

Effects of extracurricular videos on the grades of
students and their engagement

V. N. Engbers
1236660

July 11, 2022

Contents

1	Introduction	8
1.1	Set up of the experiment	8
1.2	Do the videos improve the grades of the students?	8
1.3	Do the videos engage the students with the course?	8
1.4	Do the videos improve learning retention?	9
2	Structure of the data	10
2.1	Structure of the course	10
2.2	Course data	10
2.3	Grades of the exam questions	10
2.4	Groups	10
2.5	Dependencies between years	10
2.6	Detailed grades of the midterms	11
2.7	Survey	11
2.8	Interpreting figures	11
2.9	Significance level	11
3	Statistical tests	11
3.1	Mann-Whitney U test	11
3.2	Pearson's chi-squared test	12
3.3	Kendall tau correlation coefficient	12
3.4	Chronbach's alpha	12
3.5	Picture of the decision tree	14
4	Decision trees	15
4.1	Were there differences in the grades between years 2020 and 2021?	15
4.1.1	Reliability of the midterm and exam	15
4.1.2	Are there differences between years based based on the passing rate?	15
4.1.3	Similarity of the students	16
4.2	Did the students that watched the videos score better on the renewal midterm and the renewal part of the exam?	16
4.2.1	Did the students that watched the videos perform better on the branching and Brownian midterms?	16
4.3	Did the videos improve learning retention?	16
4.3.1	Consistency of the midterms and the exam.	16
4.4	Did the videos affect motivation?	17
4.4.1	Did students spend more time on renewal theory because of the videos?	17
4.4.2	Did the students that watched the videos also spend more time on the other two subjects?	17

5	Is the passing rate of students different between years?	18
5.1	Were there more students that passed the course with an 8 or higher in one year?	18
6	Comparing the renewal midterm between years	20
6.1	Descriptive statistics	20
6.2	Statistical test	20
6.3	Conclusion and discussion	20
7	Comparing the branching and Brownian midterms between years	21
7.1	Branching Midterm	21
7.1.1	Descriptive statistics	21
7.1.2	Statistical test	22
7.2	Brownian midterm	22
7.2.1	Descriptive statistics	22
7.2.2	Statistical test	22
7.3	Conclusion on the branching and Brownian midterm:	22
8	Reliability of the midterms	23
8.1	Procedure	23
8.2	Renewal midterm 2020	23
8.3	Renewal midterm 2021	23
8.4	Branching and Brownian midterm 2020	24
8.5	Branching midterm 2021	25
8.6	Brownian midterm 2021	25
9	Re-analysing the renewal midterm	25
9.0.1	Descriptive statistics	26
9.0.2	Statistical test	26
9.1	Conclusion	26
10	Comparing the exam between years	26
10.1	Renewal theory	26
10.1.1	Descriptive statistics	26
10.1.2	Statistical test	27
10.2	Branching theory	27
10.2.1	Descriptive statistics	27
10.2.2	Statistical test	28
10.3	Brownian theory	28
10.3.1	Descriptive statistics	28
10.3.2	statistical tests	28
10.4	Probability theory	29
10.4.1	Descriptive statistics	29
10.4.2	Statistical test	29
10.5	Final Exam grades	29
10.5.1	Descriptive statistics	29

10.5.2	Statistical test	30
10.6	Overview tests	30
10.7	Conclusion	31
11	Reliability of the exam	31
11.1	Is the exam of 2020 reliable?	31
11.2	Is the exam of 2021 reliable?	32
11.3	Method for re-analysing the exams	33
12	Exam analysis with non-discriminatory questions removed	34
12.1	Renewal theory	34
12.1.1	Descriptive statistics	34
12.1.2	Statistical tests	35
12.2	Branching and Brownian theory	35
12.3	Probability theory	35
12.3.1	Descriptive statistics	35
12.3.2	Statistical test	35
12.4	Final Exam grade	36
12.4.1	Descriptive statistics	36
12.4.2	Statistical test	36
12.5	Overview tests when we remove the non-discriminatory questions	36
12.6	Conclusion of removing the non-discriminatory questions	37
13	Exam analysis with questions that had a low or high average removed.	37
13.1	Renewal theory	38
13.1.1	Descriptive statistics	38
13.1.2	Statistical test	39
13.2	Branching theory	39
13.2.1	Descriptive statistics	39
13.2.2	Statistical test	39
13.3	Brownian theory	40
13.3.1	Descriptive statistics	40
13.3.2	Statistical test	40
13.4	Probability theory	40
13.5	Final Exam grade	40
13.5.1	Descriptive statistics	40
13.5.2	Statistical test	41
13.6	Conclusion of the exam comparison while removing questions with a very high or low average	42
13.7	Conclusion about differences on the exam	42
14	Conclusion: Did the videos improve the grades in 2021?	42
14.1	Conclusion about the difference in students between years	42

15 Analysing the differences between the students that watched the videos and the students that did not.	44
15.1 Students that watched the videos	44
15.2 Students that did not watch the videos	44
15.3 Students that did not fill in the survey	44
15.4 Did the content of the videos match the content of the lecture? .	44
15.5 Groups for the midterms	44
16 Did the videos improve the renewal midterm grades of the students?	45
16.1 Summary statistics	45
16.2 Statistical test	46
16.3 Conclusion	46
17 Do the students that watched the videos perform differently on the other two midterms?	47
17.1 Branching Midterm	47
17.1.1 Descriptive statistics	47
17.1.2 Statistical test	48
17.2 Brownian Midterm	48
17.2.1 Descriptive statistics	48
17.2.2 Statistical test	48
17.3 Conclusion	49
18 Do the students that watched the videos differ on the exam?	49
18.1 Renewal theory	50
18.1.1 Descriptive statistics	50
18.1.2 Statistical test	50
18.2 Branching theory	50
18.2.1 Descriptive statistics	50
18.2.2 Statistical test	50
18.3 Brownian theory	51
18.3.1 Descriptive statistics	51
18.3.2 Statistical test	51
18.4 Probability theory	52
18.4.1 Descriptive statistics	52
18.4.2 Statistical test	52
18.5 Final exam grade	52
18.5.1 Descriptive statistics	52
18.5.2 Statistical test	52
18.6 Recap	52
18.7 Conclusion	53
19 Recap of the decision tree	54

20 Did the videos increase the motivation of the students?	55
20.1 Did the students report more motivation due to the videos? . . .	55
20.2 Did the videos help students with trying to work out new ideas?	55
20.3 Did the students that watched the videos spend more time on the course?	56
20.3.1 Renewal theory	56
20.3.2 Branching theory	57
20.3.3 Brownian theory	57
20.4 Conclusion	58
21 What was the most interesting video?	58
22 Did the videos enrich the course?	58
23 Recap of the decision tree	60
24 Consistency of the midterms with the exam	61
24.1 Test procedure	61
24.2 Is the renewal midterm consistent with the exam in 2020	61
24.3 Summary of the consistency of the other midterms with the exam in 2020	62
24.4 Is the midterm consistent with the corresponding subject on the exam?	62
24.5 Conclusion about consistency	62
25 Learning retention	63
25.1 Set up of the experiment	63
25.2 Test procedure	64
25.3 Renewal midterm	64
25.3.1 Year 2020	64
25.3.2 Year 2021	64
25.3.3 Conclusion on the renewal midterm	65
25.4 Branching processes midterms	65
25.5 Brownian midterm	65
25.6 Conclusion	65
25.7 Did the group that watched the video retain their midterm rank on the exam?	66
25.8 Renewal subject	66
25.8.1 Students that did not watch the videos	66
25.8.2 Students that watched the videos	66
25.9 Branching subject	66
25.10 Brownian subject	67
25.11 Conclusion about learning retention	67
26 Recap of the decision tree	68
27 Conclusion	69

28 Discussion	69
28.1 Do the videos improve the grades of the students?	69
28.2 Do the videos affect student engagement with the course?	69
28.3 Do the videos affect learning retention	70
29 Further recommendations	70
A Scores per quintile for the discriminatory questions on the exams	72
A.1 Histograms for the exams with questions that had a low or high average dropped	73

1 Introduction

Teaching methods have remained largely the same over time, even though there are new possibilities with technology. We are interested in whether videos can provide a good addition to a course, especially since videos on YouTube, for example TED Talks or self help videos, do quite well online.

1.1 Set up of the experiment

We have data from the course "Probability and Stochastics 1", a master course at the Eindhoven University of Technology with on average 70 students. The course has three subjects: renewal processes, branching processes and Brownian motion. For these subjects the students make a take-home midterm in groups of two students. The exam is made alone and consists of these three subjects plus a part about probability theory.

In 2021 six extracurricular videos were made for renewal theory and students were not obliged to watch these videos. The subjects of the videos were chosen by an old student of the course. We have access to grades from 2020, when there were no videos yet and to the grades in 2021. With this data, we can compare differences between the two years to see if the videos had any effect. Note that the comparison between the years is also influenced by the difficulty of the midterm or exam of that year and how well the students performed for example. Students in 2021 also had to fill in a survey on whether they watched some of the videos, this creates another comparison within the year as we can also compare the students that watched the videos with those that did not.

1.2 Do the videos improve the grades of the students?

The goal of a course is to teach students new skills and the most important measurement associated with this is the grades a student gets for the course. Our hypothesis is that videos help with studying, because you view the material in a different way than a lecture or instruction. Additionally you can watch videos at your own pace and pause them or go back, which is not possible during lectures.

1.3 Do the videos engage the students with the course?

We are also interested in whether the engagement of students increased because of the videos. This might also be reflected in the grades that students obtain, as students that are better engaged with the course might get higher grades. Research suggests that shorter videos of around 6 minutes with a presenter work well for this [1].

1.4 Do the videos improve learning retention?

We are also interested in whether videos can help with retaining the new knowledge students gained during the course. We have measurements for the midterms and the final exam. Students could study between the midterm and the exam, so the grade on the exam is potentially influenced by the videos, the knowledge obtained during the midterm, and studying for the exam. With the data we have available we were unable to draw a conclusion for this question. The results are given in section 4.3, to show what we accomplished during this project.

2 Structure of the data

2.1 Structure of the course

The course has three take home midterms that are made in groups. This type of assignment allows for more difficult questions on the midterm as students can use the internet and work together in their group. Students could also choose to work alone, instead of working together in a group. Furthermore students could also switch groups after a midterm, this option was not used however. On the exam students worked alone and they could only use scrap paper. If a student did not pass the exam a resit was possible 3 months later. The weight of each midterm in the course is 10% and the weight of the exam is 70%.

2.2 Course data

In the year 2020, 74 people signed up for the course. In 2021, 73 students signed up for the course. The data has the following columns: anonymised student ID, group numbers for each midterm, a grade for each midterm and the grade for the final exam and resit.

2.3 Grades of the exam questions

For the exams we have data on students anonymised student ID, so that we can compare this dataset with the data on the midterm. The dataset contains the grades for each sub question on the exam and which question is part of which subject.

Also note that some students got bonus points on the exam. This happened if they gave an exceptionally good answer to a question.

2.4 Groups

The midterms could be made in groups of two persons. This means that within a group, the grades of the midterms are totally dependent on each other. We need independence for some of the tests that we use when we compare the midterms, so for these tests we will take only one of the two group members. We also want to find the correlation between the midterm and exam, however there are 2 exam grades for some of the midterm grades. In this case it does matter which exam grade we choose.

2.5 Dependencies between years

There were 7 students that participated in the course in both years. These students already have some prior knowledge which causes concern. Additionally we will use tests that need independence, so for these tests the students will be removed. There were also 13 students that participated in 2019 and 2020, these

students were also removed if necessary.

This does mean that students that scored worse in 2020 were removed, so it is better to remove all the students that scored lower than a 5 on the exam when comparing the two years, which means that the power of the test is decreased when we compare the year 2020 and 2021.

2.6 Detailed grades of the midterms

For the renewal midterm of 2020, there were partial grades given for each question on the midterm. This means that we have 12 columns of data to analyse. For the other two midterms in 2020, there is no data available.

The renewal midterm in 2021 had 4 questions to analyse. The branching and Brownian midterm both had 5 questions.

