Designing Trustworthy Assessments without Proctoring

David Joyner, david.joyner@gatech.edu

Executive Director of Online Education, College of Computing, Georgia Tech

Table of Contents

The rush to emergency remote teaching due to COVID-19 has many instructors looking to perform near-replacements of in-person activities; by that I mean, we appropriately strive, given these emergency circumstances, to find 1:1 replacements for what we would be doing in person. For tests, the 1:1 replacement would appear to be digital proctoring services.

There are lots of digital proctoring services out there. I personally use ours extensively: four of my five classes use digital proctoring, covering 11 exams and over 3800 individual sessions this semester alone. I like to use it because in a purely online program, we need frequent assertions that students are doing their own work, of the kind we would get naturally if they were in our physical classrooms.

Should You Use Online Proctoring? Probably Not

I provide that disclaimer to note that I'm not a detractor with regard to online proctoring in general; it has an important role to play. That said, in the current rapid emergency shift to remote assessment, I feel digital proctoring is not an ideal solution. It is necessarily invasive, it can be technologically demanding (on both students and staff), and it creates stress at a time when students' anxiety is likely at an all-time high. I'm particularly concerned about the equity implications; remote learning is already going to create advantages for those with more stable home situations—good technology, strong internet connectivity, and dedicated room for study and test-taking. High-stakes digital proctoring will exacerbate these inequities. There are also concerns with the high costs of such services, and the technological overhead: a significant portion of my time each term is spent helping students debug issues getting proctoring running, reviewing proctoring sessions, etc. As sophisticated as the systems are, they still demand a significant amount of time from the instructor and teaching assistants.

Digital proctoring is enormously valuable in programs where students know they are signing up for an online program in the first place and can self-assess their ability to participate in the necessary activities to succeed; that is not the situation we find ourselves in.

Three Kinds of Unfair Advantages

The need for trustworthy assessments does not go away, however. We need trustworthy measures of student ability, and part of that trust is believing that the student receiving credit is the one that the assessment is measuring. This guide gives some suggestions on how to design assessments you can trust without adding digital proctoring on top. More than anything, the focus of this guide is on creating assessments that will be especially fair and tractable for students who are not seeking unfair advantages; we do not want to force honest students through onerous additional steps just to also catch dishonest students. The best assessment design will be one where honest students are not even aware of what steps were taken to prevent academic dishonesty.

First, a couple assumptions to note. I'm only discussing timed assessments; assignments, projects, etc. obviously have a significant role to play, but the shift to remote teaching doesn't change the trustworthiness of those in the same way as students were already completing those outside the classroom. Second, I'm taking for granted that you likely have a learning management system that can administer timed assessments including some features like question pools; if you do not have access these systems, you may not be able to leverage all these ideas.

With that in mind, let's talk about how to build trustworthy remote timed assessments. There are three kinds of unfair advantages I feel that digital proctoring usually helps deter. They are:

- Consulting with disallowed materials
- Collaborating with current classmates
- Seeking aid from prior students, friends, etc. who are not presently enrolled in the class

Each of these advantages demands different approaches to resolution without proctoring, so let's take them one at a time.

Consulting with Disallowed Materials

This is the difference between an open-book and closed-book test. In-person, you can ensure students' desks are empty while they take the test. With digital proctoring, you can require room scans for students to show they do not have materials nearby, and then observe the proctoring data to ensure students do not seem to be consulting materials on their desks.

There's good news and bad news on this one. The bad news is that nothing you do in an unproctored assessment can ensure they don't consult materials you would have disallowed in a traditional testing environment. You can remind students about the institute honor code all you want, but you'll ultimately only be enforcing the rules on honest students. You can employ more scalable tools like a Lockdown browser, but that ultimately just advantages students with access to other devices.

The good news, though, is that this comes with its own built-in option: you are essentially forced to make your tests open-book, open-note, etc., but you then have the freedom to design a good open-book, open-note test. The nature of designing such a test will depend on your subject matter, but there are several ways you can ensure that your test still requires the desired subject matter knowledge.

Creating Good Open-Book Tests

There are many resources out there on the internet on creating good open-book exams, including [this guide from Indiana University](#) and [this guide from Faculty Focus](#). Exams in two of my four classes are open-book, open-note, etc., and I begin them from the perspective of wanting to test familiarity with a body of content. I will largely ask two kinds of questions: first, I'll ask questions whose answers cannot be found within the content itself, but rather whose answers are dependent on broader familiarity. For example, I might ask, "Which of the following statements would Langdon Winner likely agree with?" The statements themselves are not found within the body of the papers by Winner that we read, so searching the resource will be fruitless and waste time, but students familiar with the readings would pick out relatively quickly which views are commensurate with the author's.

