

Restructuring Tests

Nathan Sanders

University of Toronto

As is common in North America, my large undergraduate courses have traditionally included major tests in the middle and at the end of the term. Barring accessibility accommodations and medical exceptions, all students used to write these tests in person, on paper, at the same time, in the same room, within a single two-hour testing window, with no or limited access to notes and books.

During the COVID-19 pandemic, “in person”, “on paper”, and “in the same room” were obviously impossible. Given the large number of my students who were international students who had returned to their home countries, “at the same time” and “within a single testing window” would have resulted in some students somewhere unfairly having to write a test late at night, in the middle of dinner, or before dawn. I am also opposed to invading my students’ privacy with proctoring software, so I wanted to rely on the honor system that they would use only authorized sources. However, given the reality that some students would inevitably cheat, I also wanted to minimize the opportunity for cheating, as well as the impact it could have on the overall distribution of grades, so that honest students would not be disadvantaged.

Finally, I wanted to minimize the work of the teaching assistants for my courses. As students themselves, they were dealing with similar academic difficulties as the undergraduate students. But as teaching assistants, they were also dealing with similar pedagogical challenges that faculty were facing in having to adapt to completely new ways of teaching and interacting with students. And as researchers, many of them were also in the midst of conducting research with human subjects, and that work had to fundamentally change during the pandemic. With all of these various practical, psychological, and emotional challenges, I felt that it was important to do what I could to shift as much of the course burden away from these teaching assistants.

With all of these considerations in mind, I completely restructured how testing worked in my courses. The most obvious change was a shift to online tests, which I implemented within our learning management system, Canvas.¹ I also chose a larger testing window (12 hours), to accommodate students in different time zones. The biggest change I made was to break up the two-hour test into four 30-minute parts. There were multiple reasons for this change. First, students’ schedules and home lives were radically different during the pandemic. Many of them lived at home instead of in campus residences, and along with this came increased

¹ Special thanks to Geoffrey Poore for his amazing text2qti Python script, available at <https://pypi.org/project/text2qti/>.

responsibilities to their households, as well as living with family members who were working at home or doing their own online schooling. All of this made it much more difficult for students to find an uninterrupted two-hour block of time when they could have a quiet space to work with consistent and reliable internet access. Breaking the test into four 30-minute parts gave them greater flexibility in scheduling so that they could work around their other responsibilities and scheduling conflicts.

The four parts had to be completed in order, but students could take as much time between parts as they needed. Thus, rather than having to cram everything in right before the test, they could focus on studying for Part 1 first, write it, then spend time refreshing their knowledge of the material in Part 2, and then after writing Part 2, they could use the washroom, have a snack, or even take a nap. With so many other stressors going on in the world and in their lives, many students reported that this structure was beneficial in reducing their testing anxiety, and it is a key component I plan to keep moving forward.

This structure was also designed to help combat cheating. Each part was drawn from large questions banks, so that each student got an overall unique version of the test. For example, Part 1 may have 15 total questions, with Questions 1–3 drawn randomly from a bank of 20 questions, Questions 4–7 from a bank of 15 questions, etc. Each bank was designed to cover the same content with questions of roughly the same difficulty. With only a 30-minute total window in which to complete each part, it would be difficult for students to work together; they would have very few identical questions, so they would effectively be trying to complete 60 minutes' worth of testing in 30 minutes. Since many of the questions were only subtly different, they would also have to waste time trying to spot these differences (or end up getting the wrong answer anyway). It would still be possible for some students to do Part 1 early and post their questions to a group chat, but the mechanics of doing so would still take precious time out of their limited 30-minute window, and other students would still need to notice any subtle differences and solve any new problems anyway. Ultimately, no approach to online teaching can completely eliminate cheating, but this method makes cheating more difficult, doing so solely through inherent design rather than invasive monitoring software. While I cannot determine how much this did in fact limit cheating, the overall average scores on these online tests were consistent with the test averages for previous in-person offerings of the courses. This suggests that the lack of monitoring was not enough to trigger a sufficient increase in cheating to skew the scores.

Finally, this method of putting the exams online, using open book/notes, and creating many smaller questions forced me to rethink many of the kinds of questions I used. For these tests, students were explicitly told they could use all of their notes and books, just like how most everything works in the real world. This meant that questions requiring rote memorization were simply no longer viable, and I (rightfully!) shifted to questions requiring more critical thinking and finding connections between concepts. I will admit, this was *hard!* Creating these questions took a lot of time, and if it were not for the circumstances, I likely would not have designed so many of these kinds of questions. But I am glad I did, as I now have large banks of these questions, and I will continue using them, and this testing structure, even once we return to the physical

classroom, for many reasons. First, despite the lack of monitoring, this method still seems robust against cheating. In addition, it allows for more flexible scheduling, both for instructors and for students. Because the tests are online, they can be scheduled outside of normal class hours, leaving more class time for discussing content. Many students also reported that they preferred this format, as it reduced their test anxiety and allowed them to fit the test around their schedule. Finally, the resulting test questions not only assess higher-level thinking, but they also minimize grading time, which helps reduce workload and allow more time to be spent on instruction rather than grading.