2.7 Survey

The students in 2021 got a survey that was non-mandatory. 43 students replied to the survey of which 5 students were not useful as these students did not make one of the midterms or the exam. Students could indicate which videos they watched, what their favourite video was and if the content of the videos matched their expectation. They could also report how long they spent studying each subject and whether they had a course on queuing theory before. The last three questions on the survey were: "Did the videos lead to new ideas?", "Did the videos make renewal processes more interesting?" and "Did the videos make the course more interesting?".

2.8 Interpreting figures

In this report, the figures from the year 2020 will be lighter in colour, while the figures in 2021 will be darker in colour. The subject: renewal processes will be green, branching processes orange, Brownian motion magenta, probability theory gray and histograms of the exam will be blue.

2.9 Significance level

We will use a significance level of 0.05 in our tests.

3 Statistical tests

3.1 Mann-Whitney U test

The assumptions of this test are that the data is independent and that the data is ordinal or continuous. Under the null hypothesis, the distributions are equal, while the alternative hypothesis is that the distributions are not equal. If the distributions are similar, the test will mainly indicate differences in the

median. If the distributions are not equal then interpreting the result of the test is harder as it can be both the distribution that causes the difference as well as the location [2].

Grades are ordinal data and we can guarantee independence by removing the students that participated multiple times in the course for example. This means that we satisfy the assumptions of the Mann-Whitney U test when we compare the grades between years.

3.2 Pearson's chi-squared test

We will often want to know whether two or more categorical variables are independent for this project, for example is passing rate associated with the year that you followed the course in? For this we will use the Pearson's chi-squared test. The two events that we analyse need to be mutually exclusive and sum up to 1. In the previous example we have 4 categories, for example passed in 2020 is one category, and for each category you need at least 5 observations. If this assumption is not met we will use the Fisher Exact test, which only has the assumption of categorical values.

3.3 Kendall tau correlation coefficient

To study learning retention we need to relate the midterms and the exam. We cannot analyse the differences between grades, because these do not have much value. The difference between a 10 on the midterm and a 7 on the exam is not the same as the difference between a 6 on the midterm and a 3 on the exam, even though the difference in grades is 3 in both cases. Thus we will look into the ranks a student got on the midterm and exam, as we do expect these to be similar. We will use the Kendall tau correlation coefficient to calculate the correlation between the ranks of the midterm and the rank of the exam.

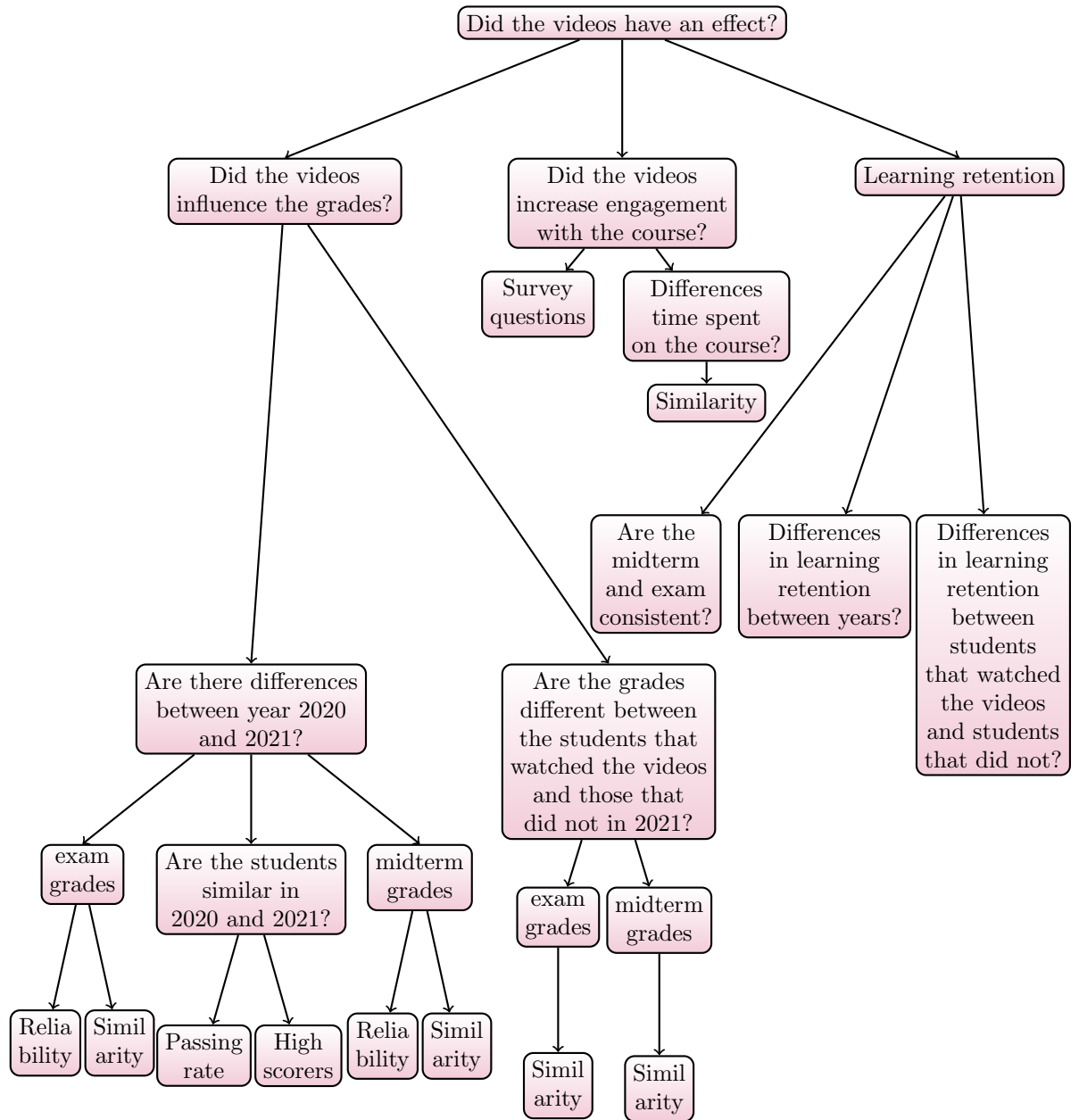
3.4 Chronbach's alpha

We want to compare the grades of the midterms and exams between the years 2020 and 2021. The questions on these exams were different however, so just comparing the two sets of grades is not fair. To see if we can compare the midterms and exams we will look into the passing rate and the reliability of the exam.

We will use the Chronbach's alpha coefficient to measure the reliability of each exam and midterm. Reliability means that under the same circumstances we would also get the same exam results. For a midterm the TU/e assessment course recommends a chronbach alpha value higher than 0.65. For the exam they recommend a value higher than 0.8.

The Chronbach's alpha has been controversial the past years, because it might be inaccurate and it's requirements are hard to fulfill. We will not discuss this further as it is outside of the scope of this report.

3.5 Picture of the decision tree



4 Decision trees

We will discuss our three research questions and how to answer these in this section. In section 3.5, the decision tree that we will use to get a conclusion for our three questions is given.

4.1 Were there differences in the grades between years 2020 and 2021?

We want to compare the grades of the renewal processes midterms between years. We also want to compare the grades on the renewal processes part of the exam to answer this question. To compare these two sets of data we will use the Mann-Whitney U test. This means that we need to account for independence as mentioned in section 2.5. We also need to take the average level of the students per year into account which is discussed in section 4.1.3 and whether the midterms and exam are similar between the two years which is discussed in section 4.1.1.

4.1.1 Reliability of the midterm and exam

The midterms and exam are different between the years, otherwise students would have prior knowledge in 2021 about the examinations. This means that just comparing the two sets of data is not fair. That is why we have to study whether the questions on the midterms and the exam are of a similar difficulty and if the questions are discriminatory. Discriminatory questions discriminate between students, so students that performed well on the test will score better on these questions than students that did not do well on the exam. After we identified the non-discriminatory questions, we will remove these from the midterms and the exam and once again compare the re-scaled grades with the Mann-Whitney U test.

4.1.2 Are there differences between years based based on the passing rate?

The passing rate of the course in both years also indicates whether the course was easier in one year. If the passing rate does not differ based on the year it indicates that the course was of a similar difficulty in both years.

It is also interesting to study high scoring students as the videos might have less effect on these students as they already score high on the midterm and exam. On the other hand the videos could also have more effect on these students as it might give them new insight which help with the material of the course. It is also another way to see if the two student populations differ.

4.1.3 Similarity of the students

We also have the branching and Brownian midterms that we can compare between the two years. For these two midterms there was no intervention, so by comparing these two midterms we can answer the question if students were better in one of the two years.

On the exam we also have the parts about branching, Brownian and probability theory, which we can use to get an indication of the level of the students.

4.2 Did the students that watched the videos score better on the renewal midterm and the renewal part of the exam?

These students were in the same year as their counter parts who did not view the video, so it is easier to compare these groups as we do not have to worry about the differences in the midterms or exam itself.

If the group of students that watched the video performed better on the renewal midterm than their counterparts, it means that the videos might have had an effect. However, we still need to check for the similarity of the two groups by comparing the branching and Brownian parts of the midterms and exam.

4.2.1 Did the students that watched the videos perform better on the branching and Brownian midterms?

We can also compare the branching and Brownian midterms to see if there are any differences between the students that watched the videos and those that did not. If these students performed better in general their grades will have to be adjusted when we compare the grades for the renewal midterm.

4.3 Did the videos improve learning retention?

The videos might also have an effect on learning retention as the videos add another way to interact with the material. We cannot answer this question by studying the differences in the midterm and exam grades, because the difference between a 5 and a 6 is different than the difference between a 7 and 8. We can however, correlate the ranks of the midterm and the exam to see whether obtaining a certain rank on the midterm also means that you get a similar rank on the exam and this might give an indication of whether the videos improved learning retention if we compare the correlation in 2020 with 2021.

4.3.1 Consistency of the midterms and the exam.

We would like it if the midterm grades can explain something about the final exam grade, as this might mean that the course is structured well or that the content of the midterms matches the content of the exam. If this is not the case, then conclusions about the exam cannot be compared to the conclusions of the midterm.

To answer this question we can look at whether scoring above average on the midterms is correlated with scoring higher on the exam.

4.4 Did the videos affect motivation?

To answer this question we can look at their self-reported motivation in the survey and whether the students thought that the videos enriched the course.

4.4.1 Did students spend more time on renewal theory because of the videos?

We also have data on the time spent by each student on each of the three subjects. With a Mann-Whitney U test we can compare the reported time spent on renewal processes of the groups that watched the videos versus those that did not. It might be possible that the students that watched the videos spent more hours on renewal theory as well, because the videos had new ideas and questions to work out or these students were more motivated.

4.4.2 Did the students that watched the videos also spend more time on the other two subjects?

The students that watched the videos might spend more time on the course in general. To answer this question we can compare the time spent on branching and Brownian theory. If the group that watched the videos also spent more time on these two subjects, then we have to take this into account for the analysis of the time spent on renewal processes.

5 Is the passing rate of students different between years?

Comparing the students between years is quite hard as the midterms and the exams were different. One of the ways to check whether these are different is to check the passing rate of the students that tried the course in both years after the first exam.

Independence: No
Non-discriminatory questions removed: No
One to one relation: No
Number of Students used in the comparison: 74-73

Table 1: The assumptions for the tests in section 5 are given here.

	Passed the course	Did not pass the course	
2020	47	27	74
2021	45	28	73
	92	55	147

Table 2: This table shows the number of students that passed and did not pass in a year.

Contingency table 2 relates the variables pass or did not pass and the years 2020 and 2021. A Pearson chi-square test with 1 degree of freedom is appropriate here as each cell in the table has a value higher than 5. The chi square statistic for this contingency table is $0.549 < 3.841$, which is smaller than the critical value. So we do not reject the null hypothesis.

5.1 Were there more students that passed the course with an 8 or higher in one year?

If there are more students in one year that performed better, the videos might have had an effect on the grades.

	Course grade ≥ 8	Course grade < 8	
2020	14	60	74
2021	5	68	73
	19	128	147

Table 3: This table shows the number of students that passed and did not pass in a year.

In table 3 we can already see a large difference in the number of students that got an eight or higher as their final grade. The chi square statistic for the Pearson chi square test is $4.7566 > 3.841$, so in this case we do reject the null hypothesis. So the proportion of students that got a high final grade is higher in 2020. The exam was the largest part of the final grade, so this result points to either students being better in 2020 or the exam being easier in 2020.

