

Title

How to use academic and digital fingerprints to catch and eliminate contract cheating during online multiple-choice examinations: a case study

Authors

Meaghan Emery-Wetherell^{1,2,*} and Ruoyao Wang^{1,3}

* Corresponding Author

1 - School of Information at University of Arizona

2 – wetherellm@arizona.edu

3 - ruoyaowang@arizona.edu

Abstract

Over four semesters of a large introductory statistics course the authors found students were engaging in contract cheating on Chegg.com during multiple choice examinations. In this paper we describe our methodology for identifying, addressing, and eventually eliminating cheating. We successfully identified 23 out of 25 students using a combination of unique academic and digital fingerprints, and identified students who used Virtual Private Networks (VPNs) to protect their online identity. There were two forms of cheating - posting questions and waiting for responses from tutors and looking for questions that had already been solved. We found that 165 questions from these examinations were posted by 10 different students but that the most common form of cheating was searching for answers that had already been posted. This paper discusses these patterns of Chegg usage, the consequences of not catching cheating early on, and

how students reacted to being caught. Also provided are R and Python code that readers may use to identify cheating students in their own courses.

Keywords

Academic Integrity, Cheating, Multiple Choice, Contract Cheating

Introduction

The switch to online courses during the COVID-19 pandemic precipitated an increase both in online cheating and collective understanding of how online cheating is conducted (Bilen and Matros 2021; Jenkins et al. 2021; Janke et al. 2021). One particularly prevalent form advertised to students through constant advertisements and social media callouts is **contract cheating**. In essence, students pay a small sum to have someone else do their work for them. This may occur through use of a site like a paper mill, hiring a stand-in for an examination or the entire class, or looking for and finding test bank questions online (Clarke and Lancaster 2006; Walker and Townley 2012; Lancaster 2019; Awdry 2021). This form of cheating is pervasive and appears to be on the rise – while estimates vary and depend on the field, approximately 15% of students self-report anonymously on having engaged in contract cheating (Fendler, Yates, and Godbey 2018; Manoharan and Speidel 2020; Lancaster and Cotarlan 2021), a number which is rising (Newton 2018). Some studies also indicate that these self-reporting studies are not entirely accurate and that the percentage of students who actually cheat could be much higher than those who self-report cheating behaviors (Nowell and Laufer 1997; Awdry 2021).

High rates of cheating are detrimental to the academic process on many levels. If extreme enough, academic integrity violations could cause universities to potentially lose accreditation and then federal funding, rendering degrees worthless (Williamson 2019). There are also ethical

ramifications of students cheating, particularly in classes that will become a large portion of their career path. Some studies have demonstrated that students who are likely to cheat are also likely to commit other unethical behaviors in their field post-graduation (Carpenter et al. 2004; LaDuke 2013; Krueger 2014; Golden and Kohlbeck 2020), and so catching students as they cheat is beneficial both for schools and for subsequent career paths, even if the students themselves might not feel it is.

Online cheating can be difficult to detect, but previous studies have demonstrated that faculty can improve their chances of detection to 83% with training (Dawson and Sutherland-Smith 2019; Dawson 2020). Unfortunately, few professors and fewer adjuncts have work hours available to dedicate to the task of finding and identifying cheating students (Crossman 2019), which may lead to lower detection and report rates. Additionally, many professors do not believe that it could happen in their class or that it is as prevalent as it is throughout a student's career (Wolverston 2016a; 2016b; Dawson 2020). Making cheating identification as easy as possible is key for both expediting detection and also determining the true scope of the problem in each course.

The primary focus of research on cheating has been descriptive, with few case studies (Ahsan, Akbar, and Kam 2022). Those few that do exist primarily focus on written or coding questions (e.g., Manoharan and Speidel 2020). Yet open-answer questions are time intensive to grade and are not as frequently used in large courses, where multiple-choice examinations are more common. When offered virtually these examinations are easier to cheat on and harder to catch, as students often have overlapping questions and potential answers (Novick et al. 2022). That said, there are some best practices known to make multiple-choice tests more difficult to cheat on like randomized question order, the use of question banks, and randomized answers

(Clarke and Lancaster 2006; Cluskey Jr, Ehlen, and Raiborn 2011; Bilen and Matros 2021) all of which can create an **academic "fingerprint"** for an examination (Manoharan and Speidel 2020).

We present here a case study of contract cheating on Chegg.com in an online asynchronous introductory statistics class at the University of Arizona, including our method for creating academic fingerprints, detecting cheating, and identifying students. We also discuss the different reactions students had to being caught and some additional aspects of the academic integrity process at University of Arizona that may guide other instructors in when and how they pursue sanctions against cheating.

