly setting restrictions that are not (or cannot be) adequately enforced is a threat to the validity of a task. If a task is designed to be conducted without access to notes, but some students access their notes, then any judgements about those students'

outcomes for that task are invalid (Dawson 2021). In this way, restrictions are not just a problem for assessment security, but also a problem for validity and fairness. The extent of this problem depends on how well restrictions can be enforced.

Enforcing restrictions in online examinations

Assessment security has always been challenging in examinations, dating back at least as far as the Chinese imperial examination system (O'Sullivan and Cheng 2022), which attempted to enforce restrictions similar to those in modern face-to-face examinations: that the examination be closed-book, with outside materials forbidden; that the examination be undertaken by the student being examined and not some third party; and that examination scripts be anonymous, to reduce the chance of bias from markers. These restrictions were breached successfully by some students, and subsequently a range of assessment security innovations were introduced to address them. Refinements to assessment security have continued over the intervening period, and current face-to-face examinations at universities have the benefit of more than a thousand years of trial and error. Unfortunately, many of those lessons do not transfer well to the context of online examinations.

When students can choose the location of their examination, there are many low-tech or even no-tech ways they can breach some restrictions. In a completely unsupervised online examination, students can breach closed book restrictions by merely opening a book. Remote proctoring claims to make this harder by monitoring students while they sit the examination through their webcam. However, there are reports online that students can successfully breach remote proctoring's closed book restrictions by using physical approaches such as having notes attached to the rear of their webcam, or writing notes onto a transparent screen protector and attaching it to their laptop screen. More high-tech approaches to breach closed-book restrictions include 'USB key injectors', which are devices that impersonate a USB keyboard and are programmed to type out some predefined text at a particular time (Dawson 2016), and hidden earpieces that can play back pre-recorded notes. We are unable to test these approaches, as although this paper's lead author has sought permission from many remote proctoring companies to try out a study where he attempts to cheat, none have agreed, and he has received legal advice that doing such a study without the companies' approval would be unwise.

While remote proctoring and lockdown vendors claim their tools are effective at stopping students from breaching restrictions, the limited peer reviewed empirical research in this area tells a different story. In one study, by Bergmans et al. (2021), 30 computer science students sat an examination, and six of them were asked to cheat in the examination. None of the students' attempts to cheat were detected by the remote proctoring software, and when recordings of the examinations were reviewed by a person, only one of the six 'cheating' students was caught. In another study, Burgess et al. (2022) experimented with four remote proctoring tools and found that the anti-cheating features in all four 'can be trivially bypassed'. These studies were conducted in artificial settings, where there were no stakes involved, and no real students – however, we are not aware of any other studies into the effectiveness of remote proctoring at detecting (as opposed to deterring) cheating. Taken together, the reluctance of remote proctoring companies to allow research into cheating, and the negative results of the two studies that have been published, suggest that restrictions are difficult to enforce in online examinations.

If an examination is designed around particular restrictions being enforced, then enforcing those restrictions becomes an essential requirement for the validity of the examination. For example, if an examination has been designed as closed-book because it requires the recall of facts that could be trivially looked up in a book, then not enforcing the closed-book restriction means assessors cannot make meaningful judgements about the capabilities of students who use their books. In addition to being a validity threat, this also creates a disparity between those students who want to cheat and are able to, and everybody else. Restrictions should

therefore be used only when necessary, and only when there is a belief they can be adequately enforced. Such a decision needs to be made with respect to the full range of restrictions possible in online examinations – but what is that range?

Dimensions of openness and restriction in online examinations

In face-to-face examinations it is common to consider a dichotomy of closed book versus open book examinations. For example, Durning et al.'s (2016) systematic review found 37 studies that compared open book examinations to closed book examinations. However, the term 'open book' is somewhat ambiguous in an online context. Some work has been done on extending open/closed book to also include 'open web' (Williams and Wong 2009), acknowledging that the web is different to a book. However, as that work is relatively old, the web it refers to was much less complex than the web of today. 'Open web' might accommodate relatively static websites like Wikipedia, but what about interactive 'homework help' sites such as Chegg, computer algebra solvers and artificial intelligence chat bots, which are all part of the web. There is a need for a more nuanced way of specifying what students are allowed to do during an examination (openness) or conversely, what they are not allowed to do (restrictions). We propose a set of three dimensions of openness or restriction in online examinations, each with a range of possible levels of openness or restriction, as summarised in Table 1. For each dimension there are a variety of potential settings, ranging from complete openness to complete restriction, and many steps in between.