6 Comparing the renewal midterm between years

We want to answer the question whether the videos had an affect on the grades for the renewal midterm.

Independence: Yes
Non-discriminatory questions removed: No
One to one relation: No
Number of Groups used in the comparison: 30-28

Table 4: The assumptions for the tests in section 6 are given here.

6.1 Descriptive statistics

The average grade for the students that passed the course for the renewal midterm in 2020 is 7.89, with a median of 8.150. The standard deviation is 1.32. In 2021, the average grade is a 8.39, with a median of 8.450. The standard deviation is 1.14. The histograms in figure 1 also show that the distributions are relatively similar, although students scored full marks more often in 2021 than in 2020.

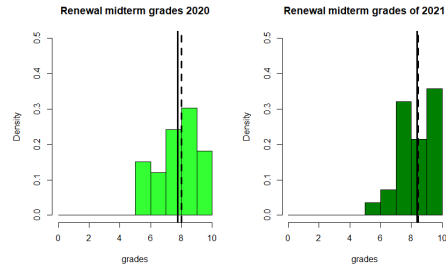


Figure 1: A histogram showing the grades for the renewal midterm in both years for the students that passed the course. The black line is the mean grade and the dashed line is the median grade.

6.2 Statistical test

The two sided Mann-Whitney U test gives a p-value of 0.1583. So there is not enough evidence to reject the null hypothesis.

6.3 Conclusion and discussion

There is no statistically significant difference between the two years, so this indicates that the videos had no major effect on the grades of the renewal processes midterm. There are still some things that first need be checked. As

a control, how did the students do on the other midterms, which we will look into in section 7. Also, the midterms were also different in both years, so the comparison may not be totally fair. To solve this problem we will also look if the midterms contains non-discriminatory questions.

7 Comparing the branching and Brownian midterms between years

There were no videos for the branching and Brownian subjects, so these two midterms can be taken as a control to assess whether the student populations were inherently different in the two years.

Independence: Yes
Non-discriminatory questions removed: No
One to one relation: No
Number of Groups used in the comparison: 30-28

Table 5: The assumptions for the tests in section 7 are given here.

7.1 Branching Midterm

7.1.1 Descriptive statistics

The mean of the branching midterm of 2020 for the students that passed the course is 8.73, the median is 9.00 and the standard deviation is 1.10. In 2021, the average grade is 9.03, the median is 9.100, with a standard deviation of 0.68. The standard deviation of the midterm in 2021 is small, which indicates that the midterm was not discriminatory. Note that both midterms have a high average grade.

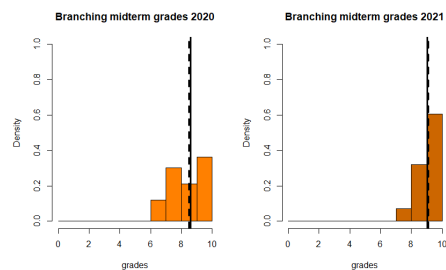


Figure 2: A histogram showing the grades for the branching midterm in both years for the students that passed the course. The black line is the mean grade and the dashed line is the median grade.

In figure 2, you can also see that the grades are clustered around 9 and 10 in 2021.

7.1.2 Statistical test

The two sided Mann-Whitney U test reports a p-value of 0.4447, so we do not reject the null hypothesis.

7.2 Brownian midterm

7.2.1 Descriptive statistics

The average grade of the Brownian midterm in 2020 for the students that passed the course is 7.48, with a median of 7.50 and a standard deviation of 1.14. In 2021, the average grade was 8.68, with a median of 8.700 and a standard deviation of 1.15.

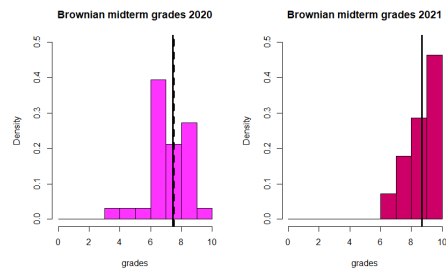


Figure 3: A histogram showing the grades for the Brownian midterm in both years for the students that passed the course. The black line is the mean grade and the dashed line is the median grade.

In figure 3 the distributions do not seem very similar, with the year 2020 being symmetric around 7-8, while the grades in 2021 are skewed to the left, even though the median and mean overlap.

7.2.2 Statistical test

The two sided Mann-Whitney U test has a p-value of 0.000341, which is significant. Thus we do reject the null hypothesis. This means that the students scored better on the Brownian midterm in the year 2021.

7.3 Conclusion on the branching and Brownian midterm:

Only the midterm on Brownian theory was made significantly better in 2021. Students also scored higher on both the renewal midterm and branching midterm

but not significantly better. Another remark is that a lot of students scored full marks during the midterms in 2021, so this points to the students being better in 2021. Although in section 5, we noticed that both the passing rate and amount of students scoring above an 8 on the course were higher in 2020. The midterms could also be easier, so we will also look into the reliability of the midterms.

8 Reliability of the midterms

First we want to identify the questions that were not discriminatory and had a high or low average. Removing these questions keeps the ordering of the students in a year the same, but the comparison between years becomes better because we adjust for either a hard or easy question. After that we can redo the comparison between the midterms as done in sections 6 and 7.

8.1 Procedure

First we will measure the overall reliability of the midterm or exam with the Chronbach alpha coefficient. Then we look at the average score of each question and the correlation of the question with the other questions in the exam. If the score is below 0.15 or above 0.85 of the total score on the question or the correlation is below 0.15 this might indicate that the question was non-discriminatory. After that the students are divided in 5 groups and for every question the mean grade for that group is calculated. By plotting these 5 values it is easier to see if a question is non-discriminatory.

Note that this analysis was done with all the students in a year, except for the students that scored a 0 for the midterm. These students scored a 0 on each question, so they make questions that are too easy harder to spot, while not adding values for the other questions. We use this set of students because it makes the differences between the low and high scoring students larger and thus it is easier to see which questions are discriminatory.

8.2 Renewal midterm 2020

The renewal midterm has an average score of 7.42. This average does include the students that redid the course or scored badly, so this is a different number than in section 6.1.

The midterm has a Chronbach's alpha coefficient of 0.788, so the midterm is considered to be reliable.

In figure 4, we can find a plot of the questions that are non-discriminatory.

8.3 Renewal midterm 2021

The Chronbach's alpha coefficient of this midterm is 0.529, which is lower than the 0.65 that is advised for midterms. There are only 4 questions in this midterm

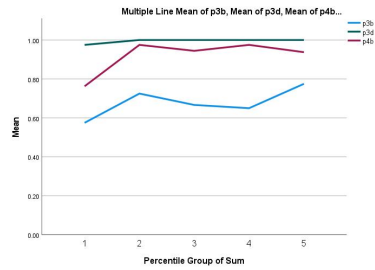


Figure 4: The grades per quintile for questions 3b, 3c and 4b of the renewal processes midterm in 2020. These three questions are not very discriminatory.

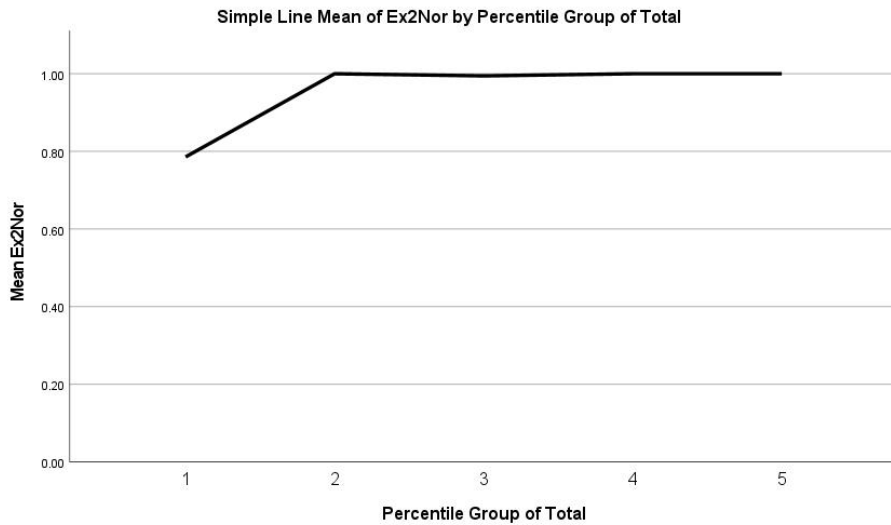


Figure 5: Question 2 does not differentiate well between students.

and these have a relatively low correlation of 0.3 with each other, this also leads to the low value of the Chronbach's alpha coefficient. Question 2 also has an average score of 0.98 out of 1. The score per quintile can be seen in figure 5. We will remove this question in the next analysis.

8.4 Branching and Brownian midterm 2020

For these two midterms there was no data available per question, so we did not check the reliability of these midterms. We will not re-analyse the midterms about branching processes and Brownian motion, while we drop the non-discriminatory questions, because we could not perform a reliability analysis for these midterms in 2020.

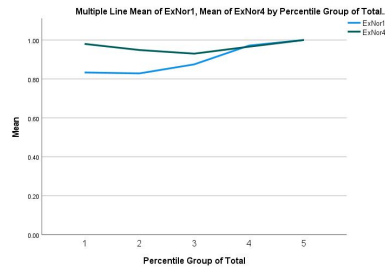


Figure 6: The grades per quintile for questions 1 and 4 of the Brownian midterm of 2021. These two questions are not very discriminatory.

8.5 Branching midterm 2021

This midterm has a Chronbach's alpha coefficient of 0.643 with 5 questions. Question 1 has an average score of 0.97 out of 1 and a correlation of 0.115 with the other questions, so this is not a good question. Question two also has a high average score of 0.94 out of 1. In general, the score of each question of this midterm is very high, which we also noticed in section 7.1.

8.6 Brownian midterm 2021

The Chronbach's alpha coefficient of this Brownian midterm is 0.546, which is low. Question 4 has an average score of 0.96 out of 1 and has a negative correlation of -0.114 with the other questions, so this is a bad question. Question 1 also has an average of 0.91 and in figure 6 you can also see that it is not a discriminatory question.

9 Re-analysing the renewal midterm

In 2020, questions 3b, 3d and 4b were removed, because these questions were non discriminatory. In 2021 question 2 was removed.

Independence: Yes
Non-discriminatory questions removed: Yes
One to one relation: No
Number of Groups used in the comparison: 30-27

Table 6: The assumptions for the tests in section 9 are given here.

9.0.1 Descriptive statistics

Note that there is one group in 2021, that did not have grades per question, so instead of 28 groups, we have 27 groups. After removing the three questions mentioned in section 9 in 2020, students had an average score of 7.50 on the renewal midterm. The median score was 7.750 and the standard deviation is 1.54. In 2021, students scored on average 8.15, with a median of 8.00 and a standard deviation of 1.26. This could indicate that students performed slightly better in 2021, which is similar to what we observed in section 6.

9.0.2 Statistical test

The two sided Mann-Whitney U test reports a p-value of 0.1478, so there is not enough evidence to reject the null hypothesis.

9.1 Conclusion

There is not enough evidence to suggest that the grades for the midterms are different between years when corrected for non-discriminatory questions. The reliability of all 3 midterms in 2021 was low which may be caused by the small number of questions in each midterm.

10 Comparing the exam between years

A lot of students scored full marks during the midterm in 2021, so we want to answer the question whether students were better in 2021. The exam also has a part about renewal theory, so we are interested in whether the videos also positively impacted the grades for this part of the exam.

Independence: Yes
Non-discriminatory questions removed: No
One to one relation: No
Number of Students used in the comparison: 45-42

Table 7: The assumptions for the tests in section 10 are given here.

10.1 Renewal theory

10.1.1 Descriptive statistics

On average, students that passed the course scored a 6.185 on the renewal part of the exam in 2020. The median grade was 6.00 with a standard deviation of 2.28. In 2021, students scored on average a 5.57, with a median of 5.375 and a standard deviation of 1.51.