Methods

Question Banks & Examination Type

All examinations were multiple-choice, but the instructor (the lead author) used different approaches to the examinations in different semesters to attempt to curb cheating. These are summarized in Table 1 and discussed below.

For the first three semesters, students were allowed to retake the examinations and have their two grades averaged out. While this was intended to turn this assessment into a formative rather than summative experience, the increased access to test questions gave students ample opportunity to post their questions on Chegg. For the fourth semester, students were only allowed to take the examination once and the time was tightened considerably (Table 1, Figure 1). On the final examination in Spring 2022, the instructor also put in restrictions so that students had only 10 questions at a time and could not go back to their questions once their answers were submitted.

Students were warned repeatedly to not cheat via statements made by the instructor in lecture videos, virtual course announcements, and descriptions on the front page of the online examination. These instructions explicitly told students to work on their own, to not use contract cheating websites (with Chegg and Coursehero provided as examples), and that the act of finding and not reporting an answer key or completed examination question online would constitute cheating. None of these warnings eliminated cheating.

All examinations used question banks for each of the 50 questions on the examination. There were 2 to 6 alternates in each question bank. Question order and answer order were also always randomized, and students took their examinations at different times. This created a distinct and unique **academic fingerprint** of who got what questions in what order at what time.

One aspect of our learning management platform hindered our ability to identify students using their academic fingerprint. If students withdrew from the course, **all their data was retroactively deleted from the course platform.** This meant that if a student cheated on an examination but dropped the course before the cheating was detected, that data was no longer stored in the class. Fortunately, this data was backed up on the university servers - upon request Information Technology added the student's data back into the course for us to access it and include it our analyses.

Chegg.com Data

While there are many contract cheating sites, this study focuses on Chegg.com (Chegg Inc n.d.), which we found was commonly used in our class. To identify the students who cheated, we searched for examination questions from the four semesters included in our study. We then provided Chegg with a list of the website addresses on Chegg (also called Uniform Resource Locators, or URLs) that contained questions from our examinations, the approximate date we

thought they had been posted, and a signed document from our dean as per Chegg's investigation requirements. Chegg returned a spreadsheet that included when the question was posted, when it was answered, when it was viewed, and the Internet Protocol (IP) addresses, email accounts, and name of the account holders who posted or viewed a question. They also removed the posts from their site.

Some important features of Chegg.com at the time of manuscript submission are:

- It is free to see the questions that have already been asked, but answers are hidden until the student pays for an account.
- A paid account allows students to view an unlimited number of answers to these previously asked questions with no limits.
- A paid account also allows students to post 20 questions per month, with additional questions costing an additional fee.
- Chegg posts do not have public information about who posted them or viewed them, or when they were posted, viewed, or answered; but this information is retained by Chegg privately.
- Chegg tutors are known to post answers very rapidly, with most answers coming in less than 20 minutes (Manoharan and Speidel 2020).
- When asking questions students can copy-paste or take screenshots with their computer or phone, which are reliably translated into text.
- Chegg questions show up on Google searches. This means that an "Open Google" policy during examinations can allow students to enter their question into Google and see a link to Chegg if that question has previously been posted.
- Chegg's privately retained metadata only backdates for the last 120 days.

- If a person posts a question and it is not viewed after 120 days, it is removed from the system entirely.
- If a post continues to get viewed, it remains on the website even though the posting student's information is purged from the system after 120 days. The posting and answering time metadata is retained as is subsequent viewership.
- This means that while we were given abundant information on who posted and looked at questions, we only had IP addresses and emails for students who posted in the last 120 days, and only had viewership information for the last 120 days.

In Fall 2020, students were able to ask more than a single question per Chegg post and so most posts had upwards of 10 or 20 questions. We analyzed the Chegg data to determine if the number of questions per post changed over time and if it had any impact on how many students were able to find that post when viewing examination answers on Chegg. We used a simple linear regression in R to check for a relationship (R Core Team 2019).

Identification Processes

The identification procedure for all cases started out with the following set of actions:

- identify that cheating has occurred on Chegg (or a similar site)
- come up with a list of the post URLs and the exact questions as they were copied over to the cheating websites; also note the posting style (e.g. screenshots)
- submit a takedown and an integrity code investigation with Chegg to obtain posting times and viewing times
- contact our learning management software Desire 2 Learn (D2L) to ask for students who had withdrawn from the course to be put back into the course to access their examination data

- pull data for examination questions and answers, and examination attempts logs

This set of actions created all the data needed to identify students, as each test has questions and timing that provides a fingerprint even when a student did not have an IP address recorded or an email with their name in it. Posts that shared a similar posting style and time of posting were considered as being posted by the same person when IP addresses or emails were not retained.

