Element of surprise and its efficacy in maintaining exam integrity in online exams

Bibek Bhatta
Working paper version: March 2023

***Abstract**:*

In view of off-site assessments being adopted by Higher Education Institutions (HEI) especially in response to Covid-19, questions can be raised regarding the extent of possible unhealthy practices employed by students when sitting for their online exams. Such unhealthy practices could take various forms including working in groups during individual assessments, accessing resources beyond what is allowed, presenting someone else's work as their own, etc. In this context, we examine the impact of element of surprise in preserving exam integrity when common monitoring protocols or other paid outsourced services cannot be implemented. Employing a three-pronged approach of comparing images, comparing text-pattern and comparing multiple choice question (MCQ) response pattern across students, our results show that introducing surprise by assigning students to random versions of exams can be helpful in thwarting possible malpractices. Though the benefit in this regard for MCQ is limited, the positive impact is more pronounced for image-sharing practices and text-sharing (copy-paste) practices. Our findings also show that active collaboration during online exams tend to take place in pairs.

Keywords: academic integrity; online assessment; textual analysis; image analysis; pattern-matching;

1. Introduction

Covid-19 pandemic has impacted our way of life including the way we teach, learn, and assess (Brammer & Clark, 2020). In the education sector, online learning became the norm to some extent during the pandemic. In the UK and abroad, many institutions including Higher Education Institutions (HEIs) resorted to online/off-site exams, in lieu of the traditional 'in-person' exam, because of the existing social distancing rules in place. Against the backdrop of Covid-19 pandemic, the issue related to academic integrity is likely to have brought additional challenges as many higher education institutions (HEI) adapted to the situation by conducting teaching and assessments online.

In view of off-site assessments being adopted by Higher Education Institutions (HEI) especially in response to Covid-19, questions can be raised on the extent of possible unhealthy practices employed by students when sitting for their online exams. Such unhealthy practices could take various forms including working in groups during individual assessments, accessing resources beyond what is allowed, presenting someone else's work as their own, etc. As the online exams by its very definition employ the utility of internet to provide questions and submitting platforms, it can be double-edged sword in that the same internet facility can be used to search up for answers on the internet. Though the level of such unhealthy practices is a matter of conjecture at this point, concerns are beginning to be voiced on how exam integrity can be preserved without question while students are taking exams off-site. For example, one of the professional accounting bodies with international reach has recently sought reassurance from UK universities on how integrity of the exam is being upheld during the Covid pandemic. Hence, it is reasonable to believe that HEIs should take extra steps to ensure that integrity of exams is being preserved. In some cases, external services can be secured and such exam monitoring activities can be outsourced, but this may not always be possible due to short notice and resource constraints.

Considering the benefits that arise from preserving exam integrity (Löfström et al., 2015), this paper examines if introducing element of surprise in online exams helps to preserve exam integrity. In this context, this paper examines the impact of element of surprise in preserving exam integrity when common monitoring protocols (like those available in on-site exam setting) or other paid outsourced services cannot be implemented. Employing a three-pronged approach of comparing images, comparing text-pattern and comparing multiple choice question (MCQ) response pattern across students, our results show that introducing surprise at short notice by assigning students to random versions of exams can be helpful in thwarting possible malpractices. Though the benefit in this regard for MCQ is limited, the positive impact is more pronounced for image-sharing practices and text-sharing (copy-paste) practices. Our findings also show that active collaboration during online exams tend to take place in pairs.

The rest of the paper proceeds as follows: Section 2 presents the research method; Section 3 provides and discusses the results. Concluding remarks and limitations are presented in Section 4.

2. Literature review

Academic dishonesty is known to be a serious problem among college students (Crown & Spiller, 1998). For a long time, academic institutions throughout the world have struggled to address the issue of academic integrity in the student body (Lin & Wen, 2007). In light of the online mode of assessments conducted by HEI recently owing to the Covid-19 pandemic, it can be argued that this has brought additional challenges to for the preservation of academic integrity.

3. Research method

This paper examines the introduction of element of surprise in exams for Masters level students of an anonymous university with the aim of finding if such elements of surprise can be helpful in lowering potential collaboration among students during online exams. For this purpose, we examine four different cohorts of students across two subjects: Subject A and Subject B. Each of these cohorts have 142 students on average. In all cases, students were given two hours to write their answers, and additional 15 minutes to upload their answer sheets to the virtual learning environment (VLE) using their university credentials. Students were required to submit their answers in word-processing document (Word), template of which was provided to them in downloadable form right before the exam.

We provide brief details below on the way the exams were conducted, so as to introduce element of surprise to two of the cohorts.

3.1. Subject A

For Subject A, the first cohort (Cohort A1), consisting of 127 students, were given the same exam paper for online exam; whereas the second cohort (Cohort A2) (with 179 students) were given "different" versions of the exam. Cohort A2 were told three days prior to the exam that they would have to choose the exam version based on the last digit of their six-digit student number (which ranged from 0-9). We call these versions "Student versions". Even though the A2 students were assigned to ten different student versions of the exam ranging from Version 0 to Version 9, all the exam versions – unbeknownst to the students – were, in effect, the same (except for the Version number). A1 sat for the exams in May 2020 and A2 sat for their exams one year later in May 2021. Both the exams were similar in

structure with predominantly numerical questions but with the requirement of some discursive element.

3.2. Subject B

Similarly, for Subject B, the first cohort (Cohort B1), consisting of 122 students, were given the same exam paper for online exam; whereas the second cohort (Cohort B2) (with 141 students) were given different versions of the exam. A stricter element of surprise was introduced to B2 cohorts (compared to A2 cohorts) in that they were not told in advance how the different exam versions would be assigned to them. Though B2 students were informed in advance that they would be assigned to different versions of the exam paper, the basis of such assignment was not revealed to them prior to the exam. This was to avoid active collaboration beforehand. At the start of the exam, students were then informed that they had to select only one version based on the second last digit of their six-digit student number. Students were advised not to download any other version other than the unique version that they were assigned to; and that if they downloaded any other version accidentally such irrelevant versions had to be deleted immediately. This was to ensure that students did not compare one version of the question with another version. B1 sat for the exams in Jan 2021 and B2 sat for their exams one year later in Jan 2022. Both the exams were similar in structure with 25 MCQ and other questions predominantly numerical in nature. Based on the second last digit of their student number, all 141 students (of Cohort B2) were assigned to one of the 10 versions (from version 0 to version 9) of the exam. Unbeknownst to the students, versions 0, 2,and 4 were the same; similarly, versions 1, 3, and 5 were the same; and versions 6 and 8 were the same; and likewise, versions 7 and 9 were the exact copies of each other. As such, though the students were made to believe that there were 10 different versions of the exam, in reality there were only four different versions that they were effectively assigned to. We refer to these four versions of exams as the “Main versions”. For B2 cohort, where the questions were in fact different, slight variations in the key inputs of the questions were introduced in the questions. Not all the questions in the different versions of the exam were different; but when the questions were different, the differences were subtle. For instance, if the tax rate in one of the questions in a given version was 17%, the corresponding question in a different version could have a tax rate of 19% with everything else being the same. An example of a

question that looks similar but is slightly different in shown in Appendix 1 with the differences highlighted in circles.

The number of students assigned to different versions of the exams for B2 is shown in Table 2.

For illustrative purposes, the four cohorts across the two subjects A and B are presented in Figure 1.

Insert Figure 1 here

3.3. Exam structure and comparability

Students could enter their answers in a number of ways in the Word document: for questions requiring numerical calculations, students could either type their answers in the provided space in the Word document, or they could write down the answers in a physical notebook and then take a picture of their answers and paste the pictures in the Word document. Additionally, they were also allowed to do the calculation in another computer package (e.g spreadsheets like Excel) and then take a screenshot of the solution and paste the screenshot in the corresponding space of the Word answer sheet. For MCQs (in Subject B), they could simply type one of the four given options for each MCQ; At any rate, students were required to submit only one Word document and the document could contain text along with pictures or screenshots.