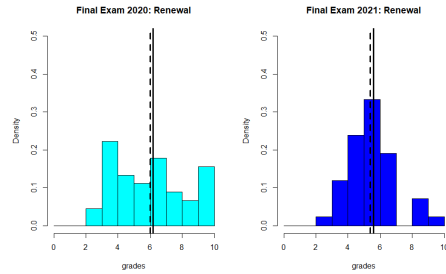


Figure 7: A histogram showing the grades for the renewal part of the exam in both years for the students that passed the course. The black line is the mean grade and the dashed line is the median grade.

The grades in 2021 had a definite peak around the value 5 which can be seen in figure 7, while in 2020, the distribution looks more uniform.

10.1.2 Statistical test

The two sided Mann Whitney U test gives a p-value of 0.3013, so we do not reject the null hypothesis.

10.2 Branching theory

10.2.1 Descriptive statistics

In 2020, the mean grade was 7.316, with a median of 7.60. The interquartile range is (6.00,9.60). The mean grade in 2021 was 7.336, with a median of 7.500. The standard deviation is 1.68.

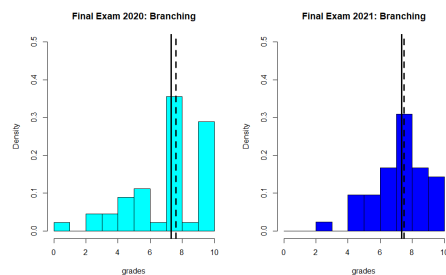


Figure 8: A histogram showing the grades for the branching part of the exam in both years for the students that passed the course. The black line is the mean grade and the dashed line is the median grade.

In figure 8, it is clear that a lot of students scored full marks on this part of the exam in 2020 or scored badly, while in 2021 there is a peak around 8 and

the distribution looks symmetric.

10.2.2 Statistical test

The two sided Mann-Whitney U test gives a p-value of 0.5199, so we do not reject the null hypothesis.

10.3 Brownian theory

10.3.1 Descriptive statistics

In 2020, the mean grade for the Brownian part of the exam is 8.133, with a median of 8.667. The interquartile range is (6.889,9.778). In 2021, the mean grade is 7.286, with a median of 7.455 and a standard deviation of 1.65.

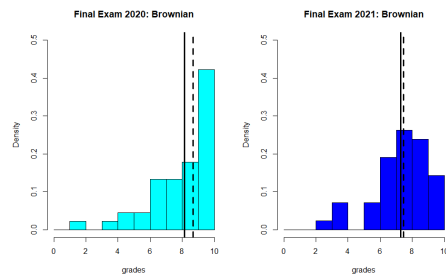


Figure 9: A histogram showing the grades for the Brownian part of the exam in both years for the students that passed the course. The black line is the mean grade and the dashed line is the median grade.

In figure 9, we can see that the students in 2020 scored full marks often, while this is not the case in 2021. This is also the reason why the data is skewed to the left in 2020.

10.3.2 statistical tests

The two sided Mann-Whitney U test gives a p-value of 0.006773, so we reject the null hypothesis. The students in 2020 scored full marks more often for this part of the exam, so in this case we can conclude that the students performed better on the Brownian part of the exam.

10.4 Probability theory

10.4.1 Descriptive statistics

In 2020, students scored on average a 4.933 for the probability part of the exam. The median grade is 5.00 and the standard deviation is 3.10, which is quite large compared to the other subjects on the exam. In 2021, the mean grade is 2.429 and the median is 2.100. The standard deviation is 1.74. The mean and median grade in 2021 is lower than in 2020 by 3 points, which is a big difference, when we compare it to the other parts of the exam. This is also one of the reasons that the final exam grade in 2021 is lower than in 2020.

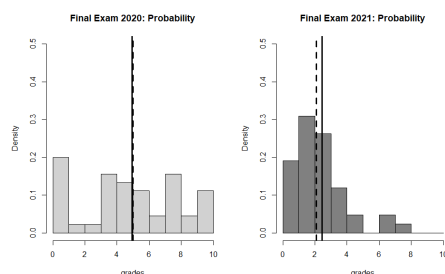


Figure 10: A histogram showing the grades for the probability part of the exam in both years for the students that passed the course. The black line is the mean grade and the dashed line is the median grade.

In figure 10, we can see that the grades in 2020 are relatively uniform, while the grades in 2021 are mainly clustered between 0 and 4. Also, none of the students in 2021 scored full marks on this subject.

10.4.2 Statistical test

The two-sided Mann-Whitney U test gives a p-value of 0.0001588, which indicates a significant difference between the two sets of grades. So the part about probability theory on the exam was made significantly better in 2020.

10.5 Final Exam grades

10.5.1 Descriptive statistics

The grade for the final exam was determined by combining the 4 parts above. The mean grade of the students that passed the course in 2020 was 7.157, with a median of 7.20 and a standard deviation of 1.38. In 2021, the average grade was 6.183, with a median of 6.235 and a standard deviation of 1.15.

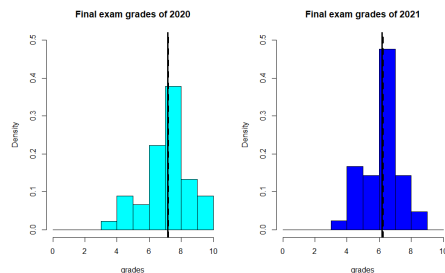


Figure 11: A histogram showing the grades of the exam in both years for the students that passed the course. The black line is the mean grade and the dashed line is the median grade.

In figure 11, we can see that the two sets of grades seem to follow the same distribution.

10.5.2 Statistical test

The two-sided Mann-Whitney U test reports a p-value of 0.0006586, which means that the exam grades differ significantly, thus we can conclude that the exam in 2020 was made better than in 2021.

10.6 Overview tests

Independence: Yes
Non-discriminatory questions removed: No
One to one relation: No
Number of Students used in the comparison: 45-42

Table 8: The assumptions for the tests in section 10.5 are given here.

Subject	p-value	Significant
Final Exam: Renewal	0.3013	No
Final Exam: Branching	0.5199	No
Final Exam: Brownian	0.006773	Yes
Final Exam: Probability	0.0001588	Yes
Final Exam grade	0.0006586	Yes

Table 9: A table showing the results of the Mann-Whitney U test for the comparisons of the exams in 2020 and 2021.

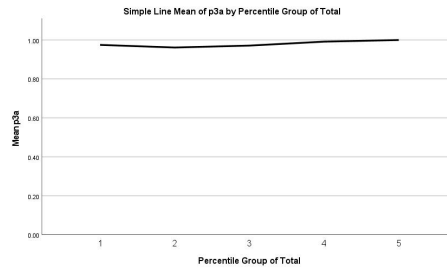


Figure 12: Question 3 of the exam in 2020 does not differentiate between students and has a very high average grade.

10.7 Conclusion

The results so far show that the students in 2020 performed significantly better on the Brownian and probability theory part of the exam. This is in contrast to the midterms, where we noticed that students in 2021 performed significantly better on the Brownian midterm. This points to the exam being harder in 2021 compared to 2020. Thus comparing the exams when the non-discriminatory questions are removed might give a better indication of whether the students are equal.

11 Reliability of the exam

In section 10.7, we remarked that the exam of 2021 might be harder, especially the probability part of the exam. We will look into the reliability of both exams with the same procedure as mentioned in section 8.1.

11.1 Is the exam of 2020 reliable?

The passing rate of this exam is 0.71, which is well within the accepted range of 0.5 up to 0.9. The Chronbach's alpha coefficient of the exam is 0.840.

All of the questions except for question 3 have a normalised score between 0.15 and 0.85, which means that the difficulty of most questions was good. Students scored on average on question 3 a 0.95 out of 1 and in figure 12 we can also see that this question does not differentiate between students because the score was equal among all quintiles.

Some other questions to note are questions 5a, 5b and 6a. The scores per quintile can be seen in figure 13. Question 5a does not differentiate well between students and this question has a low correlation with the other questions. Question 5b also does not differentiate the highest three quintiles. Question 6a differentiates only the lowest quintile from the other 4 quintiles, so it is also not a really discriminatory question.

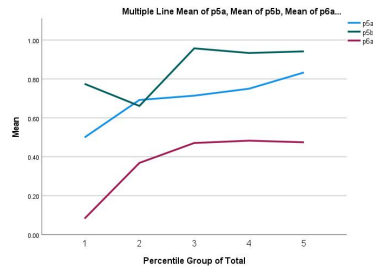


Figure 13: The questions that were removed in the exam of 2020 for the third analysis.

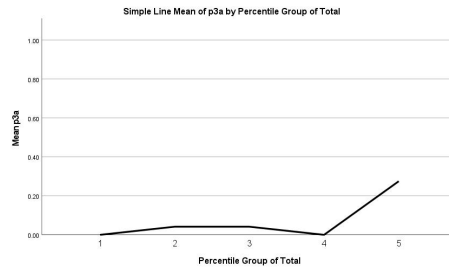


Figure 14: Question 3 about probability theory was made very badly and only top scoring students scored some points on it.

11.2 Is the exam of 2021 reliable?

This exam had a passing rate of 0.6. Using the alpha model in SPSS gives an Chronbach's alpha value of 0.818, which indicates that the exam is reliable. Questions 2b, 3 have a score below 0.15. Questions 1a, 1d and 5a have a score of 0.85 or higher, so these questions are most likely non-discriminatory. In figure 14, we can also see that question 3 was made badly by most students and only the top scoring students got awarded a small amount of points, so also this group doesn't differentiate itself well on this question.

In figure 15, we can see that 1a differentiates the lowest 40% of students from the highest 60% to a small degree. 2b differentiates the lowest 60% from the highest 40% instead. Question 2d only differentiates the lowest group quite strongly, but does not differentiate the rest of the groups. Question 4b and 5a differentiate between all the groups but only slightly, combined with their respective low and high average they do not make for great questions. The graphs for the other questions can be found in figure 36 and 37 in the appendix.

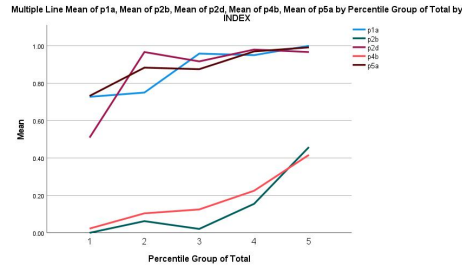


Figure 15: The questions that were removed in the exam of 2021 for the third analysis.

Question	Subject
3	Renewal processes
5a	Branching Processes
5b	Branching Processes
6a	Brownian motion

Table 10: The questions that were non-discriminatory on the exam of 2020 and their subject.

11.3 Method for re-analysing the exams

We will first analyse the exams again when question 3 is removed from both exams, as in both exams this question was non-discriminatory, which we saw in sections 11.1 and 11.2. We will also look into a third analysis where questions 3, 5a, 5b and 6a are removed in 2020 and questions 1a, 2b, 2d, 3, 4b, 5a are removed in 2021. These questions were more discriminatory than questions 3 in both exams, but still not as discriminatory as we would like. In tables 10 and 11, you can find the subject associated with each question.

Question	Subject
1a	Brownian motion
2b	Renewal processes
2d	Renewal processes
3	Probability theory
4b	Probability theory
5a	Branching processes

Table 11: The questions that were non-discriminatory on the exam of 2021 and their subject.

12 Exam analysis with non-discriminatory questions removed

In this section, we will analyse whether removing two questions that were non-discriminatory has any influence on the comparisons of the exam.

Final Exam	Mean	Median	Standard deviation
Renewal 2020	4.36	4.00	3.34
Renewal 2021	5.57	5.38	1.51
Probability 2020	4.93	5.00	3.10
Probability 2021	3.37	3.33	2.17
Final exam grade 2020	7.02	7.05	1.45
Final exam grade 2021	6.51	6.61	1.16

Table 12: The means, medians and variance of the subjects on the exam if the questions that are non-discriminatory are removed.

Independence: Yes
Non-discriminatory questions removed: Non-discriminatory questions removed
One to one relation: No
Number of Students used in the comparison: 45-42

Table 13: The assumptions for the tests in section 9 are given here.

