Title	Assessing database development skills using an on-line MCQ: Reflections on test design and academic integrity
Author(s)	Lang, Michael
Publication Date	2021-06-23
Publication Information	Lang, Michael. (2021). Assessing database development skills using an on-line MCQ: Reflections on test design and academic integrity. Paper presented at the 7th International Conference on Higher Education Advances (HEAd'21) Virtual conference, 22-23 June.
Publisher	Editorial Universitat Politècnica de València
Link to publisher's version	http://headconf.org/head21/head21-papers-published/index.html
Item record	http://hdl.handle.net/10379/17287

Assessing Database Development Skills using an On-line MCQ: Reflections on Test Design and Academic Integrity

Michael Lang
School of Business & Economics, National University of Ireland Galway, Ireland.

Abstract

This paper reports on the experiences of assessing a large postgraduate module in Database Systems using an on-line MCQ test under COVID-19 conditions with no remote proctoring. While the technology and the "higher order thinking" assessment worked well, analysis of response patterns suggested that, despite the use of randomised questions, time pressure and an honour pledge, a substantial number of students formed mini-networks during the test and colluded with each other to discuss the answers. No differences by gender or age were observed as regards tendency to engage in this behaviour, and domestic students were just as likely as international students to do it. Notably, domestic and international students seemed to stay apart, colluding almost exclusively with others of their own nationality.

Keywords: Academic integrity; Plagiarism; Culture; MCQ tests; On-line examinations; Database Systems.

1. Introduction

The practical difficulties of assuring academic integrity during the COVID-19 pandemic are by now widely acknowledged (Asgari *et al*., 2020; Senel and Senel, 2021). They are further compounded when cultural issues (Hayes and Introna, 2005; Maxwell *et al.*, 2008) and complications of on-line learning (Harris *et al.*, 2020) are added into the mix alongside the many other individual factors than can contribute to plagiarism (Moss *et al.*, 2018).

This paper reports 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. Although the technology worked very well, problematic issues surfaced during the analysis of response patterns, suggesting that collusion may have taken place during the test.

2. Teaching Context and Assessment Method

"Database Systems Development" was taken by 167 students spread across three separate programmes, commencing in September 2020. 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 and took up accommodation on or near the university campus, but a few remained in their home countries. Under normal circumstances, this module would have been taught in a lecture hall and assessed by means of an invigilated end-of-semester test in an examination centre. However, eight days into the academic year, it was announced by management that all teaching and learning would be conducted on-line for the remainder of the semester, with the sole exception of essential campus-based activities. This module was therefore taught on-line across 11 weeks, with a regular 1.5 hour slot on Microsoft Teams each week.

Because of public health guidelines, no examinations could take place in centralised indoor venues. The alternatives available under university emergency regulations were either an assignment, which students should complete in their own time and submit by a specified deadline, or a scheduled on-line examination of limited duration. For this module, the latter option was chosen as it was deemed to be more appropriate and trustworthy. It was initially proposed that the examination would be subject to remote proctoring. However, students were quite anxious about this and many of them expressed concerns:

"I live in an apartment with several others. It is not a quiet environment and it's beside a busy road so if I leave my mic turned on, it could disturb other students during the test."

"I have an unreliable internet connection and share it with other members of my household, including children doing home schooling because of COVID. What will I do if my webcam takes up all the bandwidth and the test freezes?"

"I don't have a webcam on my laptop. It is quite old and I cannot afford to buy a new one at the moment. I am worried that the on-line test might time out on me."

"It's very stressful being locked up at home and studying like this. I don't like the idea of being remotely watched in my own bedroom, it freaks me out, it's just too much."

Similar problems have been observed in other institutions (Asgari *et al*., 2020). In view of these legitimate issues about privacy, internet connectivity and additional stress at an already difficult time, it was decided not to proceed with remote proctoring. Instead, following Nguyen *et al.* (2020), 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.

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.

The test consisted of 50 multiple choice quiz (MCQ) questions, each with four possible options of which one and only one was correct. 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.* (1956) taxonomy. Such higher order MCQ questions have been found to be effective in science education (Nguyen *et al.*, 2020).

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. Students were told that they were not permitted to use an SQL interpreter during the examination. However, it was assumed that some of them would do so; accordingly, the queries to be evaluated contained logical errors (not just basic syntax errors) and used parameters that would return no results against the sample data set that had been provided. Therefore, if students attempted to cheat by asking the SQL interpreter to find the answer, it would be of little help.

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. The purpose of this was to familiarise students with the interface and also to detect any technical issues that might arise. Based on student feedback on this mock test, two changes were made to the software implementation.

Firstly, instead of using radio option buttons, they were changed to checkboxes. This was because it was discovered that a student's selected answer could inadvertently be changed if he/she used the down-arrow key to navigate to the next question; by using checkboxes, this problem was solved. Although it was then permissable to select multiple answers, students were told that only one option was correct and multiple selections would be a zero mark.

Secondly, students suggested that all of the questions should have a "mandatory" validation check so that they accidentally did not miss or skip any. Literature on the design of MCQ tests suggests that negative marking should be used to dissuade candidates from random guessing (Roberts, 2006). By enforcing a mandatory rule on questions, it might encourage guessing. However, on any other assessment format, students are encouraged to attempt all parts and are not penalised for writing something that is a bit of a guess; therefore, the fairness of negative marking on MCQ tests is debatable. Zhao (2006) demonstrates that the likelihood of passing a MCQ test of four-option questions by guessing is highly improbable for tests with large numbers of questions. Furthermore, on all other forms of assessment, students are given partial marks for providing incorrect answers that are nearly right. On a well-designed MCQ test, the correct answer may not be too far removed from very similar incorrect answers. Based on this rationale, it was decided not to use negative marking for

incorrect answers, and also to award partial marks for questions where there was a statistically skewed distribution of incorrect responses, similar to the approach suggested by Grunert *et al.* (2013).