Information

While many educators embrace open book examinations, not all of them would be happy if students had access to all of humanity's information, which may include answers to their specific examination. Some information is capable of completely invalidating some examinations. For this reason it is important to break down information further than an open book versus closed book binary.

As a first step, it is necessary to draw a distinction between websites, apps and other technologies that provide information, and others that instead provide access to people or tools. The *Information* dimension is concerned with relatively static content that already existed prior to the commencement of the examination. For example, a synchronous chat conducted through the Facebook Messenger website or app is not information in the sense we are concerned with here, it instead is better considered under the *People* dimension. Similarly, while the generative artificial intelligence ChatGPT is available through a website, it falls under the *Tools* dimension rather than *Information*. All types of media are considered information here, so this dimension spans printed and electronic text, audio, video and other media.

Some websites students would want to use during an examination are static. Journal articles, textbooks, online dictionaries and even Wikipedia do not change significantly in response to the

Table 1. Three dimensions of restrictions for online examinations.

Dimension	Definition	Possible settings
Information	Books, notes, websites, etc that may be either dynamic or static, and may exist in various different types of media	No information; only a specified list; offline only; full openness minus prohibited list; full openness
People	Working with other people to achieve the aims of the task	Who: no other people; teacher-selected peers; student-selected peers; teacher-selected others; student-selected others; any people What: feedback, collaboration, consultation, outsourcing
Tools	Software and hardware that can be used to do the task or make it easier, more efficient, or higher quality	No tools; specific tools allowed; specific tools disallowed; all tools

examination a student is sitting. However, some sites feature greater degrees of interactivity, or may update themselves over the course of the examination to be different to what the examiner expects. Examples include web forums, where a student might ask a question that is in the examination. If an answer is given to that question it can now be accessed by all students in the class. In this instance, one student's outsourcing to the forum has created an *Information* problem that threatens the assessment of the entire class. This makes dynamic websites more of a challenge.

There are many possible settings for the level of restrictions to impose on the *Information* dimension, but a minimal list would include the following. *Full openness* would mean the entire web, with the exception of sites restricted under the *People* and *Tools* dimensions, as well as whatever other information sources students had access to (e.g. printed books, notes, audio recordings). *Full openness minus prohibited list* adopts a blocklist approach and removes access to a specified set of sites but is otherwise the same as *full openness*. While it may be possible to build a list of websites that are forbidden, and allow students access to all of the web except these sites, doing so might be infeasible. Taking just the category of assignment sharing websites, there are many thousands of these sites and they are very simple to set up (Ellis, Zucker, and Randall 2018). *Access to only a specified list* takes an allowlist approach, and is likely more feasible as it involves specifying the information resources that students are allowed to access. Such a list could be constructed collaboratively within an institution or discipline group, or it could be constructed in consultation with the students in a particular course. *Offline only* permits students to access only information they already have at hand on their computer or in physical notes. Finally, *no information* denies access to any and all notes, books and other sources.

People

Traditional examinations are usually solo activities, with access to other people entirely restricted during the course of the examination. However, this is not universal. Literature on collaborative examinations stretches back at least as far as the 1990s (Stearns 1996). One specific approach, the two-stage collaborative examination (Jang et al. 2017; Kinnear 2021) requires students to first sit an examination on their own, which is graded, and then collaborate with a group of peers to respond to the same examination, which is also then graded. While these types of examinations offer access to people, they are tightly controlled in terms of who those people are. As with *Information*, the *People* dimension is not a binary of allowing everything versus disallowing everything, it is a spectrum full of choices for examination designers.

One of the most significant challenges for the *People* dimension is establishing what students have done themselves, and what other people have done for them. As in other assessment, it can be hard to disentangle or demarcate the lines between feedback, collaboration, collusion and outsourcing. Guidelines can be given to students, for example they may be told that they can ask their peers for feedback but those peers cannot share their own answers. Or students can be given almost free reign, as in Kellermann's (University of New South Wales 2020) online engineering examination, where students have access to a shared Microsoft Teams channel with their peers and the only rule is that posts must not be illegal.