12.1 Renewal theory

12.1.1 Descriptive statistics

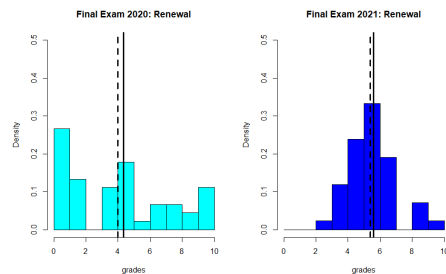


Figure 16: A histogram showing the grades of the renewal part of the exam in both years, while question 3 in 2020 was removed from the exam. The black line is the mean grade and the dashed line is the median grade.

Figure 16 and table 12 show that the median grade in 2021 is higher than in 2020. The distributions look different, with 2020 having peaks around the grades 0, 4 and 10, while 2021 has a peak around 6.

12.1.2 Statistical tests

The two sided Mann-Whitney U test reports a p-value of 0.03334, which is significant. So we reject the null hypothesis in this case. This means that the distribution of the grades of the exam in 2020 is different than the distribution of the grades in 2021.

12.2 Branching and Brownian theory

These subjects did not change, because there were no questions removed for this subject. For the results you can refer back to section 10.2 and 10.3.

12.3 Probability theory

12.3.1 Descriptive statistics

The mean and median in table 12 imply that this part of the exam was made better in 2020 than in 2021. Figure 17 also shows that the grades in 2020 are quite uniform between 0 and 10, while in 2021 most of the grades are between 0 and 5.

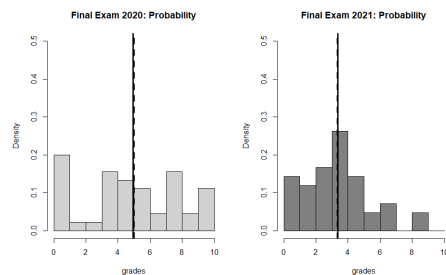


Figure 17: A histogram showing the grades of the probability part of the exam in both years, while the non discriminatory questions were removed from the analysis. The black line is the mean grade and the dashed line is the median grade.

12.3.2 Statistical test

The p-value for the two sided Mann-Whitney U test is 0.01061, so we reject the null hypothesis in this case. So we do find a statistically significant difference between the grades for the part about probability theory on the exam.

12.4 Final Exam grade

12.4.1 Descriptive statistics

The mean grade in 2020 was a 7.02, with a median of 7.05. The standard deviation is 1.45. In 2021, the mean grade was 6.51, with a median of 6.61. The standard deviation is 1.16. This difference in grades is smaller than the difference before removing the non discriminatory questions.

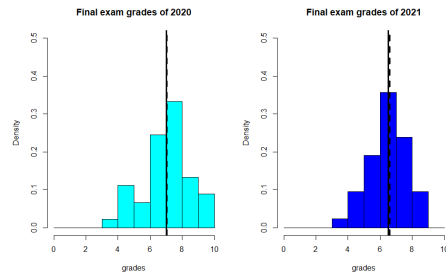


Figure 18: A histogram showing the grades of the exam in both years, while the bad questions were removed from the analysis. The black line is the mean grade and the dashed line is the median grade.

Figure 18 shows that the medians do not differ a lot and both histograms have a peak in the bracket 6 to 7.

12.4.2 Statistical test

The two sided Mann-Whitney U test gives a p-value of 0.08237, so there is not enough evidence to reject the null hypothesis.

12.5 Overview tests when we remove the non-discriminatory questions

Independence: Yes
Non-discriminatory questions removed: Yes
One to one relation: No
Number of Students used in the comparison: 45-42

Table 14: The assumptions for the tests in section 12 are given here.

Subject	p-value	Significant
Final Exam: Renewal	0.03334	Yes
Final Exam: Branching	0.5199	No
Final Exam: Brownian	0.006773	Yes
Final Exam: Probability	0.01061	Yes
Final Exam grade	0.08237	No

Table 15: A table showing the results of the Mann-Whitney U test for the comparisons of the exams in 2020 and 2021, when we remove the non-discriminatory questions.

12.6 Conclusion of removing the non-discriminatory questions

Students performed significantly better on the renewal part of the exam in 2021, while they performed worse on the Branching and probability part of the exam after we removed the non-discriminatory questions. There is not enough evidence to draw a conclusion on whether students performed better on the exam when we compare one year to the other. This gives some evidence that the videos had an effect on the part of the renewal processes theory.

13 Exam analysis with questions that had a low or high average removed.

In general questions that have a score above 0.85 or below 0.15 out of 1, are not really discriminatory. So comparing the exams when these questions are also removed can offer some more insight in whether the students performed equally well on the exam between years.

The exam of 2020 has 11 questions and 2021 has 12 questions. This means that we have fewer questions left per subject. This is mainly applicable to the subjects: renewal, branching and probability theory. In the exam of 2020, these subjects have only 2 questions after the removal of the questions with a very high or low score. The probability part of the exam in 2021 has only 1 question. Comparing this question with the two questions left in 2020 is not really meaningful because the difficulty of the questions is not averaged out.

Final Exam	Mean	Median	Variance
Renewal 2020	4.36	4.00	3.34
Renewal 2021	6.30	6.00	1.65
Branching 2020	6.18	7.00	3.76
Branching 2021	6.86	6.88	1.95
Brownian 2020	8.02	8.50	2.07
Brownian 2021	7.90	8.11	1.92
Probability 2020	4.93	5.00	3.10
Probability 2021	6.19	7.00	3.48
Final exam grade 2020	6.67	6.73	1.65
Final exam grade 2021	7.12	7.39	1.33

Table 16: The means, medians and variance of the subjects on the exam after the questions with a very high or low average score were removed.

Independence: Yes Non-discriminatory questions removed: Questions with a very high or low average removed One to one relation: No Number of Students used in the comparison: 45-42

Table 17: The assumptions for the tests in section 9 are given here.

13.1 Renewal theory

13.1.1 Descriptive statistics

From figure 19, we already get an indication that the renewal part of the exam is made better in 2020 than in 2021. This is also reflected in table 16, as the mean and median are higher in 2021.

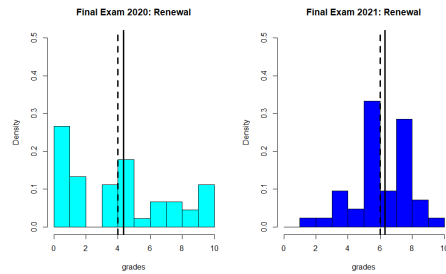


Figure 19: A histogram showing the grades of the renewal part of the exam in both years, while the questions with a very high or low score were removed from the analysis. The black line is the mean grade and the dashed line is the median grade.

13.1.2 Statistical test

The two sided Mann-Whitney U test gives a p-value of 0.002015, so we reject the null hypothesis. This means that the students performed significantly better on the renewal part of the exam in 2021, when we remove the non-discriminatory questions.

13.2 Branching theory

13.2.1 Descriptive statistics

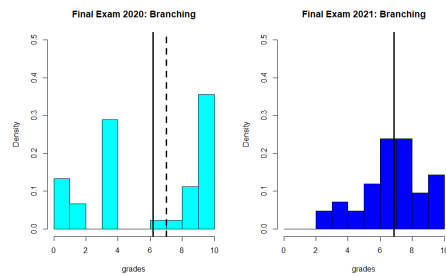


Figure 20: A histogram showing the grades of the branching part of the exam in both years, while questions with a very high or low score were removed from the analysis. The black line is the mean grade and the dashed line is the median grade.

Figure 20 shows that the distributions are very different, with the exam of 2020 showing a lot of gaps in between peaks at 1, 4 and 10. The grades of the branching part of the exam in 2021 are symmetric and have a peak at 7.

13.2.2 Statistical test

The two sided Mann-Whitney U test reports a p-value of 0.9353, so we do not reject the null hypothesis.

13.3 Brownian theory

13.3.1 Descriptive statistics

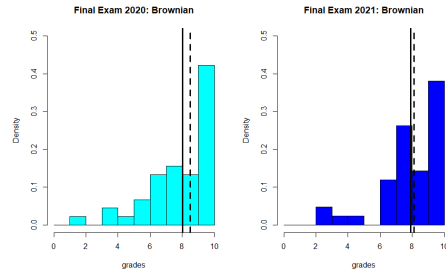


Figure 21: A histogram showing the grades of the Brownian part of the exam in both years, while some questions were removed from the analysis. The black line is the mean grade and the dotted line is the median grade.

In figure 21, we observe that the distributions are similar. Table 16 also shows that the mean and median are quite close.

13.3.2 Statistical test

The two sided Mann-Whitney U test gives a p-value of 0.5084, so we do not reject the null hypothesis.

13.4 Probability theory

For 2020 there are two questions left, while for the exam of 2021 there is only 1 question left. This means that comparing these two subjects is not meaningful anymore.

13.5 Final Exam grade

13.5.1 Descriptive statistics

After removing the questions with a very high and low score and then re scaling the grades, we get a mean exam grade of 6.67 in 2020, the median is 6.73 and the standard deviation is 1.65. In 2021, the mean grade in 7.12, with a median of 7.39 and a standard deviation of 1.33.

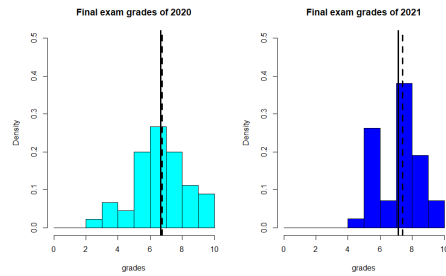


Figure 22: A histogram showing the final exam grades in both years, while questions with a very high or low average were removed from the analysis. The black line is the mean grade and the dashed line is the median grade.

The histograms in figure 22 show that the exam grades in 2021 are higher than in 2020, which is in contrast to section 10.5.2.

13.5.2 Statistical test

The two sided Mann-Whitney U test reports a p-value of 0.1661, so we do not reject the null hypothesis.

13.6 Conclusion of the exam comparison while removing questions with a very high or low average

From this section, we still draw the conclusion that the students performed similar on the exam. Also in this case the students in 2021 did perform better for the part of the exam about renewal theory, which indicates that the videos might have an effect.

13.7 Conclusion about differences on the exam

When we consider all the questions on the exam we do not find a significant difference on the renewal subject of the exam, while if we remove the non-discriminatory questions we do find a difference in favor of 2021 for the renewal part of the exam.

We will use the comparison of the exam where the non-discriminatory questions are removed. So for the exam we do find a significant difference.

14 Conclusion: Did the videos improve the grades in 2021?

We did not find significant differences between the renewal midterms. However, we did find a difference on the renewal part of the exam after we excluded the non-discriminatory questions.

14.1 Conclusion about the difference in students between years

Subject	p-value	Significant
Renewal midterm	0.1478	No
Branching midterm	0.4447	No
Brownian midterm	0.000341	Yes
Final Exam: Renewal	0.03334	Yes
Final Exam: Branching	0.5199	No
Final Exam: Brownian	0.006773	Yes
Final Exam: Probability	0.01061	Yes
Final Exam grade	0.08237	No

Table 18: A table showing the statistical tests between the midterms and exam of both years. The test used was the two sided Mann-Whitney U test. The null hypothesis is that the distribution are different. The alternative hypothesis is that the distributions are different.

For the renewal midterm we took the test with the non-discriminatory questions removed as described in section 9.1. For the branching and Brownian midterm we took the conclusion described in section 7.3. For the exam we took the conclusion described in section 12.6.

15 Analysing the differences between the students that watched the videos and the students that did not.

15.1 Students that watched the videos

There were 19 students that watched at least one video, with an average of 2.84 videos per person and a standard deviation of 1.98 videos. There were 3 students that watched all the videos and 8 students that watched only one video. All of these students also filled in how many hours they spent on each subject.

15.2 Students that did not watch the videos

There were 47 students that filled in the survey. From these, 19 watched some videos and 28 did not watch any videos. One of the students that did not watch the videos, does mention that the first video is their favourite, so maybe this student only watched the first few minutes or miss clicked this question. There is also one student that did not report how much time they spent on each subject, so for the time analysis this student will be removed.

15.3 Students that did not fill in the survey

There were also 13 students that did not fill in the survey. These students probably did not watch the videos, but we cannot assume this. We will remove these students from the analysis.

15.4 Did the content of the videos match the content of the lecture?

Twenty students reported that the content of the videos matched the content of the lectures, while 9 students reported that the content of the two did not match. This means that the videos had the right content. Note that 29 students reported on this question while there were only 19 students who watched any videos.

