Using Data Analytics to Detect Possible Collusion in a Multiple Choice Quiz Test

Michael Lang[0000-0001-7961-2285]

School of Business & Economics, NUI Galway, Ireland
michael.lang@nuigalway.ie

Abstract. This paper reports on the experiences of using an on-line MCQ test to assess students' knowledge for a postgraduate module. Because of the COVID-19 pandemic, the test was taken in a remote non-proctored environment. Although it was executed under timed conditions with students seeing questions in a randomised order, algorithmic analysis of the response patterns suggests that collusion occurred during the test. Practical implications for assessment design and administration are discussed.

Keywords: IS education; MCQ tests; academic integrity; data analytics.

1 Introduction and Background

Assuring academic integrity on computer programming assessments is a perennial challenge for university educators [1], and has been further exacerbated by the extraordinary unplanned constraints imposed by the COVID-19 pandemic [2, 3]. Information systems programmes across Europe have seen major increases in intake in recent years, many of these additional students coming from outside the EU. These larger, culturally diverse classrooms have already forced lecturers to rethink their learning design. On top of this, teaching staff have now been catapulted (by the ongoing pandemic) into an on-line environment that for many is a first-time venture into terra incognito.

Prior literature on academic integrity in computing and information systems has indicated that there may be behavioural differences across gender, national culture, maturity and English language proficiency [4, 5]. The latter is not so much of an issue in programming modules. There are several other factors that can contribute to plagiarism and academic dishonesty, including time pressure, required effort, social norms, low self-efficacy, personal morals and conscientiousness, awareness of rules, awareness of detection techniques, and perceived risk of detection [6]. Interestingly, Harris *et al* [7] found no differences in self-reported behaviour of students within on-line learning environments and those in traditional environments, as reported in previous studies [8].

This paper reflects on the experiences of assessing a postgraduate Database Systems module at an Irish university in the 2020/'21 academic year. The method of assessment was a Multiple Choice Quiz (MCQ) test which was administered on-line in a timed examination. The research questions were:

- Does nationality affect a student's tendency to cheat in an on-line non-proctored examination?
- Does gender affect a student's tendency to cheat in an on-line non-proctored examination?
- Does age affect a student's tendency to cheat in an on-line non-proctored examination?
- Does peer network affect a student's tendency to cheat in an on-line non-proctored examination?

The structure of this paper is as follows: Section 2 describes the teaching and assessment case study environment, Section 3 analyses and discusses the findings, and Section 4 presents conclusions and implications for teaching practice.

2 Description of Case Study

"Database Systems Development" was taken by 167 students spread across three separate programmes. The class was 36% female and 64% male, made up of 12 different nationalities, of which Ireland (38%), India (44%) and China (11%) were the biggest cohorts, with the remainder (7%) coming from Nigeria, Cameroon, USA, Mexico, Brazil, France, Ukraine, Pakistan and Indonesia. The median age was 25.4 years. The majority of the international students arrived in Ireland prior to semester but a few remained in their home countries. Under normal circumstances, this module would have been taught in a lecture hall. However, because of the COVID-19 pandemic, it was taught on-line across 11 weeks with a regular scheduled slot on Microsoft Teams each week. In previous years, the end-of-semester examination was in the form of an invigilated Multiple Choice Quiz (MCQ) test, taken in an examination centre. For the 2020/'21 academic year, this was replaced by an on-line MCQ test taken in the student's place of residence.

It was initially proposed that the examination would be subject to remote proctoring. However, in view of legitimate concerns raised by students about privacy, unreliable internet connectivity, possible power outages or other technical difficulties, and additional stress at an already difficult time, it was decided not to proceed with remote proctoring. Instead, following Nguyen *et al.* [10], candidates were required to declare in advance of the test that they had read the rules "and pledge, on my honour, to fully abide by them". To allow for possible connectivity problems, students were assured that there would be a degree of leniency with the examination duration, with a 30 minutes grace period before shut down. This was also intended to offset time pressure issues, a factor that has shown up in prior studies as a major contributing factor in cheating [6].

All students were issued an email one week in advance of the examination, clearly setting out the format and rules that would apply. This included a notice that algorithms would be used to detect suspicious activity and a strict warning was served not to communicate or collude with others during the test. Again, this was intended to mitigate contributory factors identified by Moss *et al* [6].

2.1 Format of Examination

The test consisted of 50 multiple choice quiz (MCQ) questions, each with four possible options of which one and only one was correct. MCQ tests are prone to cheating, and indeed this had been previously experienced in this module, with students using pre-devised signals to communicate to each other in examination halls. Problems when going on-line were therefore anticipated, especially with no proctoring and students living together or connected virtually. It was assumed, given the absence of remote proctoring, that students would use their notes, even if asked not to do so. The test was therefore designed on this basis and students were told that it would be “open book”. However, so that they would be under no illusions, they were advised that the questions would be “set in such a way that you need to understand concepts and apply your knowledge, not merely memorise material”. As such, given the time constraint of the examination (two hours), if students had to resort to consulting their notes frequently, they would place themselves under pressure and this was made known to them.