Question Difficulty

To determine if students were more likely to post a difficult question than an easy question to Chegg, we evaluated question difficulty according to Bloom's Taxonomy (Bush, Daddysman, and Charnigo 2014). Bloom's Taxonomy is an educational framework that places learning objectives and assessment tasks into 6 different categories (Remember, Understand, Apply, Analyze, Evaluate and Create). Analyze, Evaluate and Create are considered higher-order learning objectives. Questions were assigned a category by the instructor according to the primary task required in the question. We then evaluated the number of questions in each category that were asked on Chegg compared to the number of each type of question available overall using a Chi Square test for goodness of fit in R (Cochran 1952; R Core Team 2019).

Student Demographics

Previous work has found inconsistent results with student standing, student major, and likelihood of cheating (e.g. Sheard, Markham, and Dick 2003; Marsden, Carroll, and Neill 2005; Wilkinson 2009). We collected student major and class standing from the University of Arizona's enrollment system to determine if particular majors or standings of students were more

likely to cheat. We also took overall grades and grades on the assignment in question to determine if there was a distinct grade average that was associated with cheating. These demographic questions were analyzed to test whether there were particular subsets of students more likely to cheat.

Appeals and Academic Integrity Process

The process for filing and appealing cheating cases at the University of Arizona is outlined as part of their Code of Academic integrity (University of Arizona 2022). One key aspect is that prior to filing charges and recommending sanctions a professor must offer the student a chance to set up a one-on-one faculty conference. In the meeting the professor presents their evidence and the student gets their first chance to appeal their case (they can later appeal the decision with the Dean's office). Student appeals and explanations followed a few consistent patterns (see Results).

The extent of cheating in past semesters had artificially inflated grades so that the average examination grade was a low B (Figure 1). Because of the high average the lead instructor had not adjusted grades to have a higher point score ('curving' the examination), thinking the grade improvement was the result of better understanding of material. As a result, cheating students punished students who had taken the test honestly. The sanctions were therefore assigned as follows:

- Students who posted questions after the examination or who viewed one or two questions were assigned an academic integrity workshop

- Students who posted questions during the examination, or who viewed more than one or two questions during the examination were given a retroactive failing grade in the course as well as the above.
- Students who lied or continued to obfuscate the extent of their cheating (see Results for details) were given a permanent note on their transcripts in addition to the above sanctions. This note read “Grade is the result of an Academic Integrity Violation.”

Results

Note: as grades, characteristics and information revealed during student conferences is protected by Family Educational Rights and Privacy Act (FERPA), we have hidden exact numbers to prevent small-group identification in our results and discussion.

Identification Methodology

Once the data was collected, we found there was some data that made identifying students far easier than others. Many students used their name in the email handle, and quite a few of them used their university email addresses. In some cases they had logged in with someone else’s account, and we had to double check that the name in the email was an enrolled student.

After that, the next most diagnostic piece of information was IP addresses. Very few students were using Virtual Private Networks (VPNs) to disguise their IP, which meant that even if a student had used a VPN it was fairly easy to point out that they were the only one whose IP addresses consistently varied during the examination.

In the case that IP addresses and emails did not identify students, we used the actual questions themselves. Each set of students received a unique set of questions – if they posted more than approximately 4 of them, it was an easy match to a single student. It was harder to identify students who viewed questions on Chegg using the unique question method, as many of the posts had more than one question involved. In this case we identified students using their examination timing first (comparing when they were taking their examination to when the potential account was viewing questions on Chegg). In the event there were multiple individuals taking their test at the same time, we then cross-compared questions in the post with questions on the examination for the students with the right timing.

These decisions are laid out in the flowcharts in Figure 2 and Figure 3. Figure 2 shows how to identify students who posted questions and Figure 3 shows how to identify students who viewed questions but did not post them. Both flowcharts use what we found to be the most diagnostic information first to identify students then move through less convincing and more difficult fingerprinting data to use. Scripts on how to process data appropriately in both Python or R are found in Supplementary Appendix I, as well as sample data.

Identification Success and Chegg Trends

We found evidence that 10 separate students posted questions to Chegg during their examination, and we were able to identify 8 of them (Table 2). The remaining 2 students only posted one or two questions and had done so in Fall 2020, past the 120-day limit for retaining Chegg data.

These 10 students posted 165 questions from the Midterm and the Final examinations as 64 separate posts on Chegg.com (so, 165 questions but only 64 URLs). Of these questions, 150 were unique – several students asked the question multiple times or asked the same question as another student. These questions represented 27% of the possible questions offered on 7 separate examinations. 50% were posted in Fall 2020, with numbers falling precipitously over time.