3.4. Hypotheses development

Exam integrity could be compromised in various ways in off-site exams when there is no direct supervision. Such malpractices could take the form of copying from the internet, copying from local sources (e.g distributed class materials), copying from each other, outsourcing to external parties, etc. Considering that the questions were mainly numerical in nature created specifically for the given exam, it is unlikely that the students would find

ready-made answers to the questions on the internet or in the study materials. Hence, if the students are assumed to be engaging in malpractices, it is most likely that they are doing so in the form of copying from each other or helping out each other.[1]

When all students in a given cohort are given the same set of questions (i.e. Cohorts A1 and B1), it is arguably easier to cheat since students inclined to cheat can collaborate with any other willing individual within that cohort. However, when students are assigned to different versions of exam papers (Cohorts A2 and B2), students inclined to cheat would find it relatively difficult to find a willing collaborator sharing the same version of the question; as such the overall tendency to cheat is expected to decline. Hence, it can be expected that overall cheating declines when students are assigned different versions of exam papers. This leads us to our first hypothesis H1 as follows:

H1: Higher proportion of Cohort A1 (and B1) will engage in collaboration compared to Cohort A2 (and B2).

When students are assigned to different versions of exam papers giving them advance notice on the allocation process (like in Cohort A2), the element of surprise could be mitigated to some extent as students have ample time to discuss with each other and find out other likely collaborators who share the same exam version. This means that those students with the intention of engaging in unethical practices during the exam will try to find another willing candidate who shares the same version of the exam as themselves for the purpose of collaboration during the exam. This leads us to hypothesize as follows:

H2: Students within Cohort A2 will collaborate more with other students who share the same Version of the exam paper.

On the other hand, when students are kept in the dark as to how they will be assigned to different exam versions (like in B2), they are less likely to engage in collaborative effort

[1] It is also possible that students might pass on the question through internet to someone outside their academic institution, and the external partner would then prepare the answers and email the students. However, we ignore this line of enquiry.

beforehand. As such, B2 cohort is likely to witness a more pronounced reduction in collaborative practice vis-à-vis B1 cohort, as compared to reduction in collaborative practices in A2 cohort vis-à-vis A1. This leads us to our next hypothesis H3 as follows:

H3: Compared to prior cohort, reduction in collaboration in B2 will be more pronounced than in A2 cohort.

Regarding Cohort B2 where exam version allocation method was kept secret until the start of the exam, it is likely that students inclined to collaborate with others found it difficult to find like-minded individuals. Trying to find another individual sharing the same second-last digit of the student registration number and who is also willing to collaborate for the exam would have been infeasible, given the limited time allowed to finish the exam. Hence, the level of collaboration within the same Student Version for cohort B2 is expected to be significantly less than that for cohort A2. Based on this, we propose our next **hypothesis H4** as follows:

H4: The likelihood for cohort B2 students collaborating among students writing the same version of exam paper is significantly lower than that for cohort A2.

On the other hand, if the students in B2 cohort willing to engage in unethical practices were indeed able to collaborate with each other, the answer contents within the groups of students sharing the same student version of the exam should be more similar within their respective groups compared to inter-group contents. Since we do not yet know whether B2 cohort students were able to engage in ethical practices or not, we propose a competing hypotheses H4_a, referring to this as ‘cheating hypothesis’ and state as follows:

H4_a: Cheating hypothesis - For Cohort B2, answer contents within given groups are more similar to each other compared to inter-group contents.

3.5. Methodology for comparing similarity within cohorts

This paper employs three broad techniques to examine the similarity of answer papers within a given cohort of students. These three techniques are centred around the comparison of three distinct components of answer sheets, namely, answer pattern in MCQs, images, and texts. We discuss these three approaches in turn below.

3.5.1. Answer pattern comparison for MCQ

One approach of examining whether students are copying from each other, in the context of MCQ (in Subject B), is to examine the extent to which their answers match amongst each other. Say two students Adam and Eve have marked their first three answers to MCQ as A,A,B and B,D,C respectively. Since none of the answers match with each other, it is unlikely that they are copying from each other. By the same token, if a third student Sue has marked the first three answers as A,A,B, which aligns exactly with what the first student (Adam) has marked, this raises the possibility that the first (Adam) and the third (Sue) students were likely collaborating together during the exam. If the students are indeed collaborating during the exam, this kind of practice would be more pronounced for students sharing the same Student Version of the exam paper. We use this answer-matching pattern for entire set of MCQ for all students and calculate the similarity score for each student pair. For each student pair, we assign a value of 1 for each matched answer and 0 otherwise. Hence, since there are 25 MCQs, each student pair can have a maximum 'matched-score' of 25 (if all answers match with each other) and a minimum matched-score of 0 (if none of the answers match between the two students in that pair). If the students are indeed engaged in copying from each other, the matched-score would be higher among students sharing the same Student Version.

3.5.2. Image comparison

We also extract all images and screenshots from each of the student and compare the images with all the images for each student pair. We employ three different but related techniques to compare the images using image similarity index, structural similarity index and "d-hash" technique using 'opencv' package in Python[2]. This allows us to focus on those cases where images from two students are very close match of each other in terms of size and appearance. If students are indeed engaged in collaboration during the exam by sharing the same image/screenshot, the matching of images among students sharing the same Student Version is expected to be higher compared to other groups.

[2] https://pypi.org/project/dhash/ ; https://anaconda.org/conda-forge/opencv

3.5.3. Text comparison

We also compare the texts typed by all students to find any similarity among them. We extract all text characters from the documents submitted by students. [3] Since the students had to provide answers in the Word file provided by the instructor, we deleted all pre-typed texts (e.g "Answers to Section B") that were provided in the template. This leaves us with just the text typed in by students in the Word document. We also discard all non-alphanumeric characters (e.g *,%, /,(),£,$,etc) including whitespace (e.g tab spaces, empty lines, multiple spaces, etc). We consider any character or characters separated by space as one 'term'. We also retain text in their initial case (i.e. capital lower case) to perform exact match so that the same term with different capital letters (e.g. capital vs Capital) will not match together.

In an attempt to find any matching pattern, we first bring together all the text typed by students for the given cohort into one corpus. Within that corpus, we start with the highest number of consecutive terms that might appear in more than one student answer sheet. We start with an arbitrary number of N = 1200 to examine if any 1200-term combination appears more than once in the corpus. If such combination/ pattern appears more than once in the corpus, we identify the students where such a pattern occurs and then delete that specific pattern so that it is not counted again; then we move one number down to examine if any N-1 consecutive term appears more than once in the remainder of the corpus. If any given pattern of consecutive terms is not found more than once in the corpus, we simply try to find the next N-1 consecutive term and repeat the process until we reach down to N = 75. We do not document the matching for term combination that are lower than 75 in number due to the possibility that many general terms could match across students giving a false sense of collaboration. For instance, while answering a question, it is quite possible that more than one student are using the 9-term phrase "The interest rate applicable for short term loan is" without collaborating with each other. Hence to avoid this kind of matching which may not necessarily equate to collaboration, we do not match for term combination lower than 75. In this sense, we set a rather high bar for something to be detected as collaboration.

[3] For subject B, we extract text from Section B only for text analysis because Section A contains answers to MCQs.

4. Results and discussions

We present the initial findings and related discussion below.

4.1. Comparing MCQ score

Since MCQs appear only in Subject B exams, we focus on the cohorts B1 and B2 only to examine student response to MCQs.

B1 and B2 cohorts, with 122 and 179 students respectively, have 7389 and 9870 unique student pairs[4] respectively. For B1 cohorts, where all students sat for the same exam version, 34% of the student pairs had all 25 MCQ answers matching[5] with each other (see Table 1); the corresponding figure for B2 cohort – where students had to take the exam version based on their registration number – was 0.4%. It can also be deduced form Table 1 that 83% of the student pairs in B1 cohort had at least 20 MCQ responses matching with each other; whereas 25% of the student pairs from B2 had 20 or more MCQs responses matching. Untabulated results show that the average matched responses among student pairs for B1 (at 22.1) is higher than that for B2 (16.0). The interquartile range (i.e. upper quartile – lower quartile) for B1 and B2 stand at 4.0 and 6.0 respectively showing wider dispersion of matched responses among B2 compared to B1.[6] These findings provide support to our **hypothesis H1** that the answers within B1 cohort are more similar to each other compared to those of B2 cohort.