Second, I will ask questions whose answers can be found within the content, but far more efficiently for those adequately prepared. For example, I might ask, "Which of the following are not principles of universal design?" Among the options might be real principles from other sets; particularly well-prepared students could pick out the answers quickly, but students who enter with a good working knowledge of the sets would be able to quickly jump into the principles of universal design and find the specific ones mentioned. Ultimately for my content, I am not concerned with the difference between "knows this immediately" and "can find the answer to this within 10 seconds".

Narrow time windows, which we'll come back to frequently in the next couple sections, play a strong role here, too: your assessments may be designed holistically, such that while every student can Google for the answer to question 1, that search will waste precious time they will need for later questions on the assessment. The reward for getting a problem right quickly is not just the points attached to that problem, but also the time that is now available for problems further along in the assessment.

There are many other recommended approaches here; open-ended questions that test more synthesis or analysis rather than rote recall are certainly good fits for limiting the usefulness of external resources, although they might not fit with all course content. Closed-ended questions can still be useful in an open-book setting through questions like, "Which of the following are examples of this concept?" and "Which of the following statements would this person likely agree with?" Answers to these will not be

found in the content directly, but rather derive from a broader understanding of the content that cannot be developed within the exam window.

Tips for Open-Book Assessments

So, to summarize these tips for designing good, timed, open-book, unproctored assessments:

- Develop questions whose answers are not present in the content, but whose answers are clearly derived from the content.
- Pay attention to vocabulary to minimize the role that text search tools can serve.
- Experiment with how long it would take to seek an answer to ensure proper timing.

With these tips, I believe you can design an assessment that not only is non-trivial, but actually that penalizes overreliance on external resources over internal knowledge.

Collaborating with Classmates

There are two challenges presented by classmate collaboration. First, there is the issue of shared information; even if an exam is not auto-graded live, a student taking the exam early in the window might be able to share information about what is on the exam with a classmate who takes the exam later in the window. The second problem is more live collaboration, where students take their assessments together rather than individually. Both these can exist in several forms, from a student emailing her friend information that was on the test to students emergently forming collaborative documents to share test data with larger groups.

The issue of students collaborating live is similar to the issue of seeking aid from non-classmates who have no inside knowledge about the test contents, so let's tackle the first one and leave the second one for the next section.

Your goal in designing this assessment is to minimize the extent to which students are able to share useful information with one another. This can furthermore be done in two ways: minimizing the extent to which the shared information is useful, and minimizing the extent to which the information can be shared rapidly enough to make a difference.

Minimizing Useful Information-Sharing

To minimize the extent to which the shared information is useful, your best option is to heavily utilize question pools. Question pools in a learning management system allow you to write lots of questions and then dynamically assign each student a random question. This could be done at the test level (a question bank of 100 questions from which every student receives 20) or at the section level (a test made of 20 questions, each drawn from a set of 5 possible questions) depending on the demands of your test and your content.

Writing lots of questions need not be overly onerous; the questions can be near-variations of one another. Oftentimes, near variations can be even more successful. Let's imagine we are designing a test of students' knowledge of the periodic table: our question pool could have questions, "What alkali metal has the lowest atomic number?" and "What alkaline earth metal has the lowest atomic number?" Now, one student takes the test and receives the first variation, and then tells his friend, "The question about alkali metal atomic numbers, the answer is Lithium." His friend then takes the test and sees a question with "alkaline earth metal" in the title, which sounds similar enough to "alkali metal" that he likely

selects the wrong answer. Some subject matter knowledge is necessary to figure out where to apply the information you received.

These sorts of near-variations can be both easier to write and particularly effective at targeting collaborators. The goal is to make it more difficult for someone unprepared for the test to identify whether the question is the same one they expected; if that identification is tougher than answering the question authentically, then the test carries some reassurance of integrity. Plus, they are near-variations, their difficulty is likely closer, avoiding the issue where one student may randomly draw a tougher-than-average test. Random ordering can help with this as well: rather than allowing answer-sharing on the order of "1 is C, 2 is B", they can further make it so some familiarity with terminology is necessary to recognize where the shared answer has landed.