The number of posts changed over time (Figure 4A), which may reflect a change in Chegg's policies or that students hoped to hide their identity by posting questions separately. At the time of publication students now can only have a single question answered per post. Posts that had higher numbers of questions were associated with higher viewership during Fall 2021 (Figure 4B). Note that we could only evaluate viewers of questions posted in Fall 2020 or Spring 2021; Fall 2021 would have been simultaneous and students were unlikely to find the answers.

While viewership data was limited to Fall 2021, we found 15 students had viewed answers to questions on Chegg during their examinations (Table 2). This indicates approximately 10% of the course in Fall 2021 was cheating by looking for previously-answered questions. 13 of these students used their University of Arizona email or a public email address with their name in them. We were able to identify all 15 unique viewers, including those that did not have their names on the email address used.

A few accounts had University of Arizona emails belonged to students who were not in the class but had loaned their Chegg account to another student. The students who borrowed their friends accounts all stated that the account owner had been unaware of how they were using it.

Similar to Manoharan & Speidel (2020) we found the median response time for a question to be answered on Chegg was 21.8 minutes, with a minimum of 1.9 minutes. 75% of questions were answered within an hour of being posted.

Question Difficulty

Most of the question categories had approximately equal prevalence, but the most difficult category (Create) had the highest percentage of questions posted (Figure 4D). This difference was not statistically significant ($\chi^2 = 24, p = 0.24$), and the visual difference in Figure 4D may partially be because the examinations had fewer of these high-level difficulty questions in general.

Student Characteristics

68% (17) of the identified students involved were majors in our department and had this course as a mandatory prerequisite. 52% (13) of the identified students were seniors. The most common grade on the assignment cheated on was a B, and most students scored a single letter grade higher on the item than they did in the course (Figure 4C).

Appeals Discussions

The behavior of students caught posting questions versus students caught viewing questions was very different during the appeals and academic integrity process. Students who posted questions largely admitted they had cheated, though a few continued to contend that it was not them who had posted the questions.

Students who viewed the questions largely used their faculty conference time and the appeals process to contend that what they had done was not truly cheating. A few students denied the viewership data belonged to them. Several students admitted to viewing questions and that it was cheating, but stated that they thought the sanctions imposed were too extreme.

The rationales that students gave for their use of Chegg were:

- External stress related to COVID-19 or parental disapproval
- Other students were cheating as well
- The examinations were “open Google” and so using an online answer key (another person's filled out examination) was not cheating
- They had found the questions on Google, then followed them to Chegg (and therefore, it met the “open Google” criteria)
- They had paid for Chegg, but not for this purpose and therefore it did not constitute cheating
- They had used Chegg to look up similar questions but had not viewed their own, so did not know they were looking at an answer key

For this last explanation we created an additional set of computer scripts to determine if that was true. After all, looking for similar questions is not necessarily cheating on an open examination. However, we found that all students who were caught viewing Chegg data had at least one of their own examination questions written on the post that they were viewing and typically more than 4 of their own questions, indicating they had ample evidence to know it was an answer key.

Additionally, students who viewed the websites typically viewed them multiple times over the course of the examination. They would return to the same URL repeatedly, which often

correlated with them receiving that exact question on their examination at that particular time - a pattern we could see on their examination logs.

Very few students contended that it was not them who had posted or viewed questions. Those who did contended their innocence even in the face of the following evidence:

- The Chegg IP matched their home IP address, and nobody else had used that IP in the last two years
- The questions posted were the only ones they had missed on the examination
- They were the only one taking the examination at that time
- They were the only ones with those questions, or with those questions in that order

These students continued to double down regardless of the evidence presented to them.

48% of the students caught cheating appealed the sanctions with the Dean (Table 2) for the reasons listed above. This was most common with senior students who stated they were concerned about their graduation.

Discussion and Conclusions

Extent and Type of Cheating

There are many myths in academia about the extent and the reasons for cheating - like that that it is rarely done, only affects certain assignment types, or only occurs when the professor has failed to adequately prepare students for an assignment (e.g. T. Bretag 2017; T. Bretag and Harper 2017; Tracey Bretag et al. 2019; Dawson 2020). In reality cheating is

extensive and growing (Gray 2022). Certain departments have even higher proportions of cheating, like the nearly 80-90% of students found in some Engineering departments (Carpenter et al. 2004). The extent of contract cheating on Chegg was a disappointing surprise that is in line with changing trends. The best representation of this problem was Fall 2021, where we caught approximately 10% of the class looking at test questions on Chegg during the Midterm examination. We assume that prior semesters had similar degrees of cheating but were unable to catch it as Chegg removes data after 120 days.

