DATA MINING OF ONLINE QUIZ LOG FILES: CREATION OF AUTOMATED TOOLS FOR IDENTIFICATION OF POSSIBLE ACADEMIC MISCONDUCT IN LARGE STEM COURSES

Emma Spanswick[1], Marzena Kastyak-Ibrahim[1], Corey Flynn[1], Sarah Elaine Eaton[1], Nancy Chibry[1]

[1] *University of Calgary, Canada*

KEY WORDS

data mining, academic integrity, academic misconduct, automation, post-secondary, Learning Management System (LMS)

The rapid shift to remote course delivery in March 2020 during the COVID-19 pandemic presented a significant challenge for administering fair and reliable student assessments. Students and instructors who were not specifically trained for an online learning environment were forced to adapt and transition to remote mode of teaching and learning. In most cases, remote delivery implied reorganization of student assessments to online frameworks. To help with this transition, the University (located in Canada) provided faculty members with the list of the features available in the online learning management system (LMS) to consider when setting up an online assessment.

STATEMENT OF THE PROBLEM

For the University of Calgary where the majority of STEM courses and their components had been delivered in person prior to COVD-19, the online tools available were not particularly optimized for the large volumes of student assessment that were moved to online delivery, specifically the final exams of relatively large (800+ student) first- and second-year courses. The problem of practice that informed our study is that students were using online file-sharing sites to rapidly share test answers. Because our university opted not to use any kind of electronic or remote proctoring software, we wanted to see if we could find a way to identify violations of academic integrity using the tools we had available through existing university resources, namely the LMS.

LITERATURE REVIEW

Breaches of academic misconduct are common in higher education. Over more than half a century, repeated studies have shown that upwards of a third of undergraduate students engage in acts of academic misconduct every year, with results being similar in both the United States (Bowers, 1966; McCabe, 2016) and Canada (Christensen Hughes and McCabe, 2006). In addition, only a small portion of the academic misconduct identified by instructors is reported (Bowers, 1966; Coren, 2012; MacLeod and Eaton, 2020; Nadelson, 2007).

A particular topic of concern in recent years has been contract cheating (Clarke and Lancaster, 2006). Bretag et al. (2019) identified seven types of student academic outsourcing behaviours: (1) buying, selling or trading notes; (2) providing a completed assignment to another student; (3) obtaining a completed assignment from someone else; (4) providing exam assistance; (5) receiving exam assistance; (6) taking an exam for someone else; and (7) arranging for someone else to take one's exam (p. 1839).

Inappropriate or unauthorized student file-sharing has been highlighted by researchers as a growing concern (Rogerson, 2014; Rogerson and Basanta, 2016), with particular concern focused on commercial enterprises who profit from students who pay to download files, which can include completed assignments, notes, and other course materials (Wolverton, 2016). The background discussions for this study included an inquiry into the availability of course content online. We found material directly related to this specific course on four commercial file-sharing sites. We have intentionally opted not to name these companies here, though we wanted to highlight that it was easy for the research team to find copies of course assignments and other assessments online in a matter of minutes.

STUDY DESIGN

We collected data generated by the university's learning management system in form of the reports generated from internal log of the system. Each interaction between the student and the D2L Quiz is logged by the internal systems. For example, the entry into the quiz is logged, as are all page navigations. The complete information for all students who took a quiz can be downloaded as a single Excel file. This file contains a column with each student's name, attempt number, date and time stamp of each interaction, and a description of each interaction (event) (for example moving to a next page or saving a response). The last column provides the IP address indicating the location from where the quiz was accessed.) We then used statistical data mining techniques to look for connections between students' individual quiz timings for viewing and saving of randomized questions. Data mining refers to the process of extracting meaningful information from often vast amounts of raw data (e.g., Coenen, 2011 and references therein). This can be through statistical connections between various pieces of information, or through more advanced artificial intelligence frameworks such as neural networks. In all cases, data (often in very large quantities) is mined for information relevant to specific topics. Data mining techniques are used extensively in research communities that rely on large data sets and are often foundational to observational sciences (those that collect vast quantities of data from distributed sensors) such as environmental science and space science. Our LMS includes an Analytics module that mines student data within its system to provide analytical insights for student success.