It is important to note that if two students answer all the MCQs correctly, their answer pattern would match for all 25 MCQ; but this does not necessarily mean that the students were collaborating. However, if the level of complexity of exams for B1 and B2 is the same and student capability and integrity for both the cohorts is expected to be the same, the matched pattern across the cohort would be expected to be the similar under normal conditions.

Now we take a closer look within B2 cohort as they had multiple versions of exam paper among them. As shown in Table 2, exam Version 0 was assigned to 16 students; Version 1 was assigned to 14 students, and so on. As such, there are 120 unique pairs sharing exam

---

[4] Given by $\frac{n!}{(n-k)!*k!}$) where n is the number of total students and k is the number in a pair

[5] The matched answers do not necessarily mean that the answers are correct. For instance, if two students both choose answers for their first three MCQ as A, C, and C, their pattern matches and they have a matching score of three regardless of the accuracy of their answers.

[6] The standard deviations of matched responses for B1 and B2 stand at 3.8 and 4.5 respectively

Version 0; 91 unique pairs sharing exam Version 1 and so on. Out of total 9870 unique pairs among cohort B2, there are 949 unique pairs where both the students in a given pair share the same Version of the exam. The Student Version shared by highest number of pair is Student Version 4 (shared by 153 pairs) and the ones shared by least number of pairs are Student Versions 7 and 8 (both versions shared by 55 pairs each).

The average, standard deviation, and inter quartile range of matched MCQ response for all student pairs within cohort B2 stand at 16.0, 4.5 and 6 respectively, as shown in Table 3. Some, but not all, intra-group averages are higher than the overall average. For instance, the matched averages within Student Versions 0, 1, 5, 7, and 8 are higher (shown in bold) than the cohort average indicating that within-group MCQ responses are more similar than the overall cohort in some cases. Similarly, only a few Student Version groups have a lower standard deviation and inter quantile range than the overall cohort score. If we take into consideration all three metrics (i.e., average, standard deviation, and inter quartile range) for matched MCQ responses, only two student groups, viz. V_0 and V_7 have more similar MCQ responses compared to the overall B2 cohort.

In relation to hypothesis H4_a, we also run a one-way ANOVA test to examine whether there is a significant difference in means of matched MCQ response between pairs that were assigned to the same Student Version and the ones assigned to different Student Version of exam. The results of the ANOVA test, as shown in Table 4, do not show a statistical difference in the matched counts between the two groups. This does not provide support to the notion that students in A2 cohort sharing the same exam version were able to collaborate more than other students who were not sharing the same version. Hence, we reject the 'cheating hypothesis' H4_a.

4.2. Comparing images

Students were allowed to insert images and screenshots in their digital answer papers for both Subject A and Subject B for both cohorts.

For Subject A, cohort A1 – who used the same version of exam for the cohort - used a total of 331 images at an average of 2.6 images per student. Out of the 331 images, 23 were digital screenshots (by 21 individuals) while the remaining 308 images were actual photos of hand-

written work on physical notebooks.[7] Computer-aided comparison of these images, using three different techniques discussed earlier, did not yield any perfect match among image pairs. For cohort A2, who took their exam one year later and were supposedly taking different 'versions' of the exam, a total of 1217 images were used at an average of 6.8 images per student. About half of the images were screenshots from their digital devices. Comparing the images against each other yields 15 pairs of exact matches of images used across 11 students (see Table 5); all of these images were for numerical calculations. When considering just the images, the answers within cohort A2 are more similar to each other compared to those within A1; and this is contrary to **hypothesis H1**. Perhaps this is not surprising given that most students within A1 cohort were using photographs of hand-written notes, which are more easily identifiable, thus making students less likely to share their hand-written photographs with other students. It is also possible that students from A1 cohort, who sat for their online exams at the start of the pandemic, were not used to taking screenshots, as evidenced by low number of screenshots. If students chose to share answers and hand-wrote their answers in their notebook before taking a picture and uploading it, exact matches of such images are unlikely.

To test hypothesis H2, we conduct a closer examination of cohort A2 to check whether students are more likely to collaborate if they are answering the same exam version. Out of the 11 students found to have shared images, we create maximum unique pairs resulting in 55 pairs. For each unique pair, we assign a *Cheat* dummy variable of 1 if the students within a pair have shared any of images (otherwise 0). Similarly, we assign a *Same* dummy variable of 1 if both students in a pair share the same exam version (otherwise 0). To examine the hypothesis if students are more likely to collaborate if they share the same exam version, we regress *Cheat* on *Same*. Results shown in Table 5 (b) show a positive and statistically significant coefficient for *Same* thus suggesting that students sharing the same exam version are more likely to collaborate with each other compared to other students who do not share the same exam paper. This provides support to **hypothesis H2**.

For Subject B, cohort B1 – who used the same version of the exam for the cohort - used a total of 426 images at an average of 3.5 images per student. Out of these images, there were

[7] A1 cohort sat for their exams in May 2020, when Covid-19 pandemic was suddenly taking hold and many academic institutions were moving towards online exams for the first time. Perhaps this also explains why so many students did not use screenshots and relied more on taking photos of their hand-written notes, as the students were probably not used to taking screenshots of answers from their digital devices.

two pairs of exact matches across four students. Both of these images relate to numerical calculations, as shown in Appendix 2(a). For cohort B2, who took their exam one year later and took different versions of the exam, a total of 553 images were used at an average of 3.9 images per student. All three image-matching techniques yielded just one image that matched exactly between two candidates. This image (see Appendix 2)(b) relates to a formula, but not the actual working, to a given numerical problem. Based on these results, the answers within Cohort B1 are more similar to each other compared to Cohort B2, thus supporting **hypothesis H1**.

In relation to hypothesis H3, we compare the reduction of collaboration in A2 and B2 cohorts compared to their previous cohort. Compared to A1 cohort, where no students were found to have shared images with each other, A2 cohort yielded 15 pairs of exact images shared across 11 students (see Table 5, a). Hence, image-sharing practice has increased for A2 cohort compared to the previous cohort. For Subject B, image-sharing in cohort B2 cohort (two images) was lower by 50% compared to B1 cohort (four images). Compared to prior cohort, reduction in image-sharing practice in B2 is more pronounced than in A2 cohort. This provides support to **hypothesis H3**.

Regarding **hypothesis H4**, we have already shown above that A2 cohort students are more likely to share images among themselves if they are sharing the same exam version (see Table 5, b). However, we only found one pair of image shared in B2 cohort and both the students were writing different versions of the exam (see Appendix 2 , b). Hence, the likelihood for cohort B2 students collaborating among students writing the same version of exam paper seems to be significantly lower than that for cohort A2. This provides support to **hypothesis H4**. On a similar note, since the only image shared within B2 cohort were between two students writing different versions of the exam, the image-sharing practice found within B2 cohort, however limited it is, does not provide support to **hypothesis H4_a**.

4.3. Comparison of text / textual analysis

For each of the cohorts, we first bring the texts typed by the students together in one corpus. The most commonly used word ‘term’ is “the” and numerical term is “0” across the cohorts. We retain all of these terms, despite many of these terms being regarded as ‘stopwords’ in linguistic analysis literature (Loughran & McDonald, 2016). A comparison of the terms used among different cohorts is shown in Appendix 3.

For Subject A, it is found that cohort A1 – who used the same version of exam for the cohort – had 29.1% of the cohort (i.e. 37 students) engaged in text-sharing practice, as shown in Table 7. The details of the text sharing practices for all cohorts are exhibited in Appendix 5. In the case of A2 cohort, such collaboration was found to be less widespread with only 18.4% (i.e. 33 individuals) of the cohort engaging in such text-sharing practice. This provides support to hypothesis H1 that higher proportion of A1 cohort engaged in unethical collaboration compared to A2 cohort.