Minimizing Rapid Information-Sharing

There is a time factor here as well, though; with enough time, any novice can likely determine if a question is identical. That plays into the second question above: minimizing the extent to which the information can be shared rapidly enough to make a difference. The key here is to stick to short time windows, both for starting the assessment and for the length of the assessment. It is very difficult to control whether material gets out of an unproctored exam; even if you use more scalable tools like Lockdown browser, students can still use their phones to quickly take pictures of exam questions. The goal, then, is to make such information minimally valuable.

You may accomplish this by tight windows, forcing students to take the quiz simultaneously as they would in person. Although they might be collaborating separately, a tight window combined with robust question pools would make it more difficult to try to balance the information overload of various slightly different questions compared with just answering the questions directly. This also addresses the problem of addressing the supply of information rather than the demand; if the window for taking an assessment is tight, then students must choose between pursuing their own score or helping their classmates. If a window is longer, students may (for example) take screenshots throughout the quiz and only bother to share them when they are done.

Toward this end, most learning management systems will allow you to set up a quiz so students may not go back to earlier questions; this prevents the case where a student may move through pursuing their own good grade, then circle back to help their classmates with their remaining time. There are, of course, trade-offs here: this may disadvantage students with slower internet connections if your assessment relies on images that may take time to download; but preventing students from jumping around can allow you to provide more generous windows, knowing that students cannot invest the extra time into returning to earlier questions to share details with classmates.

What I would love, actually, is an extension to the question pools feature that more intelligently generated exam variations based on time through the window; if an exam is open for an hour, then it should give students who start in the first 5 minutes one variation, the next 5 minutes another, etc., such that there are questions no one has seen before for those that start at the end of the window. But I digress into feature requests.

What About Answer-Sharing Sites?

The above two sections generally assume the goal is to deter somewhat synchronous collaboration; students taking the class at the same time. A different form of classmate collaboration would happen

over time, however; students post quiz questions and solutions to one of the many answer-sharing sites online who shall remain nameless.

Many of the approaches above double as deterrents for this kind of sharing as well. Tight time windows and the inability to jump around again force students to choose between their own score and helping others. A large question pool means that students must consult a large library of possible questions to copy the answer, ideally so large that it is difficult to search-and-find within the time confines of the test itself. Near-variation of questions further supports that as it requires students to devote significant attention to ensuring they're answering the right subtle variation on the same question. Shuffling answer options can further make direct copying difficult.

The alternative, of course, is to write all-new exams between terms, which is certainly an ideal option; it means you can proactively give students the previous terms' questions, rendering answer-sharing sites irrelevant. Writing all-new assessments each term is difficult in some fields, but parameterizing the questions can help out; some tools will even let you incorporate the parameterization into the exam creation, radically raising the number of possible exam permutations, especially on mathematical questions.

Open-ended questions coupled with plagiarism-detection tools can also serve as a deterrent, although those may not be compatible with all course material or grading workflows. Another component that can help instead is withholding answer keys altogether so that complete keys cannot propagate semester to semester. It goes without saying that keys should not be released until all students have taken the assessment, but you might go a step forward and keep that entirely secret. There is a pedagogical cost to that as wrong answer feedback is valuable for learning, but I believe that can be mitigated; you can instead give students an explanation of what they got wrong without providing the original question and answer. In my classes, for instance, students are given a breakdown of their score on each chapter covered by the test, and are further given text feedback like, "Langdon Winner would argue that..." In doing so, we do not share a straightforward key that someone else could use to select the answers directly, but we still provide the feedback to correct students' misconceptions (possibly even in a more palpable way).

Tips for Minimizing Classmate Collaboration

So in short, my recommendations here are:

- Use question pools to limit the extent to which information is useful when shared
- Use short time periods to disincentivize spending quiz time sharing and to minimize how long shared information is valuable
- Prevent jumping between questions to prevent students from investing extra time into sending information to their classmates

That said, there is a downside to tight windows: it creates pressure on students to be present and ready on your schedule rather than select a time that is right for them; this latter flexibility is often one of the benefits of online education, and here we lose it for the sake of integrity. We can mitigate this in part, though, by trading our long, rare, high-stakes assessments for short, frequent, low-stakes assessments. Not only is this better practice pedagogically, but it also means that a single bad quiz grade due to a technical error, home situation, or other problem of remote assessment is not nearly as impactful.

Not only can this mitigate some of the issues with these shorter time windows, but it also dovetails with my recommendations for how to combat the third form of collaboration.