Possible settings for the *People* dimension are many and varied, and consist of two components: who the people are, and what they can do. In terms of who, there are several settings. The traditional examination adopts a *no other people* setting. We have not heard of an examination that allows *any people* of the student's choosing. Various collaborative examinations allow collaboration with *teacher-selected peers* or collaboration with *student-selected peers*. Some niche examinations allow *teacher-selected others* or *student-selected others*, for example a music performance examination where the teacher or student selects an accompanist for the student's performance. In terms of what the involvement of other people can be, students can engage with these people in different and potentially overlapping ways. Some examinations allow

feedback, such as conversations about the work at some set point, which could for example take the form of structured peer feedback sessions. *Collaboration* allows students to work with each other to produce more than they otherwise could individually, or to produce higher quality or more accurate work. *Outsourcing* involves students getting someone else to do all or part of their work, which would require them to carefully communicate the requirements of the task to others. Finally, *consultation* involves connecting with other people to get advice or expert input.

Tools

For some learning outcomes, there exist tools that can substantively demonstrate the learning outcome for the student even if the student is not capable of doing so without the tool. As a result, educators often restrict tools in examinations. Students might be allowed access to a calculator in a particular examination, but not a computer algebra solver as it would essentially do the work for the student. Spell checking tools might be allowed in one online examination, but not the ChatGPT generative artificial intelligence, which is claimed in one recent paper (Kung et al. 2022) to be capable of satisfying much of the requirements of the United States Medical Licensing Examination. As with the other dimensions, *Tools* represents a diverse range of options for examination designers.

The rise of generative artificial intelligence tools, such as ChatGPT, deserves a special mention in a consideration of *Tools*. This class of artificial intelligence can produce text, images, computer code, music, video and other types of media for students based on prompts. The ability to restrict access to ChatGPT has been cited as justification for the increased use of face-to-face examinations (Cassidy 2023). In terms of its affordances, generative artificial intelligence can be considered functionally similar to contract cheating sites, as they allow students to use natural language to request the completion of all or part of their work. In this way, generative artificial intelligence might be difficult to differentiate from the *People* dimension, especially as these tools become more sophisticated. However, from an ethical perspective it could be argued that outsourcing to a person is different to using a tool; there are also circumstances where students are expected to use these tools as part of demonstrating particular learning outcomes.

Possible settings for the tools allowable include *no tools*, that is, denying students access to all tools. Given the utility of pens and paper for supporting cognition, and the ubiquity of writing in many examinations, the *no tools* setting is likely rare. Most examinations come with the (sometimes implicit) expectation that there are *specific tools allowed*, and institutions often have clear procedures in place for the tools that are standard issue in examinations and lists of what is allowed. Alternatively, all tools can be assumed allowed except for those *specific tools disallowed*. These lists can specify individual tools, or categories of tools, such as generative artificial intelligence, writing instruments and computer algebra solvers, or take an affordance-based approach focusing on what the tools can do (Dawson 2021). Finally, *all tools* is also an available setting, which would permit students to use everything from pen and paper through to an artificial intelligence writing tool.

Criteria for setting restrictions for online examinations

How are examination designers to navigate the range of potential restrictions available? We discuss a set of five criteria that can be used to prompt thinking when selecting restrictions: learning outcomes, feasibility, consequential validity, authenticity and values.

Learning outcomes

In a criterion-referenced or standards-based system of assessment, the primary purpose of an examination is to assess the extent to which learning outcomes have been achieved. Restrictions

in an examination should be in service of the assessment of the learning outcomes. In general, the assessment of lower-level learning outcomes requires more restrictions. Consider the highest and lowest ends of the SOLO taxonomy (Biggs 1999). At the highest level, 'extended abstract', students are required to hypothesise, formulate or reflect. At the lowest levels, 'unistructural' and 'multistructural', students are required to describe, list or identify. If an examination assesses these lower-level outcomes that focus on factual recall, then allowing access to the *Information* dimension may significantly threaten the validity of assessor judgements. However, if the assessment instead assesses higher-level outcomes, focused on the application or extension of existing knowledge, then more permissive settings on the *Information* dimension may be warranted.