For Subject B, cohort B1 – who used the same version of exam for the cohort – is found to have 11.5% of the cohort (i.e. 14 students) engaged in instances of text-sharing. For B2 cohort, such text-sharing practice was slightly less widespread at 8.5% of the students (i.e., 12 students). This provides support to hypothesis H1 that higher proportion of B1 cohort engaged in unethical collaboration compared to B2 cohort.

Results for both of these subjects provide support to **hypothesis H1** that the answers within Cohort A1 (and B1) will be more similar to each other compared to answers within Cohort A2 (and B2).

To test **hypothesis H2**, we take a closer look at the text-sharing practice of cohort A2 to see if students assigned to the same exam version are more likely to collaborate with each other. Since cohort A2 were told few days in advance as to which version of the exam paper each student was assigned to, we hypothesized that A2 students inclined to collaborate with other students would have enough time and incentive to find another student taking the same version of exam. For each of the 25 N-terms that were shared (as shown in Appendix 5, b), we create every possible student pair that has shared the N-term. This yields 44 student pairs.[8] Only in seven pairs out of these 44 pairs do both the students have the same exam version, while the exam versions are different for the other 37 pairs. This shows that a vast majority of the student pairs sharing the text were working on different versions of the exam paper. This finding does not support our **hypothesis H2**.

We also examine hypothesis H2 using logistic regression. For this purpose, we create maximum unique pairs out of the 33 students who were found to collaborate; this yields 528 unique student pairs. For each unique pair, we assign a *Cheat* dummy variable of 1 if the

[8] N1036 to N83 are shared by 1 pair of students each, yielding 22 student pairs; N82 and N78 are shared among 4 students each yielding 6 pairs each;N77 is shared by 5 students yielding 10 student pairs.

students within a pair have shared any of the N-term (otherwise 0). Similarly, we assign a *Same* dummy variable of 1 if both students in a pair share the same exam version (otherwise 0). To examine the hypothesis if students are more likely to collaborate if they share the same exam version, we regress *Cheat* on *Same*. However, our results, as shown in Table 6 (Column I) show that the variable Same though positive is not statistically significant. This does not provide support to the notion that students from cohort A2 sharing the same version are more likely to collaborate with each other. Thus, we do not find support for our **hypothesis H2**.

Regarding hypothesis H3, results provided in Table 7 show that the decrease in text sharing practice compared to prior cohort in B2 (at 25.8%) is not as pronounced as the reduction achieved by A2 cohort (36.7%). Hence, our results do not support **hypothesis H3**.

To test hypothesis H4, we take a closer look at the text-sharing practice of cohorts B2 (see Appendix 5, d). Out of the 12 students found to have exchanged texts, who seem to be working exclusively in pairs, only one pair has both students answering the same version of exam (candidate SN 3 and 6 both answering SV 3) while all the remaining student pairs are assigned to different version of the exam. To further examine this hypothesis empirically, we create maximum unique pairs out of the 12 students who were found to collaborate; this yields 66 unique student pairs. For each unique pair, we repeat the process of assigning a *Cheat* dummy variable (1 if the student pair have shared text between them; otherwise 0) and a *Same* dummy variable (1 if the students within the pair share the same exam version; otherwise 0) for B2 cohort. Conducting a logistic regression of *Cheat* on *Same* does not show that students from B2 cohorts are more likely to collaborate with each other if they are sharing the same exam version (Table 6, column II). When we re-run the regression combining A2 and B2 together (see column III), we still do not find evidence that students sharing the same exam version are more likely to collaborate. Finally, in the combined regression, we allow for module-specific intercept by introducing a *Module* dummy variable of 1 if cohort is B2 (otherwise 0 for cohort A2). We also introduce an interaction term between Module and Same to explicitly test hypothesis H4. Consistent with our hypothesis H4, we expect a negative and significant intercept for the interaction term meaning B2 cohort were less likely to collaborate among students writing the same version of exam paper compared to cohort A2. Contrary to our hypothesis, the interaction term, though negative, is

not statistically significant (Table 6, column IV). Thus, we do not find support for our **hypothesis H4**.

Regarding 'cheating hypothesis' **H4_a**, we do not find evidence that answer contents among students sharing the same exam version in cohort B2 are more similar to each other compared to inter-group contents (Table 6, column II).

For ease of reference, our results for all hypotheses across MCQs, images, and text are summarized in Appendix 6.

4.4. Further discussion

As shown in Appendix 6, we find overall support for our hypothesis H1. This supports the notion that introducing different versions of exams in an online setting can lower unethical collaboration among students. Though H1 is supported in our examination of MCQ and text, the finding is mixed for usage of images. For Subject A, image-sharing practices among students emerged with the introduction of different versions of the exam (in cohort A2). It has to be remembered that A1 cohort sat for their online exams in May 2020, when the pandemic was beginning to take hold. Hence, we posit that A1 cohort were not that familiar with the practice of sharing screenshots at the beginning of the pandemic and this led to no students from A1 cohort sharing exact screenshots with others. Not many students or teachers would have foreseen few months previously that exams would be held online; as such, A1 cohort simply were not familiar with the concept of sharing screenshots during exams. One year later, when cohort A2 sat for their exams, it was widely expected that the exams would be held online and during the course of the year, when teachings were mainly online, students were able to familiarize with the concept of using screenshots. This is further supported by the heavier use of screenshots by A2 cohort compared to A1 who mainly used pictures of hand-written texts. The result for Subject B however is consistent with the results from MCQ and text, thus supporting H1. Hence, our overall results show that using different versions of exams, whether the versions are actually different in substance or just different in name only, has the potential to lower unethical collaboration among students during online exams.

We find mixed results for H2, H3 and H4. These hypotheses are related to the conjecture that assigning students to different versions of exams by giving them advance notice increases the possibility that intent students will collaborate with other students who have been assigned to the same exam version. The results hold true for the use of images only while we do not find support to this notion when examining text. In this sense, pre-assigning students to different

versions of exam is fine as long as images or screenshots are not allowed to be used in answer sheets. This technique can be effective even if the 'different' exam versions are technically the same. On a related note, as regards to 'cheating hypothesis' H4_a, our results do not show that students who are assigned to different exam versions at very short notice have a higher likelihood of collaborating with others sharing the same exam version; this is consistent across MCQ, images, and text. That we have set a high bar as to what constitutes 'collaboration' is also important here: only exact matches of pictures (e.g. pixel by pixel) and exact text higher than 74 terms are taken as instances of unethical collaboration. For example, if the students were to alter even slightly the images or the text that they were copying from each other, such cases would not be picked up as unethical collaboration in our study. Hence, the level of collaboration we have documented in this study should be seen as the floor, rather than the ceiling, of such unethical practices. Based on the collective aim of minimizing unethical practices of all forms, the results we have provided suggest that element of surprise should be embedded (for example by assigning people to different versions of exam at short notice) in order to lower unethical collaboration.

An accidental revelation of our study, purely by observation, is that students tend to collaborate in pairs. Appendix 5 shows that almost all instances of text being shared were between just the same pairs. For instance, candidate SN1 and SN7 (in Appendix 5 d) shared six blocks of text between them. None of these six blocks of text were shared with anyone else. Similarly, candidate SN5 and SN10 share one block of text between them and no one else. It is interesting that all these abovementioned students were assigned to different versions of the exam. Across the four cohorts, a total of 73 N-terms [9] were shared. Out of these 73 blocks of text, 64 were shared between just within a pair as shown in the bottom rows of Appendix 5. Further, when we examine the images being shared, the images are generally shared between two individuals only. For example, Table 5 shows that out of 11 students sharing images in A2 cohort, 10 of them shared only with one individual while only candidate 89 shared image with two individuals (ID 23 and 65). Appendix 2 also shows that the only images shared within cohort B1 and B2 were in pairs. These observations suggest that unethical practices during online exams were mainly conducted in pairs rather than in groups.