Student Response

The reasons students cheat are varied. Previous studies have found a combination of personality traits and situational characteristics that are linked with cheating, like poor self-control, busy schedules, and any other internal cost-benefit analyses that swing in the favor of cheating (Kidwell and Kent 2008; Megehee and Spake 2008; Park, Park, and Jang 2013; Olafson, Schraw, and Kehrwald 2014; Yu et al. 2018).

Nearly half of the students in our case study denied wrongdoing or contested the sanctions assigned. Other studies have found that while students recognize hypothetical scenarios as cheating, they do not consider their own similar actions as cheating (Waltzer and Dahl 2022). Some students may rationalize their actions, or believe that effective cheating is a sign of intelligence (Kidwell and Kent 2008; Peterson 2019).

Other researchers have found an association between academic misconduct and the so-called “dark triad” of personality traits of narcissism, Machiavellianism and psychopathy (Smith, Emerson, and Maudlin 2021). Students with a history of academic misconduct are also more

likely to continue to behave unethically in the workforce (Carpenter et al. 2004; LaDuke 2013; Krueger 2014; Golden and Kohlbeck 2020). Several of our students continued to assert their innocence despite overwhelming evidence. While this by no means applies to all students, the fact remains that some students will continue to lie about cheating, and a student's denial of charges should not be an indication that they did not, in fact, cheat.

Who Cheats

We found cheating was most common amongst upper-division students in our major who were trying to graduate. Because student grades on assessments were not particularly different from their grades overall (Figure 4C), we do not recommend using student characteristics to determine if they cheat or not. Many students posted their entire test, resulting in no association between test question difficulty and whether it was posted or not.

Discouraging Cheating

We found a similar phenomenon to Manoharan & Spiedel (2020) in that most of the cheating was looking for questions that were already answered, with only a few students actively posting questions. The questions posted did not show any trends in difficulty and many students were posting almost their entire examination indiscriminately instead of selecting specifically difficult questions.

One potentially encouraging trend is that the number of students actively posting questions decreased dramatically once students were warned that we were filing sanctions and that posting copyrighted class material is illegal in addition to violating our academic integrity code. As research does suggest that knowing they will be caught and punished can deter some

students from cheating (e.g. Megehee & Spake, 2008; Olafson, Schraw & Kehrwald, 2014), identifying cheating when it happens and being clear about consequences can minimize cheating in and of itself.

Cheating did not disappear until we incorporated the additional measures as recommended by Cluskey et al. (2011), Bilen & Matros (2021) and others of shortening examination time. We accomplished this by using the “locked page” feature available in D2L where students cannot scan back and forth between questions. Students find short time limits stressful and believe that may increase cheating (Novick et al. 2022). However, we found the opposite – semesters with longer exam access presented more time for contract cheating while students with shorter access time did not have time to cheat. We used 5 pages with 10 questions each, with an hour of access – giving students 12 minutes per page (though students with accommodations had longer). As the median time for a response on Chegg is 21 minutes, locking pages like this both discourages students from posting as they are aware of a tighter time budget and also decreases the likelihood they will receive a response in time.

Shorter exam time does have consequences. The average of the tightened timeline examination was approximately 2% lower than the average of student’s first attempts in previous semesters (Figure 1). The median time of completion for the locked final examination was 55 minutes but not all students were fast enough to complete the exam. 4% of students turned in a partially completed examination (missing no more than 3 questions). Moving from a loosely timed formative assessment where students had a chance to learn from their mistakes and improve to a system where students got a single shot at an examination was a disappointing move away from the lead author’s educational goals. It proved a necessary move to combat extensive online cheating.

Other Recommendations & Value Added

Students will typically cheat when they have the opportunity, motivation, capability, and can rationalize their actions (Smith, Emerson, and Maudlin 2021). As seen here, contract cheating in one semester can carry forward to other semesters. Students can view previously posted questions very readily, so the longer it takes an instructor to catch cheating the more students will be able to cheat in future semesters.

As the majority of our students used contract cheating sites first to look for questions rather than posting themselves, catching and removing questions quickly may help dramatically reduce online cheating in and of itself. Furthermore, we found a number of our questions had been scraped from Chegg and posted on other sites like Numerade and CourseHero, where the process for requiring removal of copyrighted course materials is not nearly as clear and often is not carried out. So speedy identification of questions is key to minimizing future opportunities for students to cheat.