Seeking Aid from Prior Students or Experts

The key to the second form of collaboration above was that someone else had insider knowledge of the contents of the exam, and so the goal is to obtain that insider knowledge. This third form of collaboration does not seek insider information from someone else, but rather seeks to use the expert knowledge of a former student or other more knowledgeable individual. Most often, this might be a friend or relative who is nearby and willing to help out; more nefariously, this might be someone hired online to masquerade as the student to complete the assessments.

The ultimate key to mitigating this form of collaboration is to radically increase the cost of obtaining this kind of help. By 'cost', I don't mean monetary (although that might be the form that it takes); rather, I mean that the investment necessary for a third-party collaborator to make a difference must be very large.

Short, Embedded Assessments

This brings us back to the idea of multiple short assessments. Asking your friend to sit with you during a single 3-hour final exam to help you out is not a big request. Asking your friend to help you with a 15-minute quiz three times a week for four weeks is a bigger request, without changing the amount of time actually spent in the assessment. Remote teaching gives you this remarkable ability to launch much more rapid assessments: they can be embedded in instructional material, graded automatically or speedily, and scoped to the appropriate size. With that, you could add enormously more assessment as well, which carries benefits to both pedagogy and trustworthiness.

A related approach to this, albeit one that might not work in every subject, is to create assessments where the question style or content is tightly coupled with details that are very specific to your version of a class. In one of my classes, students are required to read many academic papers on a subject; some of those papers would not be common knowledge even among practitioners in that area, and so bringing in an expert would not grant as big an unfair advantage was you might anticipate. You might put your own spin on certain vocabulary words or ask questions in format that you commonly use in your own lectures.

The goal of both of these approaches is to ensure even if the student gets help from someone else, that person basically has to take the class from scratch themselves; their expertise should either be minimally transferrable to your assessments (the second approach), or the work involved in applying their expertise to your assessments should be so high that it would be unreasonable to request it.

Deterring Paid Help

That brings us to the other potential risk of this kind of collaboration: paid help. The approaches above try to make it so that someone would not help out a friend for free because the work involved is tremendous; but that returns to our equity argument above. What about those that can simply pay for someone else to do the work for them? Remote classes make this significantly more feasible because you would not be paying someone to physically attend class in your stead, but rather just to log into a system.

Here, we generally must return to technological solutions again. Placing the assessments behind login walls means that to engage in this kind of cheating, students would need to share login credentials, a high-risk prospect especially when dealing with people who already are showing their willingness to engage in questionable behaviors. Many modern systems similarly check to see if students are accessing course materials from dramatically different IP addresses and flag them, although you may need to contact your own IT department to find out if that's an option in your system.

This makes it difficult for a student to find someone online to do considerable amounts of their work for them, but it does not mitigate having someone available in person to do so. But through these steps, we have narrowed down area of weakness to those students who can afford to hire a qualified individual to effectively complete an entire course for them while geographically nearby, a relatively daunting prospect; and one that itself is still addressable through features like class presentations and video chats with instructors or teaching assistants to discuss test answers.

Tips for Minimizing Aid from Outside Individuals

So, in short, my tips for minimizing the extent to which students can receive significant help from non-classmates (paid or as a favor) are:

- Prioritize frequent, short assessments over rare, long assessments
- Embed assessments into instructional material.
- Where possible, ask questions in a way that requires knowledge of your vocabulary or structure for teaching the content.

The overall outcome of these tips is that asking for aid from someone outside the class means effectively asking them to take the *entire* class for you, not merely to jump on and help with the occasional assessment. That's a far bigger request and less likely to be filled.

Summary

So, to summarize the suggestions I have for designing trustworthy unproctored assessments:

- Allow access to books, notes, etc. and design assessments with that in mind.
- Use question pools, especially with small variations on questions.
- Require students to begin the assessment within a short window of time.
- Require many short, low-stakes assessments (daily quizzes, for instance) rather than a small number of long, high-stakes assessments (quarterly exams, for instance).
- Require students to complete questions in order without backtracking.
- Imbue the questions on an assessment with specific details of your teaching vocabulary or structure.

Through these approaches, I feel it is wholly possible to create a trustworthy assessment strategy for your class that is natural, fair, and pedagogically beneficial for students who are authentically engaging with the material, but presents enormous challenges for students seeking to gain an unfair advantage by collaborating with classmates or outside experts.

Acknowledgements