[9] A1 has 25 N-terms shared; A2 also has 25; B1 has 13, and B2 has 11 N-terms shared among them.

5. Conclusion and limitations

In view of off-site assessments being adopted by Higher Education Institutions (HEI) especially in response to Covid-19, questions can be raised regarding the extent of possible unhealthy practices employed by students when undertaking their online exams. Such unhealthy practices could take various forms including working in groups during individual assessments, accessing resources beyond what is allowed, presenting someone else's work as their own, etc. In this context, we examine the impact of element of surprise in preserving exam integrity when common monitoring protocols or other paid outsourced services cannot be implemented. The surprise in this context relates to assigning students randomly to different versions of exam papers either at very short notice or ample notice period; such versions of exams could be really different or very similar to other versions. Using three different techniques namely image-processing, textual analysis, and analysis of multiple-choice question pattern across student pairs, we find that introduction of element of surprise lowers the degree of active collaboration among students during online exams and this approach is generally more effective in lowering text-sharing (or copy-paste) practices among students. Overall, we find text-matching pattern across collaborating students to decrease from 53% of their total typed text to 43% when element of surprise is introduced.

For multiple choice questions, our findings show limited benefit from element of surprise in online exams. There are instances where intra-group matching pattern is not statistically different to inter-group pattern. We also find that sharing of image is somewhat lowered when students are assigned to different exam versions. Finally, we find economically

significant decline in pattern-match across students when they are assigned to different versions of the exam.

However, there is limited evidence that students determined to collaborate in exams may be willing to do so with other students regardless of which exam versions they are assigned to. Further, student collaborations tend to take place in pairs rather than among many individuals, regardless of element of surprise.

Like any other research, our findings should be observed within the context provided. First, we have examined different cohorts from two different years. One benefit of this approach is that student conduct could be examined over exams of similar structure and difficulty level. It is possible that the cohorts were inherently different in character. Though the assumption in this study is that the overall character of students was similar across the cohorts over the two-year period, explicitly allowing for possible differences in student characters is something that could be explored further. Cultural traits of students, depending on the composition of international students for example, could be taken into account to explore this issue further. Second, our sample is limited in the sense that we have just four cohorts to compare results. Ideally, a similar approach could be employed across a much larger cohort.

Third, students could share responses to numerical questions with one another, hand-write such responses in their own notebook, and take pictures of such responses and upload online as their own work. Since the handwriting is different for each individual, these images would look different for each individual thus making it extremely difficult to detect such malpractices. Hence, the level of collaboration we have documented in this study should be seen as the proverbial tip of the iceberg.

Despite the shortcomings mentioned above, the approach taken in this study can be useful to other studies examining possible student malpractices in online exams; and the findings can be helpful to academic institutions in their effort to lower student malpractices in online exams.

References

Brammer, S., & Clark, T. (2020). COVID-19 and Management Education: Reflections on Challenges, Opportunities, and Potential Futures. *British Journal of Management*, *31*(3), 453–456. https://doi.org/https://doi.org/10.1111/1467-8551.12425

Crown, D. F., & Spiller, M. S. (1998). Learning from the Literature on Collegiate Cheating: A Review of Empirical Research. *Journal of Business Ethics*, *17*(6), 683–700. https://doi.org/10.1023/A:1017903001888

Lin, C.-H. S., & Wen, L.-Y. M. (2007). Academic dishonesty in higher education—a nationwide study in Taiwan. *Higher Education*, *54*(1), 85–97. https://doi.org/10.1007/s10734-006-9047-z

Löfström, E., Trotman, T., Furnari, M., & Shephard, K. (2015). Who teaches academic integrity and how do they teach it? *Higher Education*, *69*(3), 435–448. https://doi.org/10.1007/s10734-014-9784-3

Loughran, T., & McDonald, B. (2016). Textual Analysis in Accounting and Finance: A Survey. *Journal of Accounting Research*, *54*(4), 1187–1230. https://doi.org/https://doi.org/10.1111/1475-679X.12123

6. Tables and figures

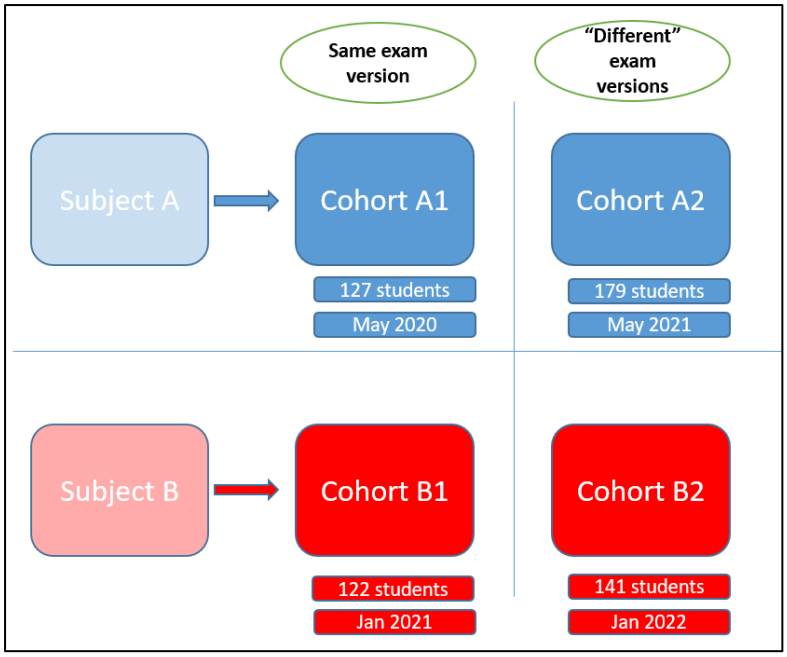


Figure 1: The four cohorts of students across two subjects. Two cohorts A2 and B2 were assigned to different versions of exam paper.

*Table 1**: Matched MCQ patterns among student pairs of Subject B***

This table shows the number of student pairs with exact number of matches on multiple choice questions.

B1 Cohort				B2 Cohort			
Matched	No. of pairs	% of total pairs	Cumulative %	Matched	No. of pairs	% of total pairs	Cumulative %
1				1	1	0.0%	0.0%
2				2			
3				3	11	0.1%	0.1%
4				4	38	0.4%	0.5%
5				5	80	0.8%	1.3%
6	5	0.1%	0.1%	6	165	1.7%	3.0%
7	7	0.1%	0.2%	7	158	1.6%	4.6%
8	25	0.3%	0.5%	8	193	2.0%	6.5%
9	94	1.3%	1.8%	9	337	3.4%	10.0%
10	104	1.4%	3.2%	10	418	4.2%	14.2%
11	21	0.3%	3.5%	11	463	4.7%	18.9%
12	26	0.4%	3.8%	12	484	4.9%	23.8%
13	56	0.8%	4.6%	13	455	4.6%	28.4%
14	252	3.4%	8.0%	14	546	5.5%	33.9%
15	48	0.7%	8.6%	15	555	5.6%	39.6%
16	73	1.0%	9.6%	16	660	6.7%	46.2%
17	205	2.8%	12.4%	17	862	8.7%	55.0%
18	215	2.9%	15.3%	18	986	10.0%	65.0%
19	154	2.1%	17.4%	19	1032	10.5%	75.4%
20	280	3.8%	21.2%	20	887	9.0%	84.4%
21	421	5.7%	26.9%	21	681	6.9%	91.3%
22	902	12.2%	39.1%	22	506	5.1%	96.4%
23	900	12.2%	51.3%	23	222	2.2%	98.7%
24	1091	14.8%	66.1%	24	88	0.9%	99.6%
25	2502	33.9%	100.0%	25	42	0.4%	100.0%
	7381	**100%**			**9870**	**100%**	

*Table 2**: Exam versions and unique student pairs for Cohort B2***

This table shows the number of students and student pairs by Student version of exam

I	II	III
Student Version	Number of students	Unique pairs
0	16	120
1	14	91
2	12	66
3	16	120
4	18	153
5	13	78
6	16	120
7	11	55
8	11	55
9	14	91
Total	***141***	***949***

*Table 3**: Average intra-group matches for MCQ for B2 cohort***

This table presents the average intra-group matched MCQ pattern among different Student Version of exam papers along with the number of students within each group. V_0, V_1 etc represent Student Version 0, 1 etc of the exam papers. V_none represents all student pairs who do not share the same Student Version of the exam paper.