Conclusion

Cheating in a classroom has impacts beyond the immediate benefit to the student. It can compromise their understanding of key material, can affect whether an instructor modifies examination scores for other students, and can hide areas of weakness in content that an instructor might otherwise be able to address. As online contract cheating sites retain questions for long periods of time and these questions are then scraped and propagated online to other sites, it behooves an instructor to catch posted questions quickly.

As seen here, using question banks, randomized order, locked pages with short time limits, and being clear about the consequences and likelihood of capture can all minimize if not eliminate cheating. A powerful motivator for honesty is knowing the cost is not worth the benefit – and knowing that students get caught and punished even when trying to hide their tracks can only help increase the public perception that cheating is not worth it.

Catching and dealing with contract cheating is time-consuming. It was a minimum of 4 hours per student when we were still perfecting the methods in Supplemental Appendix I. Looking for, identifying, and discussing consequences with students still takes us approximately 1-2 hours per student. Though frustrating, this is considerably less time than it takes to rewrite an entire examination of question banks. Ultimately, catching and preventing cheating is a frustrating but omnipresent aspect of teaching, one that appears to be a growing service allotment as online cheating continues to rise.

Ethics Statements

The author's Institutional Review Board approved the research protocol (STUDY00001197) on 5-19-2022 as ethical and of minimal risk.

Funding

This research was conducted without external funding.

Author Contributions

M. Emery-Wetherell wrote the paper, taught the course, wrote the examinations, and conducted all meetings with students and the dean. R. Wang and M. Emery-Wetherell worked together to create the identification methodology and identify students. R. Wang provided all code in Python and M. Emery-Wetherell provided all code in R.

Acknowledgements

P. Jensen (University of Arizona) for his assistance and support throughout the process of filing academic integrity code violations. R. Ronald (University of Arizona) for her support throughout the appeals process. P. Biggs (University of Arizona) for his assistance in tracking down missing student data and IP addresses. Two anonymous reviewers provided helpful feedback to this article. The support staff at Chegg.com for providing student data, even though their site is used sometimes for cheating. The students who admitted to cheating and made following the code of conduct process significantly easier. Perhaps most importantly, students who did not cheat.

Figure and Table Captions

Figure 1:

A scatter and line plot showing the average examination scores across four semesters. The black line is the overall score, with the averages of retakes included, while the grey line shows only the average score on a student's first examination attempt. The dashed line shows when a new final examination was written to deter cheating, and the dotted line shows where both examinations were rewritten and adjusted to avoid cheating (see Table 1 for adjustment details). The overall

examination scores show a steady increase until the examinations were re-written, causing a nearly 20% decrease in average midterm score from Fall 2021 to Spring 2021.

Figure 2:

A flowchart of the decision-making process when trying to identify a student who has posted a question on Chegg using data from Chegg and from D2L. Test is used here as a shorthand for examination.

Figure 3:

A flowchart of the decision-making process when trying to identify a student who has been viewing questions on Chegg using data from Chegg and from D2L. Test is used here as a shorthand for examination.

Figure 4:

A four-panel figure showing key results. Panel A shows the distribution by semester of how many questions students asked per post, showing a shift in the number of questions per post. Panel B shows the relationship between the number of questions per post and a student's likelihood of seeing that particular post when searching for answers on Chegg. The linear regression in panel B is statistically significant ($p < 0.001$) with a moderate R^2 of 0.45. Panel C shows the letter grade improvement on the examination in question compared to their overall course grade. Panel D shows the overall percentage of available questions in that category that were posted to Chegg (not the percentage a student got on an examination).

Table 1.

Student population and variations on examination rules made over the four semesters. Student population shows the total enrollment, the percentage that passed and the percentage that passed with an A in the course. All grade data has been updated to remove students who were given a retroactive failing grade as the result of cheating. The variations on examinations including rewriting entirely new tests (changing the examination version), no longer allowing a second attempt (retake) and locking pages. The time limit shown is minutes allowed per attempt, and number of days the examination was available to students is recorded under Open Days.

Table 2.

Overall numbers of students who posted questions, viewed questions, and used other students accounts to view questions. Columns show the number of distinct students found cheating, the number that we were unable to identify, the number of identified students who filed appeals, and the semesters when the behavior was recorded. The minimum number of students in a course was 129 in Fall 2020; the maximum was 157 students in Spring 2021.

Supplemental Appendix 1:

Provided are scripts in both R and python that work with D2L data and Chegg data. Data from other learning platforms like Canvas should be adaptable to this format, and fake data is provided that shows patterns of cheating as seen in the class. Details on the scripts and files are provided in the ReadMe file.