The first half of the test examined knowledge of database design concepts. Most of these questions put forward four assertions and the candidate was required to use his/her understanding of the theory to decide which assertion was true or false. A few other questions in this section presented data modelling scenarios with four possible choices, again requiring the candidate to consider options and make a decision, thus requiring higher order thinking skills at the Analyse→Synthesise→Evaluate end of the Bloom *et al.* [9] taxonomy. Such higher order MCQ questions have been found to be effective in science education [10].

The second half of the test was based on knowledge of the Structured Query Language (SQL). The questions used a database schema that students were given in advance and asked to print out. The lectures and course exercises also used this same database so students were familiar with it and were expected to have practiced upon it. Instead of being asked to write SQL code from scratch, students’ knowledge was assessed by other means: (1) “fill in the blanks” questions that required them to complete an SQL query by inserting the correct missing words in the correct order, (2) being asked to evaluate four different ways of solving a problem and selecting which of these ways is valid or not valid, (3) inspecting code snippets and being asked to detect which line(s) contain errors, if any, and (4) inspecting code and being asked what output it would generate.

The quiz was administered using Microsoft Forms, which required users to authenticate themselves using their university Office365 account credentials. Several days in advance, they were given a “mock” test with 10 sample questions and the precise rules and instructions that would apply.

2.2 Student Performance

In advance of the test, there was quite a degree of trepidation amongst students about it being MCQ format and being a timed examination (all their other modules, with just one exception, used take-home assignments). After the test, the feeling was much more positive. One student said that *"My overall experience with the exam was really nice, I enjoyed each and every question asked, they were certainly tricky but it's just required to use a bit of brain and knowledge. I would say that if a student has gone through your class notes that you've shared all along and did the exercise queries, they can achieve an excellent score in it".* The performance scores on questions were very strong, with just 8 of the 50 having less than 50% correct. The median across the test was 79% correct, ranging from 10% on a question that only a very few students got right up to 100% on one question that they all answered correctly. Although this level of performance might suggest that the test was easy with a low level of discrimination amongst the possible answers, it can be alternatively explained by the fact that many of the students had some prior knowledge of databases and had obtained first class honours undergraduate degrees.

3 Analysis and Discussion

With a MCQ test in a class of high performers, it can be difficult to detect collusion because the majority of students will consistently pick the correct answers. Instead, it can be revealing to compare patterns of incorrect responses [11]. In theory, the probability of two students randomly picking the same incorrect answer from two four-option questions is (3 matching pairs) / (9 possible combinations) = 0.33. Therefore, if students have five matching incorrect answers, the probability is 0.33^5 which is less than a 1 in 250 chance. However, this is not a random process; incorrect answers do not have equal probability of being chosen because some may be easier to eliminate than others. For the purposes of this exercise, it was therefore assumed that a threshold of 8 similar incorrect responses would be used.

The MCQ test responses were exported from Microsoft Forms into a MySQL database, where they were transposed and a paired-list of possible collusion suspects was generated. Additionally, data on age, nationality and gender were linked to the responses. It was assumed that students took the test at their registered addresses so those details were converted into latitude and longitude coordinates using Google Maps and also imported.

The highest number of shared incorrect responses was 16. The statistical probability of this occurring (based on equal weighting for each option) is 1 in 50 million so, if going just by mathematics, it would be an absolute certainty that this pair of students colluded. However, these two students were in two different Masters programmes, live thousands of kilometres apart and, to the best of the examiner's knowledge, have never communicated with each other. Both students failed the test and had several other incorrect responses.

Out of a class of 167 students, 70 (42%) had eight or more similar incorrect answers to other students. Of those 70, several had eight or more similar incorrect answers in common with more than one student. The results of the similarity analysis were then exported from MySQL and imported into a Neo4J database, where graph queries were executed to detect suspicious clusters. This revealed several mini-networks of students who had numerous identical incorrect responses (see Fig. 1).

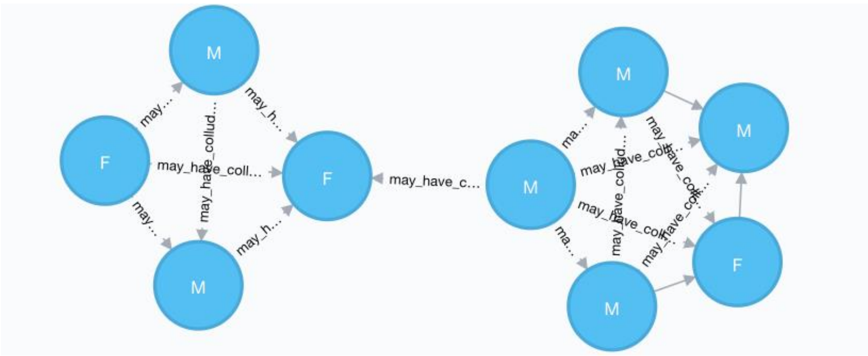

Fig. 1. Examples of mini-networks of collusion suspects detected by Neo4J graph query.