15.5 Groups for the midterms

There were 55 students that made all the midterms and the exam. For the midterms these students worked in 34 groups. Twelve students worked alone and the rest worked in groups of two. There was one student that did not make the Branching midterm and exam in one group of two students, this student did work on the renewal processes midterm, so we will consider this a group instead of the other student working alone. In table 19, you can find how many students that worked alone, filled in the survey. In table 20, you can find a

Did not fill in the survey	Did not watch the videos	Watched the videos
3	5	4

Table 19: How many students watched the videos while working alone

	NF	NW	W
NF	0	-	-
NW	3	4	-
W	7	8	0

Table 20: The amount of groups based on whether a student did not fill in the survey(NF), the student did not watch videos(NW) or the student watched the video(W). The place in the table indicates the intersection of the group.

similar division of the students that worked in groups of 2. To test whether the videos had an effect on the midterm we will split the groups of the midterm in two groups. One group will be the groups where at least one student watched the videos. In the other group we have the groups where none of the students watched the videos. There were 3 students that worked alone and did not fill in the survey, these will be removed from the analysis, which leaves us with 31 groups for the midterm. The 10 groups where only one of the students filled in the survey will be classified according to the student that did fill in the survey. So in the group that did not watch the videos we have $5 + 3 + 4 = 12$ groups. In the group that watched the videos we have $4 + 7 + 8 = 19$ groups. This means that we do have a small amount of groups to compare, so the power of the tests will be small.

16 Did the videos improve the renewal midterm grades of the students?

Independence: No
Non-discriminatory questions removed: No
One to one relation: No
Number of Groups used in the comparison: 12-19

Table 21: The assumptions for the tests in section 16 are given here.

16.1 Summary statistics

For the groups where none of the student watched the videos, the average score on the renewal midterm is 8.66, with a standard deviation of 0.92. The groups that did watch the videos scored on average 7.91, with a standard deviation of

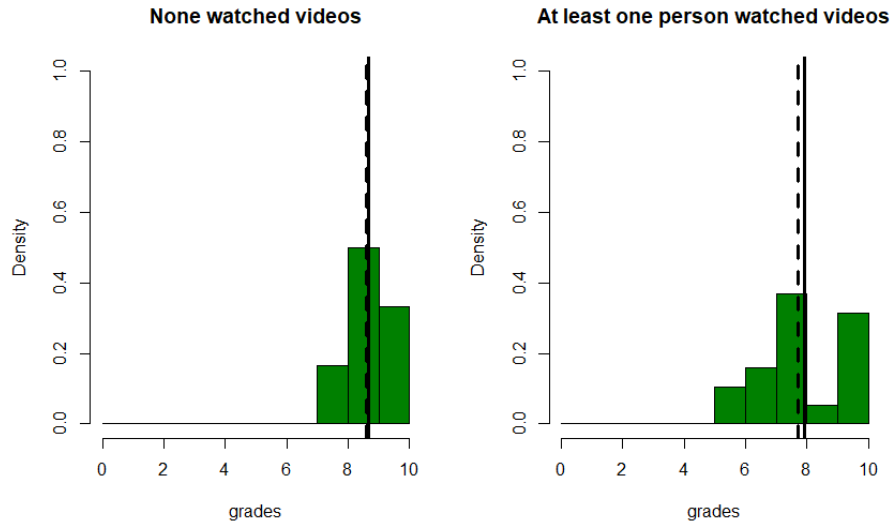


Figure 23: Here you can see the histograms of the grades for the renewal midterm for the three different groups. The black line is the mean and the dashed line is the median.

1.38. Figure 23 shows that the groups that did watch the videos also scored twice below a 6, which lowers the average grade.

16.2 Statistical test

The Mann-Whitney U test gives a p-value of 0.1078, so we do not reject the null-hypothesis.

16.3 Conclusion

We observed that the groups where at least one person watched the videos scored lower on average on the midterm, although this is not a significant difference. After doing a post-hoc power analysis, we get a power of 0.52 for this test, which is lower than the advised 0.8. So if we collect more data this might lead to a different conclusion. Also note that there were two groups where at least one student watched the videos that scored below a 6 on this midterm. While every group where none of the students watched a video scored above a 7. If we increase the sample size these outliers might even out.

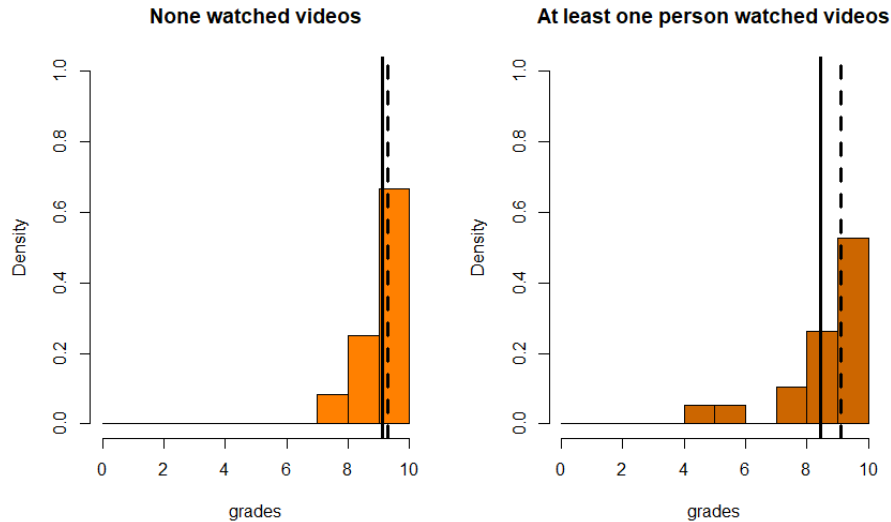


Figure 24: Here you can see the histograms of the grades for the branching midterm for the three different groups. The black line is the mean and the dashed line is the median.

17 Do the students that watched the videos perform differently on the other two midterms?

Independence: No
Non-discriminatory questions removed: No
One to one relation: No
Number of Groups used in the comparison: 12-19

Table 22: The assumptions for the tests in section 17 are given here.

17.1 Branching Midterm

17.1.1 Descriptive statistics

For the groups where none of the student watched the videos, the average score is 9.108, with a standard deviation of 0.773. The groups that did watch the videos scored on average 8.447, with a median of 9.100. The inter quartile range is (8.400,9.250). Figure 24, shows that for both groups, half of the midterm groups scored within the bracket of 9 to 10. For the groups that watched the videos there are also two groups that scored a 4.4 and a 5.2, which lowers the average grade.

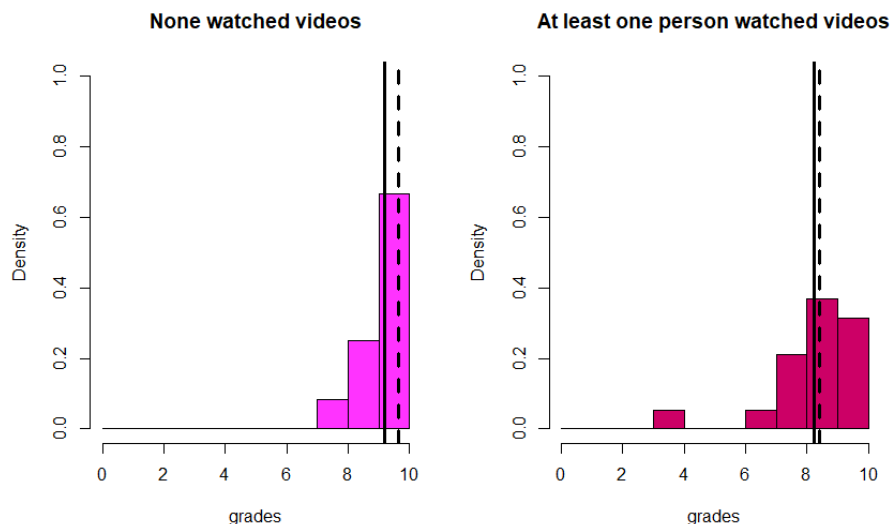


Figure 25: Here you can see the histograms of the grades for the Brownian midterm for the three different groups. The black line is the mean and the dashed line is the median.

17.1.2 Statistical test

The Mann-Whitney U test gives a p-value of 0.1545, so we do not reject the null hypothesis.

17.2 Brownian Midterm

17.2.1 Descriptive statistics

For the groups where none of the student watched the videos, the average score for the Brownian midterm is 9.204, with a standard deviation of 0.930. The groups that did watch the videos scored on average 8.224, with a standard deviation of 1.538. In figure 25, we can also see that the groups where none of the students watched the videos scored between a 9-10 more often than the groups where at least one student watched the videos. There is also one group where at least one student watched the videos that scored a 3.3.

17.2.2 Statistical test

The two sided Mann-Whitney U test gives a p-value of 0.0365, so we reject the null hypothesis. In this case the groups that did not watch the videos performed significantly better than the groups where at least one student watched some videos.

17.3 Conclusion

The groups where none of the students watched the videos performed significantly better on the Brownian midterm. We also observed that the groups that did not watch the videos performed better on the renewal and branching midterms, although this was not a significant difference.

Independence: No
Non-discriminatory questions removed: No
One to one relation: No
Number of Groups used in the comparison: 12-19

Table 23: The assumptions for the tests in section 17.3 are given here.

Subject	Test	P-value	Significant?
Renewal midterm	Mann-Whitney U test	0.1078	No
Branching midterm	Mann-Whitney U test	0.1545	No
Brownian midterm	Mann-Whitney U test	0.0365	Yes

Table 24: The p-values for comparing the groups where none of the students watched the videos versus the groups where at least one student watched the videos with a two sided Mann-Whitney U test.

18 Do the students that watched the videos differ on the exam?

In this section, we will compare the group of students that watched the videos with the group of students that did not watch the videos. From the 43 students that filled in the survey and made the exam, 19 indicated that they watched the videos. This is a relatively small sample size for the Mann-Whitney U test, so the power of the test will be quite small.

In this case we are also comparing groups of students that made the same exam, so we can use all the questions in the exam, even question 3b which had a very low score as discussed in section 12.

Independence: No
Non-discriminatory questions removed: No
One to one relation: No
Number of Students used in the comparison: 24-19

Table 25: The assumptions for the tests in section 18 are given here.

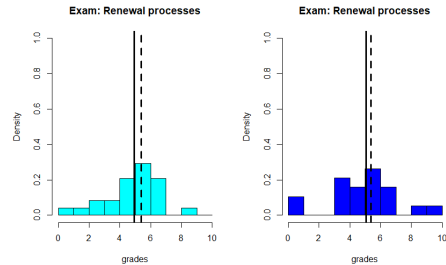


Figure 26: These histograms show the grades for the renewal part of the exam. The left one is the group of students that did not watch the videos. The right histogram is for the students that did.

18.1 Renewal theory

18.1.1 Descriptive statistics

The students that did not watch the videos scored on average a 4.94, with a median of 5.38 and a standard deviation of 1.80. The students that watched the video scored on average a 5.04, with a median of 5.38 and a standard deviation of 2.34. In figure 26, we can see that the distributions look relatively similar.

18.1.2 Statistical test

The Mann-Whitney U test reports a p-value of 0.9609, so we do not reject the null hypothesis.

18.2 Branching theory

18.2.1 Descriptive statistics

The students that did not watch the videos scored on average a 6.83, with a median of 7.15 and a standard deviation of 2.23. The students that did watch the video scored on average a 6.40, with a median of 7.40. The inter quartile range is (5.25,7.90). In figure 27, we can see that the grades for the students that watched the video are skewed to the left. Most students scored between a 7 and 8, but there were also students that scored badly on this part of the exam.

18.2.2 Statistical test

The Mann-Whitney U test reports a p-value of 0.8067, so we do not reject the null hypothesis.

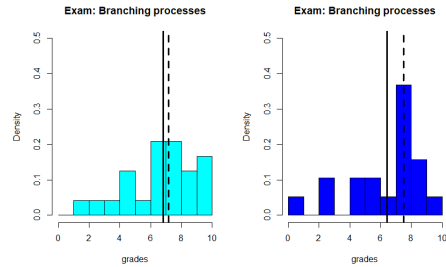


Figure 27: These histograms show the grades for the branching part of the exam. The left one is the group of students that did not watch the videos. The right histogram is for the students that did watch some videos. The black line is the mean grade and the dashed line is the median.