References

- Ahsan, K., S. Akbar, and B. Kam. 2022. "Contract Cheating in Higher Education: A Systematic Literature Review and Future Research Agenda." *Assessment & Evaluation in Higher Education* 47 (4): 1–17.
- Awdry, R. 2021. "Assignment Outsourcing: Moving beyond Contract Cheating." *Assessment & Evaluation in Higher Education* 46 (2): 220–35.
- Bilen, E., and A. Matros. 2021. "Online Cheating amid COVID-19." *Journal of Economic Behavior & Organization* 182: 196–211.
- Bretag, T. 2017. "Good Practice Note: Addressing Contract Cheating to Safeguard Academic Integrity." Tertiary Education Quality and Standards Agency, Good Practice Note: Tertiary Education Quality and Standards Agency.
- Bretag, T., and R. Harper. 2017. "Assessment Design Won't Stop Cheating, but Our Relationships with Students Might." *The Conversation*, May 11, 2017.
<https://theconversation.com/assessment-design-wont-stop-cheating-but-our-relationships-with-studentsmight-76394>.
- Bretag, T., R. Harper, K. Rundle, P.M. Newton, C. Ellis, S. Saddiqui, and K. van Haeringen. 2020. "Contract Cheating in Australian Higher Education: A Comparison of Non-University Higher Education Providers and Universities." *Assessment & Evaluation in Higher Education* 45 (1): 125–39.

- Bush, H.M., J Daddysman, and R. Charnigo. 2014. "Improving Outcomes with Bloom's Taxonomy: From Statistics Education to Research Partnerships." *Journal of Biometrics & Biostatistics* 5 (4): 6155–80.
- Carpenter, D.D., T.S. Harding, C.J. Finelli, and H.J Passow. 2004. "Does Academic Dishonesty Relate to Unethical Behavior in Professional Practice? An Exploratory Study." *Science and Engineering Ethics* 10: 311–24.
- Chegg Inc. n.d. "Honor Code Investigation." Accessed June 9, 2022.
<https://www.chegg.com/honorcode/investigation-request>.
- Clarke, R., and T. Lancaster. 2006. "Eliminating the Successor to Plagiarism? Identifying the Usage of Contract Cheating Sites." In *Proceedings of 2nd International Plagiarism Conference*, 19–21. Citeseer.
- Cluskey Jr, G. R., C.R. Ehlen, and M.H. Raiborn. 2011. "Thwarting Online Exam Cheating without Proctor Supervision." *Journal of Academic and Business Ethics* 4 (1): 1–7.
- Cochran, W.G. 1952. "The X² Test of Goodness of Fit." *The Annals of Mathematical Statistics* 24 (3): 315–45.
- Crossman, Katherine. 2019. "Is This in My Contract?: How Part-Time Contract Faculty Face Barriers to Reporting Academic Integrity Breaches." *Canadian Perspectives on Academic Integrity* 2 (1): 32–39.
- Dawson, P. 2020. "The Prevention of Contract Cheating in an Online Environment." Australian Government: Tertiary Education Quality and Standards Agency (TEQSA).[Article].

- Dawson, P., and W. Sutherland-Smith. 2019. "Can Training Improve Marker Accuracy at Detecting Contract Cheating? A Multi-Disciplinary Pre-Post Study." *Assessment & Evaluation in Higher Education* 44 (5): 715–25.
- Fendler, R.J., M.C. Yates, and J.M. Godbey. 2018. "Observing and Deterring Social Cheating on College Exams." *International Journal for the Scholarship of Teaching and Learning* 12 (1): 4.
- Golden, J., and M. Kohlbeck. 2020. "Addressing Cheating When Using Test Bank Questions in Online Classes." *Journal of Accounting Education* 52: 100671.
- Gray, B.C. 2022. "Ethics, EdTech, and the Rise of Contract Cheating." In *Academic Integrity in Canada*, 189–201. Springer, Cham.
- Janke, S., S. C. Rudert, A. Petersen, T. M. Fritz, and M. Daumiller. 2021. "Cheating in the Wake of COVID-19: How Dangerous Is Ad-Hoc Online Testing for Academic Integrity?" *Computers and Education Open* 2: 100055.
- Jenkins, B.D., J.M. Golding, A.M. Le Grand, M.M. Levi, and A.M. Pals. 2021. "When Opportunity Knocks: College Students' Cheating Amid the COVID-19 Pandemic." *Teaching of Psychology*, Online First.
- Kidwell, L.A., and J. Kent. 2008. "Integrity at a Distance: A Study of Academic Misconduct among University Students on and off Campus." *Accounting Education: An International Journal* 17 (S1): S3–16.
- Krueger, L. 2014. "Academic Dishonesty among Nursing Students." *Journal of Nursing Education* 53 (2): 77–87.