Some examinations assess outcomes across multiple levels. This creates challenges for setting restrictions. For example, if a statistics examination includes low-level outcomes about memorisation of particular formulae, as well as higher-level outcomes involving the critique of published research papers' methods sections, then the assessment designer has to decide if they will target the restrictions to the higher or lower-level outcomes. This can be addressed through setting multi-part examinations with different restrictions in each stage based on the outcomes being assessed.

New *Tools* are increasingly able to demonstrate outcomes for students either partially or completely. This may require changes to the outcomes that are assessed. Bearman and Luckin (2020) argue that, in this new context, the capability to make decisions about the quality of work, or 'evaluative judgement' (Tai et al. 2018), becomes a crucial learning outcome as it is likely to remain a human responsibility.

Feasibility

While it is possible for examination designers to set whatever restrictions they would like, not all restrictions are enforceable. In a face-to-face examination it might be possible to stop students from accessing particular *Tools*; in a supervised online examination this is more difficult, and in an unsupervised online examination many types of restriction are completely infeasible to enforce. *Information* is particularly difficult to restrict in any type of online examination conducted remotely, as examiners are unable to adequately inspect the space the examination is being conducted within. *People* can also be challenging to restrict in an online examination, as there are ways to communicate with other test-takers that may be difficult to detect, such as the use of hidden earpieces.

Examination designers should not set restrictions that cannot be adequately enforced. Doing so creates a significant validity problem. If an examination has been designed around students not having access to *Information* but the students still access it, accurate judgements about student capability cannot be made. This also creates an equity problem as not all students necessarily have the capability or the inclination to access the forbidden information. The same is true for unenforced restrictions in the *People* and *Tools* dimensions.

In considering feasibility, assessment designers also need to consider if allowing or denying is better for a given dimension. In general, it is easier to deny everything except a set of allowed *Information*, *Tools* and *People*. The alternative, allowing everything except a set of prohibited items, requires that list to be exhaustive. This is a much larger challenge, both in initially setting up the blocklist, and in keeping it up to date.

Enforcement of restrictions can be implemented by making breaching the restrictions difficult, or through making breaches easy to detect and prove (Dawson 2021). Deterrent effects, whereby students think they can get caught breaching restrictions, can also be used. However, the field of cybersecurity would suggest that 'security theatre' approaches that are dependent more on perception than actual detection capabilities can have short-lived effectiveness (Schneier 2018). The degree of enforcement of restrictions that is deemed adequate will vary from context to context, and it is worth remembering that perfect enforcement is not possible, and some more effective approaches may require more resourcing than is available.

Consequential validity

In assessment, consequential validity refers to the impacts of the assessment beyond its ability to allow judgements to be made about student learning (Sambell, McDowell, and Brown 1997; St-Onge et al. 2017). Restrictions carry with them impacts for consequential validity. For example, students deploy different study strategies for closed book versus open book examinations (Johanns, Dinkens, and Moore 2017), and these have implications for long-term learning (Agarwal and Roediger 2011). In choosing restrictions, examination designers need to think about their impact on students' lives beyond the immediate act of assessment.

Remote proctoring deserves particular consideration as it carries complex implications for consequential validity. There are claims in the literature that remote proctoring has implications for privacy (Karim, Kaminsky, and Behrend 2014; Selwyn et al. 2023), discomfort and anxiety for students (Silverman et al. 2021), creating a culture of distrust and punishment (Swauger 2020) and potential discrimination (Logan 2020). However, to some students those are reasonable sacrifices to be made in order to sit an examination at a location of their choosing (Balash et al. 2021) – which is a positive for consequential validity. The degree to which these factors play out in different contexts will vary, but they are significant to the consequential validity of remote proctored examinations.