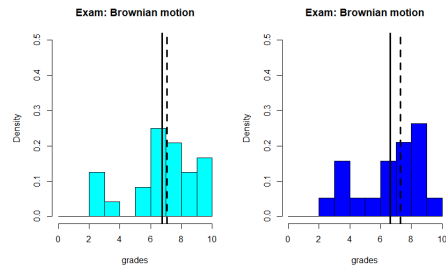


Figure 28: These histograms show the grades for the Brownian part of the exam. The left one is the group of students that did not watch the videos. The right histogram is for the students that did.

18.3 Brownian theory

18.3.1 Descriptive statistics

The students that did not watch the videos scored on average a 6.74, with a median of 7.05. These grades had a standard deviation of 2.04. The students that watched the videos scored on average a 6.61, with a median of 7.27 and a standard deviation of 2.17. In figure 28, we can see that the distributions look relatively similar.

18.3.2 Statistical test

The Mann-Whitney U test reports a p-value of 0.9902, so we do not reject the null hypothesis.

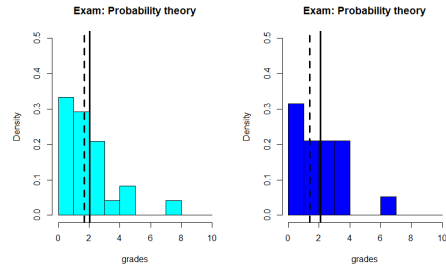


Figure 29: These histograms show the grades for the probability part of the exam. The left one is the group of students that did not watch the videos. The right histogram is for the students that did.

18.4 Probability theory

18.4.1 Descriptive statistics

The students that did not watch the videos scored on average a 2.05, with a median of 1.70 and a standard deviation of 1.71. The students that did watch the videos scored on average a 2.07, with a median of 1.40. The interquartile range is (0.50,3.20). In figure 29, we can see that the distributions look relatively similar.

18.4.2 Statistical test

The Mann-Whitney U test reports a p-value of 0.9314, so we do not reject the null hypothesis.

18.5 Final exam grade

18.5.1 Descriptive statistics

The average exam grade of the students that did not watch the videos was 5.95, with a median of 6.20 and a standard deviation of 1.71. The students that did watch the videos scored a 5.81 on average. The median is 6.44. The interquartile range is (4.77,7.02). In figure 30, we can see that the distribution look relatively similar.

18.5.2 Statistical test

The Mann-Whitney U test gives a p-value of 0.8737, so we do not reject the null hypothesis.

18.6 Recap

Here we want to answer whether the students that watched the videos about renewal theory perform better on the exam compared to the students that did

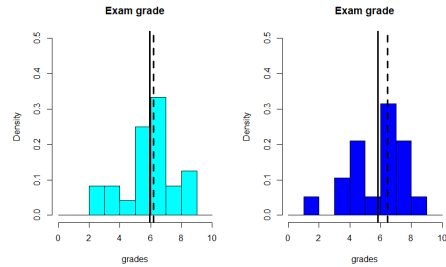


Figure 30: These histograms show the grades for the final exam grade. The left one is the group of students that did not watch the videos. The right histogram is for the students that did.

not watch the video. The two groups made the same exam at their own, so the students are independent.

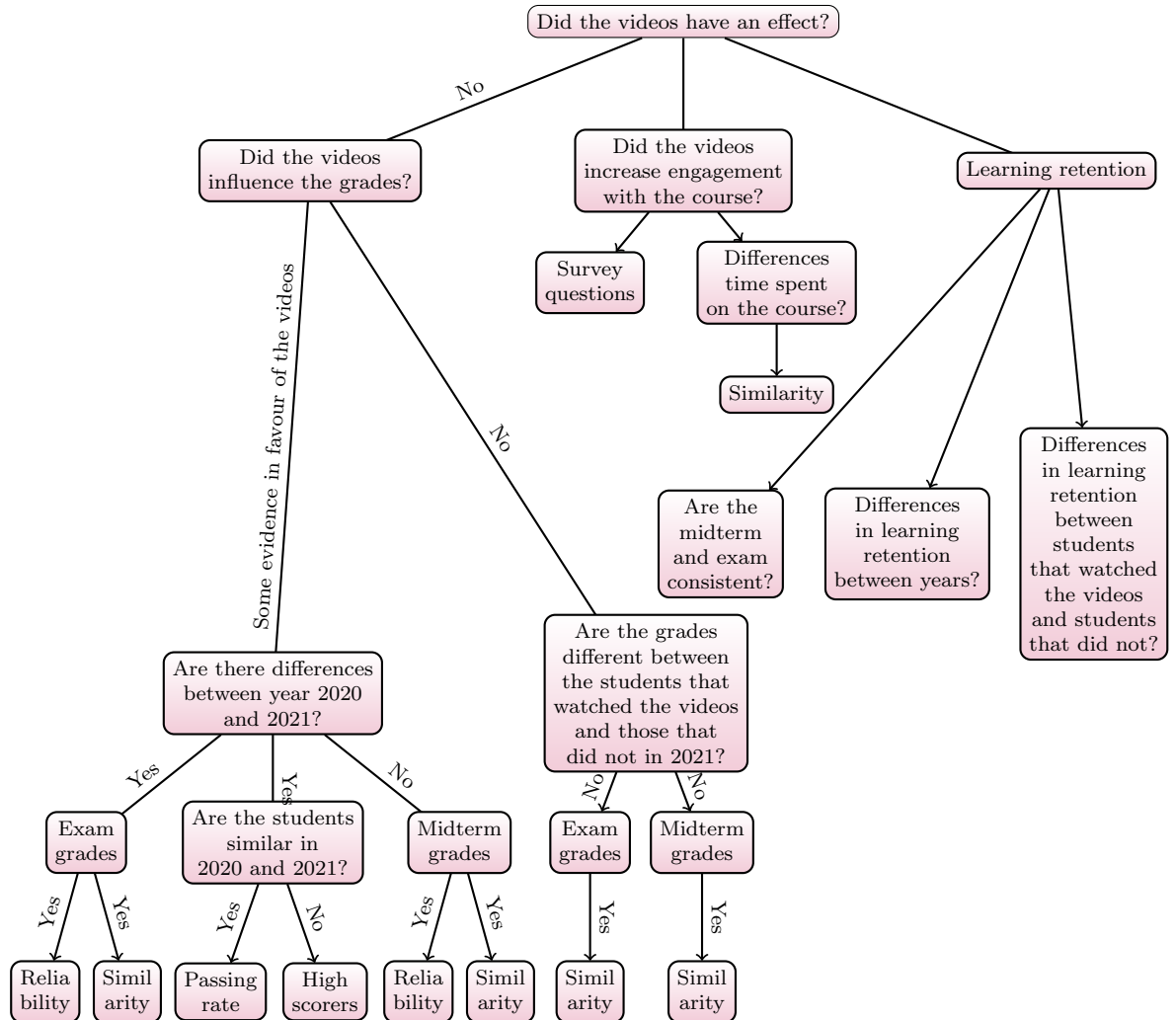
Subject	P-value	Significant?
Renewal theory	0.9609	No
Branching theory	0.8067	No
Brownian theory	0.9902	No
Probability theory	0.9314	No
Final exam grade	0.8737	No

Table 26: The p-values for comparing the subjects on the exam between the students that watched the videos and those that did not.

18.7 Conclusion

The students performed almost the same on the exam, so it does not seem that the students differ in skill. We also did not find evidence that the group that watched the videos performs significantly better on the renewal part of the exam.

19 Recap of the decision tree



20 Did the videos increase the motivation of the students?

We will look into different questions that were asked on the survey to answer this question.

20.1 Did the students report more motivation due to the videos?

The question in the survey was: "Do you feel that the videos made renewal processes more interesting for you? Or in lack of interest, did you appreciate the topic more?". Students had to answer "yes" or "somewhat" if the feeling of either question was positive, otherwise they could answer "no". From the 47 students that filled in parts of the survey only 31 students answered this question. All the students that watched a video replied to this question. The

Did the videos increase motivation?	Yes	Somewhat	No
Student watched the videos	5	9	5
Student did not watch the videos	0	4	7
total	5	13	12

Table 27: A contingency table showing whether students felt more motivated by the videos, where the students are divided based on whether they watched the videos.

values in table 27, are quite small for the group of students that did not watch the videos. We will use a Pearson chi square test, to see if there is a dependency between motivation and whether a student watched the videos. The null hypothesis is that the motivation is not correlated with whether a student watched the videos. We get a χ^2 -value = 5.515 for the Pearson chi squared test, with two degrees of freedom. This gives a p-value of 0.0634, which is not significant. So we do not reject the null hypothesis. In this case we do see that students that watched the videos answered positively 14 out of 19 time, while students that did not answered positively 4 out of 11 times. So this might also be a lack of power for this test.

It also does not seem like the whole student population thought that the videos increased motivation, as only 5 of the 31 students replied with a yes to this question.

20.2 Did the videos help students with trying to work out new ideas?

The question in the survey was: "Did watching any of the videos result in you trying to work out any of the exercises, questions, or ideas you heard there?", which could be answered with true or false. Nineteen students that did not

watch the videos all answered false for this question, which is consistent with our expectation. From the 19 students that watched the video, eight students answered that the videos inspired them to work out new questions or ideas. This means that the videos didn't help that much with motivating students to learn new ideas that go beyond the course.

20.3 Did the students that watched the videos spend more time on the course?

Independence: No
Non-discriminatory questions removed: No
One to one relation: Yes
Number of Students used in the comparison: 27-19

Table 28: The assumptions for the tests in section are given here.

20.3.1 Renewal theory

The students that did not watch the videos spent on average 30.15 hours on renewal theory, with a median of 30 and a standard deviation of 12.04. The students that watched the videos spent a mean of 37.42 hours, with a median of 40.00 and a standard deviation of 11.90 hours.

From the averages we get that the group that watched the video also spent

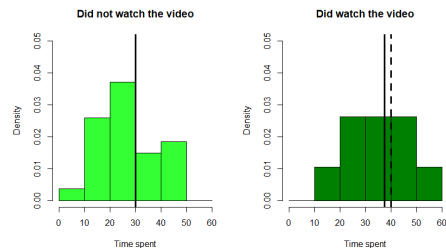


Figure 31: Time spent on the renewal theory part of the course. The group that watched the videos also spent more time.

some more time on this subject. The difference in time spent is 7 hours, while on average students watched 3 videos, which is 1.5 hours of videos, so this does indicate that they spent more time in general on renewal processes. We will compare the time spent between the two groups with a two-sided Mann-Whitney U test. This test reports a p-value of 0.05329, which is not significant. So we do not reject the hypothesis that these groups spent the same amount of time on renewal theory.

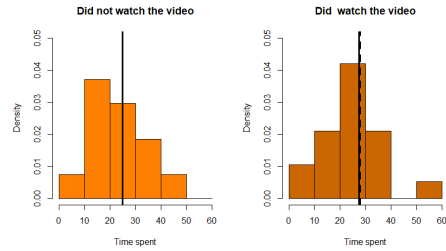


Figure 32: Time spent on the branching theory part of the course. For this subject there is not much difference between the two groups

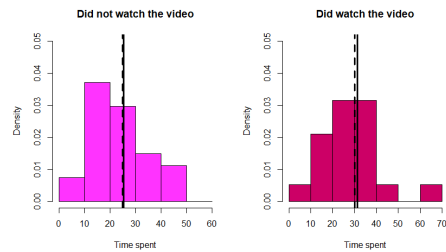


Figure 33: Time spent on the Brownian theory part of the course. For this subject there is not much difference between the two groups

20.3.2 Branching theory

The students that did not watch the videos spent on average 24.89 hours on branching theory, with a median of 25.00 hours and with a standard deviation of 11.42 hours. The group that did watch the videos spent 27.58 hours with a median of 28.00 hours and a standard deviation of 11.49 hours. One student spent 60 hours on this subject, which is quite remarkable. The time spent on this subject is also the least out of the three subjects, while this subject had the highest grades of the three.

The two-sided Mann-Whitney U test reports a p-value of 0.4191, which is not significant. So we do not reject the null hypothesis. This is also what we expect from the averages.