Matched Student Versions	Average Matched Answers	Std. Dev	Min	Max	Inter Quartile Range	Count
V_0	**18.5**	**3.7**	9	25	**4**	120
V_1	**16.5**	**4.0**	8	25	6	91
V_2	13.6	5.2	4	23	7	66
V_3	15.5	**4.1**	5	25	**5**	120
V_4	14.2	**4.4**	6	23	7	153
V_5	**16.7**	**4.4**	7	25	6	78
V_6	15.5	5.1	4	24	6	120
V_7	**19.0**	**2.3**	15	23	**4**	55
V_8	15.4	**3.8**	6	21	6.5	55
V_9	**16.2**	4.5	8	25	7	91
SV_none	16.0	4.5	1	25	6	8921
Overall	**16.0**	**4.5**	**1**	**25**	**6**	**9870**

*Table 4**: Analysis of variance***

This table shows the ANOVA test for matched scores between student pairs that were assigned to the same Student Version and the ones assigned to different Student Version of exam

Source	SS	df	MS	F	Prob > F
Between groups	2.3	1	2.3	0.11	0.737
Within groups	200026.7	9868	20.3		
Total	200028.9	9869	20.3		

Bartlett's test for equal variances: chi2(1) = 0.118 Prob>chi2 = 0.731

Table 5: Images for cohort A2

a)This table shows the details of images that were exact match of each other for cohort A2. *Student_1 ID* represents anonymised roll number of first student in the pair; *Student_2 ID* represents anonymised roll number of second student in the pair; *Student_1 Version* represents the Student Version of exam paper for the first student in the pair; *Student_2 Version* represents the Student Version of exam paper for the second student in the pair; *Student_1 Image* represents the matched image in the pair of the first student; *Student_2 Image* represents the matched image in the pair for the second student in the pair.

Student_1 ID	Student_2 ID	Student_1 Version	Student_2 Version	Student_1 Image	Student_2 Image
23	89	1	3	image14.png	image14.png
23	89	1	3	image6.png	image6.png
27	91	4	3	image2.png	image2.png
27	91	4	3	image1.png	image3.png
27	91	4	3	image1.png	image1.png
65	89	3	3	image2.png	image2.png
65	89	3	3	image3.png	image3.png
65	89	3	3	image1.png	image1.png
131	66	5	5	image3.png	image3.png
131	66	5	5	image2.png	image2.png
131	66	5	5	image6.png	image6.png
131	66	5	5	image4.png	image4.png
137	154	4	4	image8.png	image7.png
137	154	4	4	image7.png	image6.png
159	63	2	9	image10.png	image10.png

b) This table shows the results from maximum likelihood estimation. Dependent variable is Ln(P) - Ln(1-P) where P is the probability of a student collaborating with another student in the exam. *Same* is a dummy variable of 1 if both students in a student pair share the same exam version. Statistical significance is reported against 10% (*), 5% (**) and 1% (***) significance levels with *t*-statistics shown in brackets, after clustering the error terms using Huber-White approach

	A2 Images
Same	2.420**
	(2.48)
Constant	-2.708***
	(-4.50)
No. of observations	55
Pseudo R-sq.	0.156
Log-likelihood	-16.00

Table 6: Logistic regression

This table shows the results from maximum likelihood estimation. Dependent variable is Ln(P) - Ln(1-P) where P is the probability of a student collaborating with another student in the exam. *Same* is a dummy variable of 1 if both students in a student pair share the same exam version. Module is a dummy variable of 1 for B2 cohort and 0 for A2 cohort. Statistical significance is reported against 10% (*), 5% (**) and 1% (***) significance levels with *t*-statistics shown in brackets, after clustering the error terms using Huber-White approach.

	I A2	II B2	III Both A2 and B2	IV Both A2 and B2
Same	0.390 (0.77)	0.00 (0.00)	0.343 (0.74)	0.390 (0.77)
Module				0.390 (0.77)
Module*Same				-0.390 (-0.31)
Constant	-2.692*** (-14.26)	-2.303*** (-4.87)	-2.645*** (-15.11)	-2.692*** (-14.26)
No. of observations	528	66	594	594
Pseudo R-sq.	0.00213	0	0.00172	0.00356
Log-likelihood	-128.5	-20.11	-148.9	-148.6

*Table 7**: Percent of students sharing same text, by cohort***

Cohort	No. of Students	Collaborating students	% student collaborating	Increase/ Decrease compared to previous cohort
A1	127	37	29.1%	
A2	179	33	18.4%	-36.7%
B1	122	14	11.5%	
B2	141	12	8.5%	-25.8%

7. Appendices

Appendix 1: Showing different versions of the Question 2 between Main version 0 and Main version 3

Question 2 from Main version 0

QUESTION 2

2a)

Coleraine Industries has $75 million in free cash that it can use for a share buy-back. Suppose instead Coleraine invests the funds in an account paying 9% interest per year. The corporate tax rate is 17%; for investors, the tax rates on capital gains and interest income are 12% and 15% respectively.

Requirement

Question 2 from Main version 3

QUESTION 2

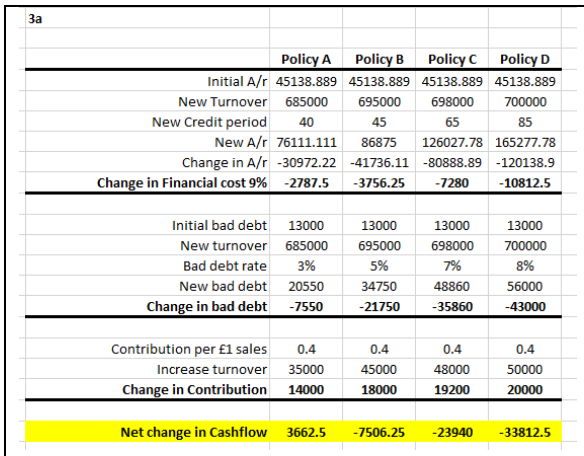
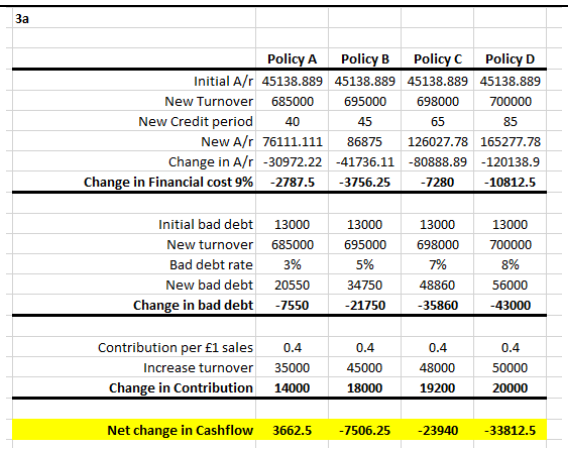
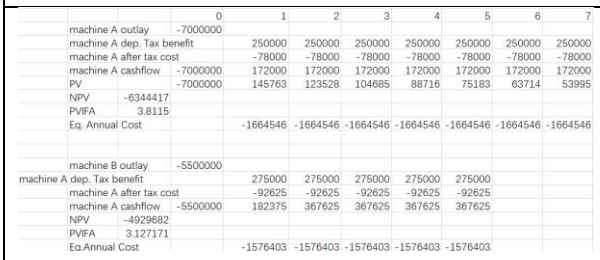
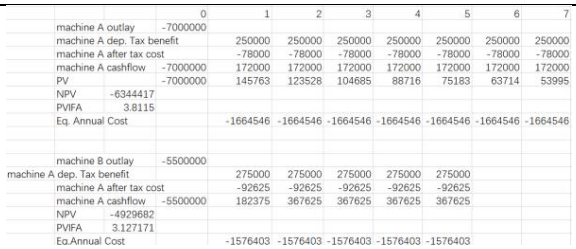
2a)

Coleraine Industries has $75 million in free cash that it can use for a share buy-back. Suppose instead Coleraine invests the funds in an account paying 8% interest per year. The corporate tax rate is 17%; for investors, the tax rates on capital gains and interest income are 12% and 15% respectively.