RESULTS

Within the context of the course described here, rates of academic misconduct showed a dramatic increase from 2018 through 2020. Data from the final assessments (administered in person) from previous years were compared to our findings from online assessments. Our analysis shows that compared to the 2019 course offering, there was a threefold increase in academic misconduct cases. The significance of this work is that, although we make no claims about differentiation between an increase in the rate of detection versus actual misconduct cases, we found that the tools developed in our study here have dramatically increased our ability to identify and provide evidence for breaches of academic integrity. In the Winter 2020 course offering, there were 33 cases (4% of students enrolled in the course) identified as potential academic misconduct ones because of a student completing the exam in less than 25% of the time and/or answering at least one challenging question (often requiring calculations) correctly in under one minute which was impossible for even a professor to do. All the case were investigated by the Associate Dean and three of them (9%) were dismissed. In the previous three years, the highest

percentage of cases was 2% (mostly associated with using unauthorized material during the final exam) of the enrollment and more than 10% of the cases were dismissed during the investigation by the Associate Dean. In the Summer 2020 course offering, the case detection rate was 9% and none of the cases was dismissed. We will share the technical details of our results in our presentation.

SIGNIFICANCE

We believe we have developed a new method of data mining LMS activity logs to identify suspicious activity during exam/quiz administration. We make not claims that suspicious activity on exams equates with misconduct. Instead, through this study we examined variables such as quiz duration, IP addresses, question duration and question order/timing to flag students who performed outside expected norms. In a large (800+) undergraduate course, our data mining flagged ∼10% of test takers, half of which were pursued for formal investigation of academic misconduct. These findings show that data collected by most LMSs can be used to flag student misconduct and can assist in the development of fair and resilient evaluation methods even in an online environment. We wish to share the results of our study so others can replicate it at their own institutions as a viable alternative to paying for surveillance technology such as electronic proctoring software.

REFERENCES

Bowers, W. J. (1966). *Student dishonesty and its control in college.* (Doctor of Philosophy). Columbia University, New York.

Bretag, T., Harper, R., Burton, M., Ellis, C., Newton, P., Rozenberg, P., van Haeringen, K. (2019). Contract cheating: A survey of Australian university students. *Studies in higher education*, 44(11), 1837–1856. doi: https://doi.org/10.1080/03075079.2018.1462788

Christensen Hughes, J. M., and McCabe, D. L. (2006). Academic misconduct within higher education in Canada. *The Canadian Journal of Higher Education*, 36(2), 1–21. Retrieved from http://journals.sfu.ca/cjhe/index.php/cjhe/article/view/183537/183482

Clarke, R., and Lancaster, T. (2006). *Eliminating the successor to plagiarism: Identifying the usage of contract cheating sites.* Paper presented at the Second International Plagiarism Conference, The Sage Gateshead, Tyne and Wear, United Kingdom.

Coenen, F. (2011). Data mining: past, present and future. *Knowledge Engineering Review*, 26(1), 25–29.

Coren, A. (2012). The theory of planned behaviour: Will faculty confront students who cheat? *Journal of Academic Ethics*, 10(3), 171–184. doi:10.1007/s10805-012-9162-7

MacLeod, P. D., and Eaton, S. E. (2020). The paradox of faculty attitudes toward student violations of academic integrity. *Journal of Academic Ethics.* doi:10.1007/s10805-020-09363-4

McCabe, D. (2016). Cheating and honor: Lessons from a long-term research project. In T. Bretag (Ed.), *Handbook of Academic Integrity* (pp. 187–198). Singapore: Springer Singapore.

Nadelson, S. (2007). Academic misconduct by university students: Faculty perceptions and responses. *Plagiary*, 67–76. Retrieved from http://hdl.handle.net/2027/spo.5240451.0002.008

Rogerson, A. M. (2014, June). *Detecting the work of essay mills and file swapping sites: some clues they leave behind.* Paper presented at the 6th International Integrity and Plagiarism Conference, Newcastle-on-Tyne, United Kingdom.

Rogerson, A. M., and Basanta, G. (2016). Peer-to-peer file sharing and academic integrity in the Internet age. In T. Bretag (Ed.), *Handbook of Academic Integrity.* Singapore: Springer.

Wolverton, B. (2016, September 2). The new cheating economy. *The Chronicle of Higher Education*, 63(1), A40+.