- LaDuke, R.D. 2013. "Academic Dishonesty Today, Unethical Practices Tomorrow?" *Journal of Professional Nursing* 29 (6): 402–6.
- Lancaster, T. 2019. "Social Media Enabled Contract Cheating." *Canadian Perspectives on Academic Integrity* 2 (2): 7–24.
- Lancaster, T., and C. Cotarlan. 2021. "Contract Cheating by STEM Students through a File Sharing Website: A Covid-19 Pandemic Perspective." *International Journal for Educational Integrity* 17 (1): 1–16.
- Manoharan, S., and U. Speidel. 2020. "Contract Cheating in Computer Science: A Case Study." In *2020 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*, 91–98. IEEE.
- Marsden, H., M. Carroll, and J.T. Neill. 2005. "Who Cheats at University? A Self-Report Study of Dishonest Academic Behaviours in a Sample of Australian University Students." *Australian Journal of Psychology* 57 (1): 1–10.
- Megehee, C.M., and D.F. Spake. 2008. "The Impact of Perceived Peer Behavior, Probable Detection and Punishment Severity on Student Cheating Behavior." *Marketing Education Review* 18 (2): 5–19.
- Newton, P.M. 2018. "How Common Is Commercial Contract Cheating in Higher Education and Is It Increasing? A Systematic Review." In *Frontiers in Education*, 3:67. Frontiers.
- Novick, P.A., J. Lee, S. Wei, E.C. Mundorff, J.R. Santangelo, and T.M. Sonbuchner. 2022. "Maximizing Academic Integrity While Minimizing Stress in the Virtual Classroom." *Journal of Microbiology & Biology Education*, e00292-21.

- Nowell, C., and D. Laufer. 1997. "Undergraduate Student Cheating in the Fields of Business and Economics." *The Journal of Economic Education* 28 (1): 3–12.
- Olafson, L, G Schraw, and N. Kehrwald. 2014. "Academic Dishonesty: Behaviors, Sanctions and Retention of Adjudicated College Students." *Journal of College Student Development* 55 (7): 661–64.
- Park, E-J., S. Park, and I-S. Jang. 2013. "Academic Cheating among Nursing Students." *Nurse Education Today* 33 (4): 346–52.
- Peterson, J. 2019. "An Analysis of Academic Dishonesty in Online Classes." *Mid-Western Educational Researcher* 31 (1).
- R Core Team. 2019. "R: A Language and Environment for Statistical Computing." Windows. Vienna, Austria: R Foundation for Statistical Computing. <http://www.R-project.org/>.
- Sheard, J., S. Markham, and M. Dick. 2003. "Investigating Differences in Cheating Behaviours of IT Undergraduate and Graduate Students: The Maturity and Motivation Factors." *Higher Education Research & Development* 22 (1): 91–108.
- Smith, K.J., D.J. Emerson, and S. Maudlin. 2021. "Online Cheating at the Intersection of the Dark Triad and Fraud Diamond." *Journal of Accounting Education* 57.
- University of Arizona. 2022. Code of Academic Integrity. <https://deanofstudents.arizona.edu/policies/code-academic-integrity>.
- Walker, M., and C. Townley. 2012. "Contract Cheating: A New Challenge for Academic Honesty?" *Journal of Academic Ethics* 10 (1): 27–44.

- Waltzer, T., and A. Dahl. 2022. "Why Do Students Cheat? Perceptions, Evaluations, and Motivations." *Ethics & Behavior*, 1–21.
- Wilkinson, J. 2009. "Staff and Student Perceptions of Plagiarism and Cheating." *International Journal of Teaching and Learning in Higher Education* 20 (2): 98–105.
- Williamson, M. 2019. "Contract Cheating and Academic Integrity in Higher Education: What Can Universities, Governments and Quality Assurance Agencies Do to Understand, Prevent and Respond to the Challenge?" *Kuwait International Law School Journal, Special Supplement* 4 (1): 219–74.
- Wolverston, B. 2016a. "The New Cheating Economy." *The Chronicle of Higher Education*, August 28, 2016. <https://www.chronicle.com/article/The-New-Cheating-Economy/237587>.
- . 2016b. "In a Fake Online Class with Students Paid to Cheat, Could Professors Catch the Culprits?" *The Chronicle of Higher Education*, October 26, 2016. <https://www.chronicle.com/resource/how-students-cheat-in-a-high-t/6122>.
- Yu, H., P.L. Glanzer, B.R. Johnson, R. Sriram, and B. Moore. 2018. "Why College Students Cheat: A Conceptual Model of Five Factors." *The Review of Higher Education* 41 (4): 549–76.