Authenticity

When thinking about restrictions, assessment designers may wish to consider the types of *Information, People* and *Tools* that graduates will use when they graduate. Building on ideas of authentic assessment (Villarroel et al. 2018), examination designers may wish to consider the notion of *authentic restrictions*, which asks: 'does this restriction apply to professionals in the discipline as they complete this task?' (Dawson 2021, 136) and suggests that a justification should be provided where inauthentic restrictions are in place. Possible justifications might include, for example, the need to develop automaticity with the fundamental capabilities of a discipline, or the need to function quickly in an emergency. Authenticity is not in and of itself a good thing, nor a panacea against cheating (Ellis et al. 2020); however, by challenging the need for restrictions it can help to reduce the attack surface of the task.

Values

Finally, the values of the institution, discipline and assessment designer likely play an important role in the selection of restrictions – just as they do in other assessment design decisions (Bearman et al. 2017). For example, an institution might pride itself in student academic freedom, and view restrictions as anathema to its mission. Another might value equity above all else and seek a careful palette of restrictions it regards as promoting equity. Some disciplines value automaticity in recall of basic facts and the application of basic formulae and may wish to restrict access to *Information* and *Tools*. One individual assessment designer may value deeper learning, and another might value content coverage, and these would shape the restrictions they choose.

Discussion and conclusion

This paper has argued that there is a need for a fresh look at restrictions in examinations, and in particular online examinations. It advocates for a shift beyond binaries of open/closed and a more conceptual view of what is restricted than just books and the web. The three dimensions proposed, *Information*, *People* and *Tools*, each exist on a spectrum rather than being all or nothing. Educators can use the three dimensions when thinking about the restrictions they set, and the

five criteria of learning outcomes, feasibility, consequential validity, authenticity and values to guide them. Within these three dimensions, a diverse range of examinations can be specified. A multi-stage first-year anatomy examination might allow *no information* in order to assess lower-level outcomes, while also allowing *feedback with student-selected peers* after the first part of the examination is over, then *access to specific tools* to allow students to use a spell checker to check their work for an extended written section. A computer science examination might afford *full openness* in terms of information to simulate the authentic reality of writing computer software, as well as *all tools* for the same reason, but completely restrict students to *no other people* to assess specifically what each student is capable of. These two examples demonstrate the subtleties required when considering openness and restrictions and the need to move beyond simple binaries. We make no claims of completeness of our three dimensions. There may be other high-level dimensions, and there will likely be other settings possible within each dimension.

While we have discussed some potential criteria for how to think about restrictions, there remains limited evidence in terms of the impact of different types of restrictions. One of the most studied is the impact of open book versus closed book in examinations (Durning et al. 2016; Johanns, Dinkens, and Moore 2017). The challenge remains how to evidence claims that are made about restrictions, and how to produce evidence, both on the effectiveness of restrictions and also of the potential harms and unexpected benefits that restrictions may introduce.

We have discussed restrictions as things that are selected by assessment designers; however, it is worth noting that they operate within a context of procedures and technologies that shape what the designers do (Bennett et al. 2017). In particular, the enforcement of restrictions in online examinations is often done through technologies developed by large private companies. It is their software developers and product managers who ultimately define what restrictions are possible, and what types of restrictions can be turned on and off. This has parallels with the world of video games, where the rules of games are codified by programmers, and they can be difficult to change even if they ruin the fun for players (Consalvo 2007).

The existing binary notions of open/closed book, and the extended open web concept introduced by Williams and Wong (2009) left a lot of room for ambiguity in communications about restrictions in examinations. This has resulted in a research literature where it is unclear if, for example, students were allowed the use of contract cheating sites in studies that report on open web examinations. The particulars of the restrictions in any given study of online examinations should shape the interpretation of that study's findings. These challenges with communications extend to teachers communicating with each other about examinations if they do not share a common understanding of the restrictions in place for any given examination. Possibly the most disadvantaged are students, who do not usually have access to clear communications about the restrictions in place for any given task. We think that our proposed framework provides an opportunity to clarify the restrictions in any given examination, which can support clearer and more replicable research and practice.