Requirement

Appendix 2: Image comparison

a) Matched images of B1 cohort

 3a Policy A / Policy B / Policy C / Policy D Initial A/r 45138.889 45138.889 45138.889 45138.889 New Turnover 685000 695000 698000 700000 New Credit period 40 45 65 85 New A/r 76111.111 86875 126027.78 165277.78 Change in A/r -30972.22 -41736.11 -80888.89 -120138.9 **Change in Financial cost 9%** **-2787.5** **-3756.25** **-7280** **-10812.5** Initial bad debt 13000 13000 13000 13000 New turnover 685000 695000 698000 700000 Bad debt rate 3% 5% 7% 8% New bad debt 20550 34750 48860 56000 **Change in bad debt** **-7550** **-21750** **-35860** **-43000** Contribution per £1 sales 0.4 0.4 0.4 0.4 Increase turnover 35000 45000 48000 50000 **Change in Contribution** **14000** **18000** **19200** **20000** **Net change in Cashflow** **3662.5** **-7506.25** **-23940** **-33812.5**	 3a Policy A / Policy B / Policy C / Policy D Initial A/r 45138.889 45138.889 45138.889 45138.889 New Turnover 685000 695000 698000 700000 New Credit period 40 45 65 85 New A/r 76111.111 86875 126027.78 165277.78 Change in A/r -30972.22 -41736.11 -80888.89 -120138.9 **Change in Financial cost 9%** **-2787.5** **-3756.25** **-7280** **-10812.5** Initial bad debt 13000 13000 13000 13000 New turnover 685000 695000 698000 700000 Bad debt rate 3% 5% 7% 8% New bad debt 20550 34750 48860 56000 **Change in bad debt** **-7550** **-21750** **-35860** **-43000** Contribution per £1 sales 0.4 0.4 0.4 0.4 Increase turnover 35000 45000 48000 50000 **Change in Contribution** **14000** **18000** **19200** **20000** **Net change in Cashflow** **3662.5** **-7506.25** **-23940** **-33812.5**
Candidate 59, image 4	Candidate 98, image 5
 0 1 2 3 4 5 6 7 machine A outlay -7000000 machine A dep. Tax benefit 250000 250000 250000 250000 250000 250000 250000 machine A after tax cost -78000 -78000 -78000 -78000 -78000 -78000 -78000 machine A cashflow -7000000 172000 172000 172000 172000 172000 172000 172000 PV -7000000 145763 123528 104685 88716 75183 63714 53995 NPV -6344417 PVIFA 3.8115 Eq. Annual Cost -1664546 -1664546 -1664546 -1664546 -1664546 -1664546 -1664546 machine B outlay -5500000 machine A dep. Tax benefit 275000 275000 275000 275000 275000 machine A after tax cost -92625 -92625 -92625 -92625 -92625 machine A cashflow -5500000 182375 367625 367625 367625 367625 NPV -4929682 PVIFA 3.127171 Eq.Annual Cost -1576403 -1576403 -1576403 -1576403 -1576403	 0 1 2 3 4 5 6 7 machine A outlay -7000000 machine A dep. Tax benefit 250000 250000 250000 250000 250000 250000 250000 machine A after tax cost -78000 -78000 -78000 -78000 -78000 -78000 -78000 machine A cashflow -7000000 172000 172000 172000 172000 172000 172000 172000 PV -7000000 145763 123528 104685 88716 75183 63714 53995 NPV -6344417 PVIFA 3.8115 Eq. Annual Cost -1664546 -1664546 -1664546 -1664546 -1664546 -1664546 -1664546 machine B outlay -5500000 machine A dep. Tax benefit 275000 275000 275000 275000 275000 machine A after tax cost -92625 -92625 -92625 -92625 -92625 machine A cashflow -5500000 182375 367625 367625 367625 367625 NPV -4929682 PVIFA 3.127171 Eq.Annual Cost -1576403 -1576403 -1576403 -1576403 -1576403
Candidate 48, image 5	Candidate 88, image 4

b) Matched image of B2 cohort

$3 \times \left(\frac{3*F*\sigma^2}{K*4}\right)^{1/3}$	$3 \times \left(\frac{3*F*\sigma^2}{K*4}\right)^{1/3}$
Candidate 70, 1mage 4 (SV 5)	Candidate 108, image 1 (SV 4)

Appendix 3: Terms used by students , by cohort

Cohort	No. of Students	Unique terms used	Most common number	Most common word	Terms used only once (no.)	Average terms
A1	127	6057	0 (1952)	the (8871)	2331	1123
A2	179	4755	0 (1799)	the (3826)	1773	450
B1	122	3306	0 (1069)	the (2030)	1295	356
B2	141	2102	1 (1317)	of (758)	755	230

Appendix 4: Matching terms for current year when N = 496

I 75 8 1 17 5 17 million 5 17 1 0 12 4 5496 million II 75 8 1 0 15 5 1 million 1 spent on fees 1 1 0 17 1 0 12 0 7304 fees 5 1 0 7304 0 7304 5 98 million This means if the fee of raising new funds is 5 98million the actual cost to shareholders is0 6204 million i Number of time periods 24 12years 2 Coupon payments 40 8 1000 2 Semiannual interest rare 4 8 2 P 40 PVIFA 4 24 40 16 9355 644 42 1000 332 58 1 r 24 1000 332 58 1 r 24 r 0 047 YMT 0 047 2 0 094 ii Number of time periods 8 4years 2 Coupon payments 40 8 1000 2 Semiannual interest rare 4 8 2 PV 40 PVIFA 4 8 35 6 7327 235 6445 1000 774 3555 1 r 8 r 0 032 YTC 0 032 2 0 064 Effective tax advantage of debt in this case is shown as follows 1 1 0 15 1 0 19 1 0 28 0 04375 4 375 The leveraged recap will increase Belfast s enterprise value by T D 4 375 220 million 9 625 million There are 60 million share after the announcement share price should increase by 9 625 60 0 16 Calculation of Cost of equity Using the dividend growth model P D0 1 g r g P 95 D0 7 g 0 03 r r 10 59 Calculation of Cost of pref shares rp 7 5 50p 60p 3 75 60 6 25 Calculation of Cost of debt pre tax 8 100 95 rd 8 42 Market values Weight Equity E 855 000 855 2460 E E P D Pref shares P 180 000 180 2460 P E P D Debt D 1425 000 1425 2460 D E P D 2460 000 WACC 855 2460 r_E 180 2460 r_P 1425 2460 r_D 1 t E P D WACC 3 68 0 46 4 8 14 1 2 3 4 5 6 7 Machine Green outlay 7000000 machine Green dep Tax benefit 7000000 7 22 220000 220000 220000 220000 220000 220000 220000 Machine Green after tax cost 60000 1 22 46800 46800 46800 46800 46800 46800 46800 Machine Green Flows 7000000 173200 173200 173200 173200 173200 173200 173200 Discounted CF 7000000 154643 138074 123280 110072 98278 87749 78347 NPV of Greeen 6209557 FVIFA 13 5 7 4 3546 Eq Annual Cost NPV of Green PVIFA 1425977 1217821 1217821 1217821 1217821 1217821 1217821 Machine Value outlay 6000000 machine Value dep Tax benefit 6000000 5 22 264000 264000 264000 264000 264000 Machine Value after tax cost70000 1 22 54600 54600 54600 54600 54600 Machine Value Flows 6000000 209400 209400 209400 209400 209400 Discounted CF 6000000 1551111 11489712 85108977 630436869 4669902732 NPV of Value 5392489401 FVIFA 13 5 5 3 4747 Eq Annual Cost NPV of Value PVIFA 1551929491 1551929491 1551929491 1551929491 1551929491