Not surprisingly, six clusters were found amongst cohabiting students, including one group of four and another of three. Amongst students that were geographically distributed, several clusters of four or more students with similar answer patterns were also identified. Interestingly, these clusters were mostly of the same nationality, with Irish students and Indian students not mixing but rather forming their own groups. Chinese students barely featured at all, seemingly keeping to themselves (which might be because most of them were not in Ireland). Overall, 63% of Irish students and 31% of Indians were found to have suspicious response patterns. On the face of it, this seems to suggest that nationality does make a difference. However, self-efficacy may be a moderating factor here because many of the Indian students had substantial prior experience in database design and development. For other nationalities, the numbers are too small to be statistically meaningful.

While the presence of these clusters cannot be taken as absolute proof that collusion took place during the test, and perhaps the excuses of "I and my friend both got it wrong because we studied together" or "we both picked the second most likely options" can explain some of the similarity, it does seem quite likely that, even though students were seeing the questions in a different random order, some of them were in communication with each other during the test and comparing responses.

As regards gender and age, no differences were observed; males and females were just as likely to engage in this behaviour, as were younger students and mature students.

4 Conclusions and Implications

This work-in-progress paper set out to explore if the factors of nationality, gender, age and peer network affected a student's tendency to cheat in an on-line non-proctored

MCQ examination situation. Gender and age were found not to have any impact, but nationality and peer networks did appear to be at play with several cases of suspected collusion. Because the evidential base was quite weak and the cost of investigating suspicious cases is high, none of these cases were brought forward for further action.

The lessons learned from implementing an on-line MCQ test with randomised order suggest that this could work effectively if administered in an invigilated examination centre. It could also function well if the on-line class were geographically distributed, unlike the very unusual situation that occurred during COVID-19 where international students travelled abroad, only to become engaged in distance education while actually living on or near campus with other international classmates. However, if students are permitted to take tests on-line in non-proctored environments, it seems likely based on the experiences reported herein that many of them will engage in behaviour to cheat the test. The use of constraints such as tighter time limits, no facility to navigate back to review answers, and screen-at-a-time design have all been suggested in the literature, but these could just drive students towards more innovative ways of beating the system. Further study is required on the phenomenon of "classroom culture", and how students from diverse backgrounds coalesce to form on-line peer learning communities with shared social norms, ethical values and honour codes.

References

1. Sheard, J., Butler, M., Falkner, K., Morgan, M., Weerasinghe, A.: Strategies for maintaining academic integrity in first-year computing courses. Proc. ACM Conf. Innovation and Technology in Computer Science Education, Bologna, Italy, pp. 244-249, (2017).
2. Senel, S., Senel, H. C.: Remote Assessment in Higher Education during COVID-19 Pandemic. International Journal of Assessment Tools in Education 8(2), 181-199 (2021).
3. Pappas, I. O., Giannakos, M. N.: Rethinking Learning Design in IT Education During a Pandemic. Frontiers in Education, 6:652856, (2021). doi: 10.3389/feduc.2021.652856.
4. Adair, J. H., & Linderman, J. L.: Computer ethics and academic honesty: Developing a survey instrument. ACM SIGCSE Bulletin 17(1), 93-98 (1985).
5. Atkinson, D., Nau, S. Z., Symons, C.: Ten years in the academic integrity trenches: Experiences and issues. Journal of Information Systems Education, 27(3), 197-208 (2016).
6. Moss, S. A., White, B., Lee, J.: A systematic review into the psychological causes and correlates of plagiarism. Ethics & Behavior 28(4), 261-283 (2018).
7. Harris, L., Harrison, D., McNally, D., Ford, C.: Academic Integrity in an Online Culture: Do McCabe's Findings Hold True for Online, Adult Learners?. Journal of Academic Ethics 18(4), 419-434 (2020).
8. McCabe, D. L., Treviño, L. K., Butterfield, K. D.: Cheating in academic institutions: A decade of research. Ethics & Behavior, 11(3), 219-232 (2001).
9. Bloom, B.S., Engelhart, M.D., Furst, E.J., Hill, W.H., Krathwohl, D.R.: Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain. David McKay, New York (1956).
10. Nguyen, J. G., Keuseman, K. J., Humston, J. J.: Minimize Online Cheating for Online Assessments During COVID-19 Pandemic. Journal of Chemical Education 97(9), 3429-3435 (2020).
11. Ercole, A., Whittlestone, K. D., Melvin, D. G., Rashbass, J.: Collusion detection in multiple choice examinations. Medical Education 36(2), 166-172 (2002).