This framework was designed with examinations in mind, especially online examinations; however, it could also be applied to assessment more broadly. With the rise of generative artificial intelligence, there have been calls to ban tools like ChatGPT in some educational contexts (e.g. New York: Elsen-Rooney 2023). Those seeking to place restrictions on assessment in general may find it useful to structure their thinking around the types of restrictions that can be put in place, which we regard as controls on *Information*, *People* and *Tools*, and reflect on the criteria of learning outcomes, feasibility, consequential validity, authenticity and values when making their decisions.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on Contributors

Phillip Dawson is a professor and Co-Director of the Centre for Research in Assessment and Digital Learning (CRADLE) at Deakin University. His research focuses on assessment, feedback, and cheating.

Kelli Nicola-Richmond is an associate professor and Associate Head of School, Teaching and Learning of the School of Health & Social Development at Deakin University. Her research focuses on evaluative judgement, exams and cheating, Long COVID and student underperformance and failure in clinical placement.

Helen Partridge is a professor and Pro Vice-Chancellor, Teaching and Learning at Deakin University. Her research focuses on the interplay between information, technology and learning.

ORCID

Phillip Dawson http://orcid.org/0000-0002-4513-8287
Kelli Nicola-Richmond http://orcid.org/0000-0003-4874-5055
Helen Partridge http://orcid.org/0000-0002-8365-870X

References

Agarwal, P. K., and H. L. Roediger. 2011. "Expectancy of an Open-Book Test Decreases Performance on a Delayed Closed-Book Test." *Memory (Hove, England)* 19 (8): 836–852. doi:10.1080/09658211.2011.613840.

National Council on Measurement in Education 2014. *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association, American Psychological Association.

Balash, D. G., D. Kim, D. Shaibekova, R. A. Fainchtein, M. Sherr, and A. J. Aviv. 2021. "Examining the Examiners: Students' Privacy and Security Perceptions of Online Proctoring Services." In *Proceedings of the Seventeenth Symposium on Usable Privacy and Security*, 633–652. Virtual Conference: USENIX Association.

Bearman, M., P. Dawson, S. Bennett, M. Hall, E. Molloy, D. Boud, and G. Joughin. 2017. "How University Teachers Design Assessments: A Cross-Disciplinary Study." *Higher Education* 74 (1): 49–64. doi:10.1007/s10734-016-0027-7.

Bearman, M., P. Dawson, D. Boud, S. Bennett, M. Hall, and E. Molloy. 2016. "Support for Assessment Practice: Developing the Assessment Design Decisions Framework." *Teaching in Higher Education* 21 (5): 545–556. doi:10.1080/13562517.2016.1160217.

Bearman, M., and R. Luckin. 2020. "Preparing University Assessment for a World with AI: Tasks for Human Intelligence." *In Re-Imagining University Assessment in a Digital World*, edited by Margaret Bearman, Phillip Dawson, Rola Ajjawi, Joanna Tai and David Boud. Cham, Switzerland: Springer.

Bennett, S., P. Dawson, M. Bearman, E. Molloy, and D. Boud. 2017. "How Technology Shapes Assessment Design: Findings from a Study of University Teachers." *British Journal of Educational Technology* 48 (2): 672–682. doi:10.1111/bjet.12439.

Bergmans, L., N. Bouali, M. Luttikhuis, and A. Rensink. 2021. "On the Efficacy of Online Proctoring Using Proctorio." In "*Proceedings of the.*" 13th International Conference on Computer Supported Education (CSEDU 2021) - Volume 1, 279–290. SCITEPRESS – Science and Technology Publications, Lda. doi:10.5220/0010399602790290.

Biggs, J. 1999. "What the Student Does: Teaching for Enhanced Learning." *Higher Education Research & Development* 18 (1): 57–75. doi:10.1080/0729436990180105.

Burgess, B., A. Ginsberg, E. W. Felten, and S. Cohney. 2022. "Watching the Watchers: Bias and Vulnerability in Remote Proctoring Software." In *31st USENIX Security Symposium (USENIX Security 22)*, 571–588. Boston, MA: USENIX Association.

Carless, D. 2009. "Trust, Distrust and Their Impact on Assessment Reform." *Assessment & Evaluation in Higher Education* 34 (1): 79–89. doi:10.1080/02602930801895786.

Cassidy, C. 2023. "Australian universities to return to 'pen and paper' exams after students caught using AI to write essays." The Guardian, accessed 22 January 2023. https://www.theguardian.com/australia-news/2023/jan/10/universities-to-return-to-pen-and-paper-exams-after-students-caught-using-ai-to-write-essays.