Appendix 5: This appendix presents the number of N-terms appearing more than once in the texts for a given year. *SN* is student number; *SV* is Student Version of exam paper; *Terms* represent the total number of terms used by each student. *N* followed by digits refers to the number of consecutive terms; for example, *N362* represent 362 consecutive matching terms, and so on. *FINAL* is the total number of matching terms for the reported student. *Similarity* represents the degree of similarity text for the specific student and is calculated as *FINAL* scaled by *Terms*. **1** denotes that the corresponding N-term for the student matches exactly with that of another student. *Shared among* shows the number of students sharing the corresponding N-term.

a) Matched patterns for A1 cohort

SN	SV	Terms	N362	N235	N215	N137	N127	N118	N115a	N115b	N114	N112	N100	N98	N94	N91	N87	N86	N85	N84	N83	N82a	N82b	N79a	N79b	N78	N77	FINAL	Similarity
1	0	901	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	91	0.10
2	0	1132	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	296	0.26
3	0	901	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	79	0.09
4	0	668	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	215	0.32
5	0	2171	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	854	0.39
6	0	1518	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	115	0.08
7	0	869	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	87	0.10
8	0	773	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	115	0.15
9	0	881	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	577	0.65
10	0	1477	0	1	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	1	0	576	0.39
11	0	1394	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	82	0.06
12	0	1254	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0.08
13	0	813	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	86	0.11
14	0	2332	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	115	0.05
15	0	1795	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	78	0.04
16	0	765	0	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	318	0.42
17	0	1586	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	114	0.07
18	0	1674	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	185	0.11
19	0	1334	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	294	0.22
20	0	1135	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	292	0.26
21	0	1358	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	235	0.17
22	0	726	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	85	0.12
23	0	722	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	181	0.25
24	0	1210	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	114	0.09
25	0	1069	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	233	0.22
26	0	1093	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	215	0.20
27	0	2359	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	127	0.05
28	0	955	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	91	0.10
29	0	852	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	84	0.10
30	0	929	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	194	0.21
31	0	1023	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	77	0.08
32	0	1443	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	85	0.06
33	0	815	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	77	0.09
34	0	989	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	79	0.08
35	0	1711	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	84	0.05
36	0	1266	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	233	0.18
37	0	1145	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	137	0.12
Shared among			**2**	**2**	**4**	**2**	**2**	**2**	**2**	**5**	**2**	**2**	**2**	**2**	**3**	**2**	**5**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**		

b) Matched patterns for A2 cohort

SN	SV	Terms	N1036	N237	N236	N227	N226	N192	N180	N177	N176	N168	N161	N152	N147	N128	N113	N112	N105	N98	N96	N91	N88	N83	N82	N78	N77	FINAL	Similarity
1	1	741	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	320	0.43
2	9	444	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	180	0.41
3	5	822	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	77	0.09
4	3	1064	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	417	0.39
5	4	1187	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	176	0.15
6	9	556	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	82	0.15
7	1	344	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	83	0.24
8	4	1399	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1262	0.9
9	3	681	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	82	0.12
10	2	829	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	96	0.12
11	3	827	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	320	0.39
12	0	428	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	78	0.18
13	7	944	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	289	0.31
14	5	1590	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	574	0.36
15	2	1492	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	77	0.05
16	7	1329	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	268	0.2
17	3	275	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	98	0.36
18	5	680	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	383	0.56
19	2	1451	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	113	0.08
20	3	757	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	237	0.31
21	3	1313	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1262	0.96
22	2	1678	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	574	0.34
23	6	1556	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	181	0.12
24	2	871	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	268	0.31
25	7	879	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	264	0.3
26	5	705	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	383	0.54
27	7	840	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	1	0	1	560	0.67
28	1	940	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	193	0.21
29	9	153	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	78	0.51
30	9	520	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	105	0.2
31	1	1252	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	113	0.09
32	1	1254	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	194	0.15
33	0	204	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	96	0.47
Shared among			**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**4**	**4**	**5**		

c) Matched patterns for B1 cohort

SN	SV	Terms	N543	N277	N171	N138	N125	N118	N116	N112	N109	N104	N94	N83	N81	FINAL	Similarity
1	0	187	0	0	0	0	0	0	1	0	0	0	0	0	0	116	0.62
2	0	287	0	0	0	0	1	0	0	0	0	0	0	0	0	125	0.44
3	0	705	1	0	0	0	0	0	0	0	0	0	0	0	0	543	0.77
4	0	1403	0	1	0	1	0	0	1	0	0	1	0	1	0	718	0.51
5	0	1007	0	0	0	1	0	0	0	0	0	0	1	1	0	315	0.31
6	0	543	1	0	0	0	0	0	0	0	0	0	0	0	0	543	1.00
7	0	1326	0	1	0	0	0	1	1	0	0	0	1	0	0	605	0.46
8	0	747	0	0	0	0	0	0	0	0	0	1	0	1	0	187	0.25
9	0	491	0	0	0	0	1	1	1	0	0	0	0	0	0	359	0.73
10	0	410	0	0	0	0	0	0	0	0	0	0	0	0	1	81	0.20
11	0	1984	0	0	0	0	0	0	0	0	0	0	0	0	1	81	0.04
12	0	852	0	0	1	0	0	0	0	1	1	0	0	0	0	392	0.46
13	0	692	0	0	1	0	0	0	0	1	1	0	0	0	0	392	0.57
14	0	543	1	0	0	0	0	0	0	0	0	0	0	0	0	543	1.00
Shared among			3	2	2	2	2	2	4	2	2	2	2	3	2		

d) Matched patterns for B2 cohort

SN	SV	Terms	N496	N125	N121	N119	N115	N104	N102	N97	N96	N94	N91	FINAL	Similarity
1	4	1090	0	0	**1**	0	**1**	0	0	**1**	**1**	**1**	**1**	614	0.56
2	3	496	**1**	0	0	0	0	0	0	0	0	0	0	496	1.00
3	3	757	0	0	0	0	0	0	**1**	0	0	0	0	102	0.13
4	0	590	0	0	0	**1**	0	0	0	0	0	0	0	119	0.20
5	9	763	0	**1**	0	0	0	0	0	0	0	0	0	125	0.16
6	3	710	0	0	0	0	0	0	**1**	0	0	0	0	102	0.14
7	3	675	0	0	**1**	0	**1**	0	0	**1**	**1**	**1**	**1**	614	0.91
8	3	455	0	0	0	0	0	**1**	0	0	0	0	0	104	0.23
9	7	520	0	0	0	**1**	0	0	0	0	0	0	0	119	0.23
10	5	893	0	**1**	0	0	0	0	0	0	0	0	0	125	0.14
11	9	265	0	0	0	0	0	**1**	0	0	0	0	0	104	0.39
12	1	496	**1**	0	0	0	0	0	0	0	0	0	0	496	1.00
Shared among			**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**		

Appendix 6: Summary of hypothesis and test results; Y= support for hypothesis; N = no support for hypothesis;

n/a = not applicable

Hypothesis	Related modules	MCQ (Cohorts: B1,B2)	Image	Text
H1	A1,A2;B1,B2	Y	N/Y	Y
H2	A2	n/a	Y	N
H3	A1,A2,B1,B2	n/a	Y	N
H4	A2, B2	n/a	Y	N
H4_a	B2	N	N	N

H1: Higher proportion of Cohort A1 (and B1) will engage in collaboration compared to Cohort A2 (and B2).
H2: Students within Cohort A2 will collaborate more with other students who share the same Student Version of the exam paper.
H3: Compared to prior cohort, reduction in collaboration in B2 will be more pronounced than in A2 cohort.
H4: The likelihood for cohort B2 students collaborating among students writing the same version of exam paper is significantly lower than that for cohort A2.
H4_a: Cheating hypothesis - For Cohort B2, answer contents within given groups are more similar to each other compared to inter-group contents.