Guidance on the avoidance of cheating on MCQ tests recommends to present questions in a randomised order, one at a time, and not to allow back tracking. In this test, the questions and options were randomised but all questions were presented on a single screen and back tracking was permitted. This was again informed by student suggestions; as one student put it by email, *"In an exam, I sometimes skip past tricky parts and come back to them later, and I also like to check back over all my answers at the end of the test if I have time, so I hope we will be allowed to do this in the Databases MCQ test"*.

3. Findings and Discussion

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). In anonymous feedback captured at the end of teaching but before the test, one of them commented that *"the decision to have MCQ as a final test was a bit disappointing. I was expecting some sort of project where we could put into practice what we have learned for the module"*. However, another remarked that *"it was made clear why we would be doing an MCQ so I am happy to be doing this"*. Students were provided with a test bank of preparatory exercises to work on, but without the carrot of credit going directly for this, many of them did not feel incentivised to complete the exercises. 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.

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 (Ercole *et al.,* 2002). 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. Mogull (2004) uses a different basis for computation of probabilities, based on the number of students who correctly answer any given question. This approach is flawed in that it assumes that if 9 out of 10 students were to get a question right, that the probability of any given student getting it wrong would be 10%; but for a weaker student amongst stronger peers, this does not hold. Furthermore, it omits to consider the effect of two or more students who share the same misunderstandings because they study together. For the purposes of this exercise, it was therefore assumed that a threshold of 10 similar incorrect responses would be used, or 8 similar in the cases of persons who shared accommodation with one or more classmates (which was about a third of all students).

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 about 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, 35 had ten or more similar incorrect answers to other students. Of those 35, a few had ten 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. 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 and those of other nationalities barely featured at all, seemingly keeping to themselves (which might be because most of them were not in Ireland). 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. Of course, it should be borne in mind that COVID-19 is a great leveller, with everybody feeling the pressure and perhaps behaving out of normal character.

4. Conclusions

MCQ tests are prone to cheating (Manoharan, 2019), and indeed this was 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. 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.

The comparison and clustering algorithms that were used to identify collusion suspects in this study are not complex but demand technical know-how to implement. Further time is then required to follow-up individually on each suspected breach. This places a very substantial overhead on examiners if all cases are to be pursued. Preventative measures must be the way to go on this rather than costly detection. It is likely that many of the teaching and assessment practices which were hurriedly adopted during COVID-19 will be refined and retained. Further study is required on the phenomenon of "classroom culture", and how students from diverse backgrounds coalesce to form on-line learning communities with shared social norms, ethical values and honour codes (Shafaei *et al*., 2016).

References

Asgari, S., Trajkovic, J., Rahmani, M., Zhang, W., Lo, R. C., & Sciortino, A. (2020). An Observational Study of Engineering Online Education During the COVID-19 Pandemic. *arXiv preprint arXiv:2010.01427*. doi:10.35542/osf.io/ursmb.

Bloom, B.S., Engelhart, M.D., Furst, E.J., Hill, W.H., & Krathwohl, D.R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain*. New York: David McKay.

Ercole, A., Whittlestone, K. D., Melvin, D. G., & Rashbass, J. (2002). Collusion detection in multiple choice examinations. *Medical Education*, 36(2), 166-172.

Grunert, M. L., Raker, J. R., Murphy, K. L., & Holme, T. A. (2013). Polytomous versus dichotomous scoring on multiple-choice examinations: development of a rubric for rating partial credit. *Journal of Chemical Education*, 90(10), 1310-1315.

Harris, L., Harrison, D., McNally, D., & Ford, C. (2020). Academic Integrity in an Online Culture: Do McCabe's Findings Hold True for Online, Adult Learners?. *Journal of Academic Ethics*, 18(4), 419-434.

Hayes, N., & Introna, L. D. (2005). Cultural values, plagiarism, and fairness: When plagiarism gets in the way of learning. *Ethics & Behavior*, 15(3), 213-231.

Manoharan, S. (2019). Cheat-resistant multiple-choice examinations using personalization. *Computers & Education*, 130, 139-151.

Maxwell, A., Curtis, G. J., & Vardanega, L. (2008). Does culture influence understanding and perceived seriousness of plagiarism?. *International Journal for Educational Integrity*, 4(2), 25-40.

Mogull, R. G. (2004). A Device to Detect Student Cheating. *Journal of College Teaching & Learning*, 1(9), 17-22.

Moss, S. A., White, B., & Lee, J. (2018). A systematic review into the psychological causes and correlates of plagiarism. *Ethics & Behavior*, 28(4), 261-283.

Nguyen, J. G., Keuseman, K. J., & Humston, J. J. (2020). Minimize Online Cheating for Online Assessments During COVID-19 Pandemic. *Journal of Chemical Education*, 97(9), 3429-3435.

Roberts, T. S. (2006). The use of multiple choice tests for formative and summative assessment. In *Proceedings of 8th Australasian Conference on Computing Education* (pp. 175-180).

Senel, S., & Senel, H. C. (2021). Remote Assessment in Higher Education during COVID-19 Pandemic. *International Journal of Assessment Tools in Education*, 8(2), 181-199.

Shafaei, A., Nejati, M., Quazi, A., & Von der Heidt, T. (2016). 'When in Rome, do as the Romans do': Do international students' acculturation attitudes impact their ethical academic conduct?. *Higher Education*, 71(5), 651-666.

Zhao, Y. (2006). How to design and interpret a multiple-choice-question test: a probabilistic approach. *International Journal of Engineering Education*, 22(6), 1281-1286.