Cizek, G. J., and J. A. Wollack. 2017. "Exploring Cheating on Tests: The Context, the Concern, and the Challenges." In *Handbook of Quantitative Methods for Detecting Cheating on Tests* edited byGregory J Cizek and James A Wollack, 3–19. Abingdon: Routledge.

Consalvo, M. 2007. *Cheating: Gaining Advantage in Videogames*. Cambridge, MA: MIT Press.

Dawson, P. 2016. "Five Ways to Hack and Cheat with Bring-Your-Own-Device Electronic Examinations." *British Journal of Educational Technology* 47 (4): 592–600. doi:10.1111/bjet.12246.

Dawson, P. 2021. *Defending Assessment Security in a Digital World: Preventing e-Cheating and Supporting Academic Integrity in Higher Education*. Abingdon, Oxon: Routledge.

Durning, S. J., T. Dong, T. Ratcliffe, L. Schuwirth, A. R. Artino, J. R. Boulet, and K. Eva. 2016. "Comparing Open-Book and Closed-Book Examinations: A Systematic Review." *Academic Medicine : journal of the Association of American Medical Colleges* 91 (4): 583–599. doi:10.1097/ACM.0000000000000977.

Ellis, C., K. van Haeringen, R. Harper, T. Bretag, I. Zucker, S. McBride, P. Rozenberg, P. Newton, and S. Saddiqui. 2020. "Does Authentic Assessment Assure Academic Integrity? Evidence from Contract Cheating Data." *Higher Education Research & Development* 39 (3): 454–469. doi:10.1080/07294360.2019.1680956.

Ellis, C., I. M. Zucker, and D. Randall. 2018. "The Infernal Business of Contract Cheating: Understanding the Business Processes and Models of Academic Custom Writing Sites." *International Journal for Educational Integrity* 14 (1): 1. doi:10.1007/s40979-017-0024-3.

Elsen-Rooney, M. 2023. "NYC education department blocks ChatGPT on school devices, networks." Chalkbeat New York, accessed 22 January 2023. https://ny.chalkbeat.org/2023/1/3/23537987/nyc-schools-ban-chatgpt-writing-artificial-intelligence.

Harper, R., T. Bretag, and K. Rundle. 2021. "Detecting Contract Cheating: Examining the Role of Assessment Type." *Higher Education Research & Development* 40 (2): 263–278. doi:10.1080/07294360.2020.1724899.

Hodges, C. B., S. Moore, B. B. Lockee, T. Trust, and M. A. Bond. 2020. "The difference between emergency remote teaching and online learning."

Jang, H., N. Lasry, K. Miller, and E. Mazur. 2017. "Collaborative Exams: Cheating? Or Learning?" *American Journal of Physics* 85 (3): 223–227. doi:10.1119/1.4974744.

Johanns, B., A. Dinkens, and J. Moore. 2017. "A Systematic Review Comparing Open-Book and Closed-Book Examinations: Evaluating Effects on Development of Critical Thinking Skills." *Nurse Education in Practice* 27: 89–94. doi:10.1016/j.nepr.2017.08.018.

Karim, M. N., S. E. Kaminsky, and T. S. Behrend. 2014. "Cheating, Reactions, and Performance in Remotely Proctored Testing: An Exploratory Experimental Study." *Journal of Business and Psychology* 29 (4): 555–572. doi:10.1007/s10869-014-9343-z.

Kinnear, G. 2021. "Two-Stage Collaborative Exams Have Little Impact on Subsequent Exam Performance in Undergraduate Mathematics." *International Journal of Research in Undergraduate Mathematics Education* 7 (1): 33–60. doi:10.1007/s40753-020-00121-w.

Kung, T. H., M. Cheatham, A. Medenilla, C. Sillos, L. De Leon, C. Elepaño, M. Madriaga, et al. 2022. "Performance of ChatGPT on USMLE: Potential for AI-Assisted Medical Education Using Large Language Models." *PLOS Digital Health* 2 (2):e0000198. doi: 10.1371/journal.pdig.0000198.

Logan, C. 2020. "*Refusal, Partnership, and Countering Educational Technology's Harms*." Hybrid Pedagogy. https://hybridpedagogy.org/refusal-partnership-countering-harms/

Manadhata, P. K. 2013. Game Theoretic Approaches to Attack Surface Shifting. In *Moving Target Defense II*, edited by Sushil Jajodia, Anup K. Ghosh, V. S. Subrahmanian, Vipin Swarup, Cliff Wang and X. Sean Wang, 1-13. New York, NY: Springer.

O'Sullivan, B., and L. Cheng. 2022. "Lessons from the Chinese Imperial Examination System." *Language Testing in Asia* 12 (1): 52. doi:10.1186/s40468-022-00201-5.

Sambell, K., L. McDowell, and S. Brown. 1997. "But is It Fair?": An Exploratory Study of Student Perceptions of the Consequential Validity of Assessment." *Studies in Educational Evaluation* 23 (4): 349–371. doi:10.1016/S0191-491X(97)86215-3.

Schneier, B. 2018. *Click Here to Kill Everybody: Security and Survival in a Hyper-Connected World*. New York, NY: WW Norton & Company.

Scoles, J., M. Huxham, and J. McArthur. 2013. "No Longer Exempt from Good Practice: Using Exemplars to Close the Feedback Gap for Exams." *Assessment & Evaluation in Higher Education* 38 (6): 631–645. doi:10.1080/02602938.2012.674485.

Selwyn, N., C. O'Neill, G. Smith, M. Andrejevic, and X. Gu. 2023. "A Necessary Evil? The Rise of Online Exam Proctoring in Australian Universities." *Media International Australia:1329878X211005862* 186 (1): 149–164. doi:10.1177/1329878X211005862.

Silverman, S., A. Caines, C. Casey, B. Garcia de Hurtado, J. Riviere, A. Sintjago, and C. Vecchiola. 2021. "What Happens When You Close the Door on Remote Proctoring? Moving toward Authentic Assessments with a People-Centered Approach." *To Improve the Academy* 39 (3):115–131. doi:10.3998/tia.17063888.0039.308.

Sotola, L. K., and M. Crede. 2021. "Regarding Class Quizzes: A Meta-Analytic Synthesis of Studies on the Relationship between Frequent Low-Stakes Testing and Class Performance." *Educational Psychology Review* 33 (2): 407–426. doi:10.1007/s10648-020-09563-9.

St-Onge, C., M. Young, K. W. Eva, and B. Hodges. 2017. "Validity: One Word with a Plurality of Meanings." *Advances in Health Sciences Education : Theory and Practice* 22 (4): 853–867. doi:10.1007/s10459-016-9716-3.

Stearns, S. A. 1996. "Collaborative Exams as Learning Tools." *College Teaching* 44 (3): 111–112. doi:10.1080/87567555.1996.9925564.

Swauger, S. 2020. *Our Bodies Encoded: Algorithmic Test Proctoring in Higher Education*. In Critical Digital Pedagogy, edited by Jesse Stommel, Chris Friend and Sean Michael Morris. Hybrid Pedagogy Inc.

Tai, J., R. Ajjawi, D. Boud, P. Dawson, and E. Panadero. 2018. "Developing Evaluative Judgement: Enabling Students to Make Decisions about the Quality of Work." *Higher Education* 76 (3): 467–481. doi:10.1007/s10734-017-0220-3.

University of New South Wales. 2020. *"Case Studies: Online Assessment* " accessed 16January 2023. https://www.teaching.unsw.edu.au/academic-integrity/case-studies.

Villarroel, V., S. Bloxham, D. Bruna, C. Bruna, and C. Herrera-Seda. 2018. "Authentic Assessment: Creating a Blueprint for Course Design." *Assessment & Evaluation in Higher Education* 43 (5): 840–854. doi:10.1080/02602938.2017.1412396.

Williams, J. B., and A. Wong. 2009. "The Efficacy of Final Examinations: A Comparative Study of Closed-Book, Invigilated Exams and Open-Book, Open-Web Exams." *British Journal of Educational Technology* 40 (2): 227–236. doi:10.1111/j.1467-8535.2008.00929.x.