20.3.3 Brownian theory

The students that did not watch the videos spent on average 25.48 hours, with a median of 25.00 hours and a standard deviation of 11.47 hours. The group that did watch the videos spent 31.26 hours, with a median of 30.00 hours and a standard deviation of 13.54 hours.

The difference in study time between the two groups is 6 hours, which is large considering that the group that did not watch the videos only spent 25 hours

on this subject. The Mann-Whitney U test reports a p-value of 0.1639, which is not significant. So we do not reject the null hypothesis.

20.4 Conclusion

We did not find a significant difference in time spent between the students that watched the videos and those that did not. However, the group that watched the videos spent some more time on the course in general, so this indicates that the group is more motivated. The standard deviations of the time spent on each subject is quite large, when coupled with the low sample size, means that the power of the test is low.

21 What was the most interesting video?

There were 6 videos and it is interesting to see which type of video is the most popular, so that in the future this type of content is made. In total 12 students reported a favourite video. Video 6, Maria’s restaurant, was the most popular with 10 votes. In this video Maria explained an example of renewal processes, so the video is about applying the theory. Video 1, about stationary renewal processes, got 7 votes. This video introduced some new theory. Video 3, Little’s law, was the least liked video with only one vote. There were only 4 students that watched this video. During a bachelor course about queuing systems this subject is already introduced, so that might explain why students did not like the video.

22 Did the videos enrich the course?

It is also interesting to see whether the students thought that the videos enriched the course, some students see the added value of the videos for other students, but do not watch the videos themselves.

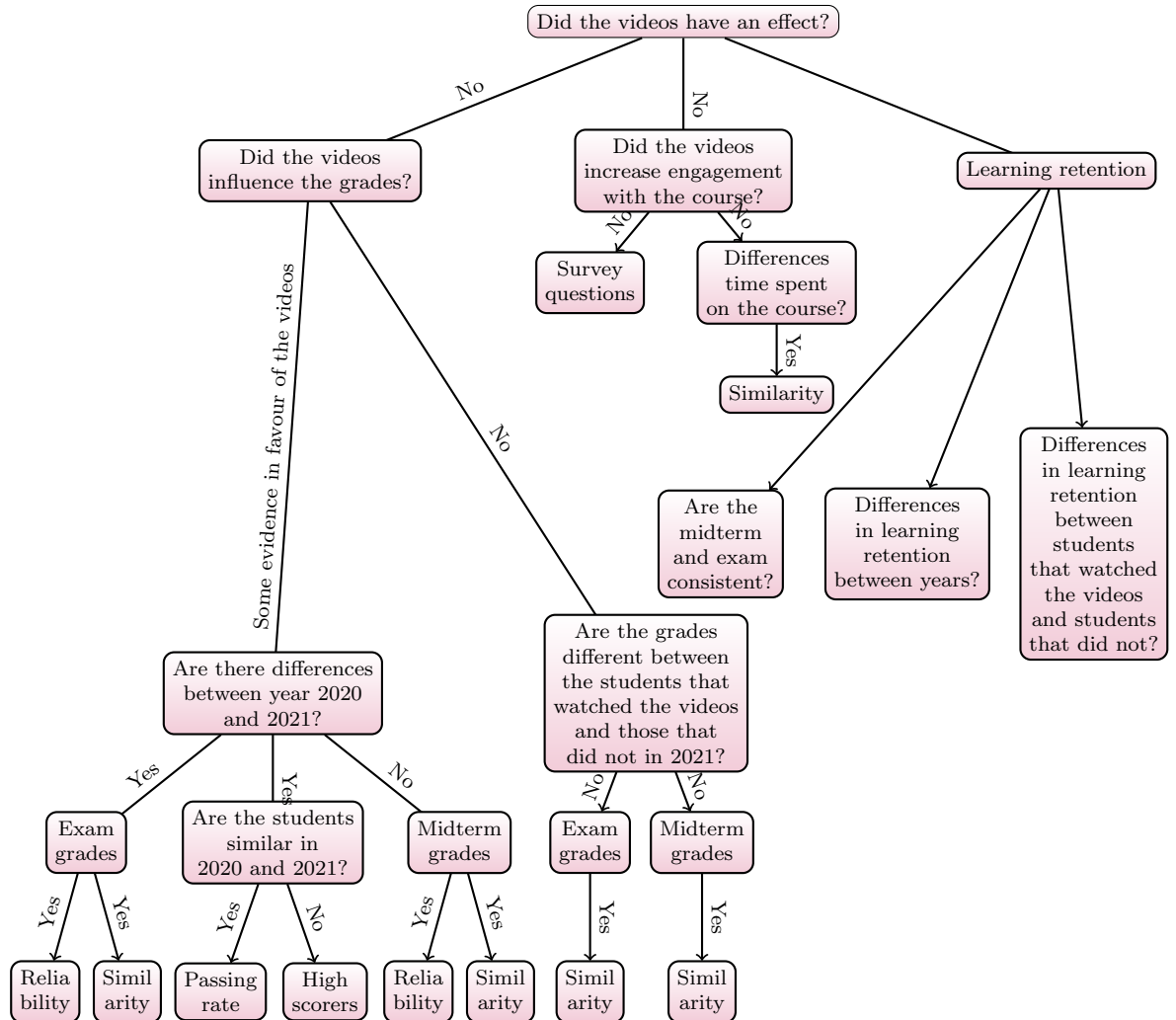
To study this the following question was asked in the survey: ”Do you feel that the videos made the course any more interesting? Or do you feel that the videos were of any added value to the course?”. The answers possible were: ”yes”, ”somewhat” or ”no”. The same group as in section 20.1 plus one student who answered no, but did not answer the question about motivation, answered this question. This gave rise to the following table: As you can see in table

Did the videos enrich the course?	Yes	Somewhat	No
Student watched the videos	9	6	4
Student did not watch the videos	0	6	6
Total	9	12	10

Table 29: Did the students think that the videos enriched the course. The students were split based on whether they watched the video.

29, students that watched the videos also thought that they enriched the course more. The total amount of answers is also quite uniform over each of the 3 categories, so this does not give a clear indication whether students thought that the videos enriched the course in general. The students that watched the videos did find them motivating, as 15 out of 19 students answered positively on this question.

23 Recap of the decision tree



24 Consistency of the midterms with the exam

It is also interesting to see if the grades of the midterms are consistent with the grades of the exam, because then the midterms help with preparing for the exam as well as being a good way to assess students.

We can still compare the midterm and exam by looking at the group of students that scored above average on the midterm and see if they also scored above average on the exam, as we expect that students that perform well on the midterm also perform well on the exam.

24.1 Test procedure

First we split the students into two groups, one group that scored equal or above the median on the renewal midterm and the other group that scored below the median grade. We do the same for the exam, which gives us 4 groups in total. We can then use the Pearson chi square test to relate these two variables. We want a 1 to 1 relation between the midterm and the exam for this analysis, so we will use the highest scoring student of a group in this analysis.

24.2 Is the renewal midterm consistent with the exam in 2020

Independence: No
Non-discriminatory questions removed: No
One to one relation: Yes
Number of students used in the comparison: 2020: 37 and in 2021: 35

Table 30: The assumptions for the tests in section 24 are given here.

	Exam 2020 \geq median	Exam 2020 $<$ median
Renewal midterm 2020 \geq median	12	9
Renewal midterm 2020 $<$ median	9	8

Table 31: The amount of students that scored higher on the renewal midterm of 2020 are divided into the groups of whether they scored higher on the exam as well.

For contingency table 31 we get a p-value of 0.6658 with the Pearson- χ^2 test, so we do not reject the null-hypothesis.

	Contingency table		p-value	Significant?
Renewal 2020	$\frac{12}{9}$	$\frac{9}{8}$	0.6658	No
Branching 2020	$\frac{14}{6}$	$\frac{5}{12}$	0.01383	Yes
Brownian 2020	$\frac{13}{7}$	$\frac{6}{11}$	0.0716	No
Renewal 2021	$\frac{12}{6}$	$\frac{7}{10}$	0.1303	No
Branching 2021	$\frac{12}{6}$	$\frac{7}{10}$	0.1303	No
Brownian 2021	$\frac{13}{5}$	$\frac{7}{10}$	0.0636	No

Table 32: For each contingency table the top row is whether the student scored above or equal to the median of the midterm. The bottom row is for the students that scored below the medium. The first column are the students that scored above or equal to the median grade on the exam and the second column in each contingency table are the students that scored below the median on the exam. The p-values for the Pearson chi square test per table are also given.

24.3 Summary of the consistency of the other midterms with the exam in 2020

We will not write out the full analysis. You can find the results in table 32. In table 32, we can see that, except for the branching midterm of 2020, scoring higher on the midterms does not necessarily imply that you score higher on the exam as well.

24.4 Is the midterm consistent with the corresponding subject on the exam?

In the previous section, we looked at the whole exam, but we can also look at the subjects on the exam as these are about the same theory. This gives rise to table 33.

As we can see in table 33, it does not seem like scoring higher on the midterms is correlated with scoring higher on the specific subjects on the exam.

24.5 Conclusion about consistency

We did not find enough evidence to conclude whether scoring above average on the midterm is correlated with scoring above average on the exam.

	Contingency table		p-value	Significant?
Renewal 2020	$\frac{13}{6}$	$\frac{8}{10}$	0.1412	No
Branching 2020	$\frac{13}{12}$	$\frac{6}{6}$	0.9093	No
Brownian 2020	$\frac{14}{4}$	$\frac{6}{11}$	0.03282	Yes
Renewal 2021	$\frac{15}{8}$	$\frac{4}{8}$	0.07228	No
Branching 2021	$\frac{11}{7}$	$\frac{8}{9}$	0.4042	No
Brownian 2021	$\frac{14}{4}$	$\frac{6}{11}$	0.0111	Yes

Table 33: For each contingency table the top row is whether the student scored above or equal to the median on the midterm. The bottom row is for the students that scored below the median. The first column are the students that scored above or equal to the median on the corresponding part on the exam and the second column in each contingency table are the students that scored below the median on the corresponding part of the exam. The p-values for the Pearson chi square test per table are also given.

25 Learning retention

We are also interested to see if the videos influenced the learning retention of students. Were students able to better recall the things that they learned through the videos during the exam for example?

25.1 Set up of the experiment

We cannot compare the midterm and exam grades directly, as we already mentioned in section 24. We expect however, that students obtain a similar rank on the midterm compared to the exam. We can calculate the correlation between the two rank sets of the midterm and exam with a Kendall tau test. We do have mixed ranks for the midterms because students worked in groups. Thus, we will use the students that scored best in their group to get a 1 to 1 relationship between the midterm grades and the exam. In our analysis, we will also include the students that participated in both years, because we will only compare ranks within a year. We will compare both the ranks of each midterm with the rank of the exam and with the rank on that specific subject on the exam.

We can also look at other studies regarding learning retention, one study showed that preschoolers forgot around 50% of the words that they had to memorise in a time span of 2 days [4]. We only have grades for the midterm and the exam which have a time span of 6 weeks between them, so the students will proba-

bly forget more of the things they learned during the midterm. Students also studied for the exam, so after a bad midterm they might have studied harder and vice versa. With this in mind, we can still look at the results, drawing conclusions however will be harder.

25.2 Test procedure

First we removed the students that scored the lower exam grade of their group, because then we have a 1 to 1 relation between the midterm and the exam. In most cases the group member that scored the highest grade scored a similar exam grade compared with the students that worked alone. We did not test this thoroughly though. After that we assigned ranks based on the midterm and the exam and used a Kendall tau correlation test on the ranks.

<p style="text-align: center;">Independence: Yes Non-discriminatory questions removed: Yes One to one relation: Yes Number of Students used in the comparison: 30-28</p>

Table 34: The assumptions for the tests in section 25.3, 25.4 and 25.5 are given here.

25.3 Renewal midterm

25.3.1 Year 2020

In 2020 the Kendall tau correlation between the ranks on the renewal midterm and the whole exam is 0.156, with a p-value of 0.231. We do not reject the null hypothesis, so there is not enough evidence to say that the correlation between the ranks is nonzero.

The Kendall tau correlation coefficient for the renewal midterm and the specific part on the exam in 2020 is 0.162, with a p-value of 0.233. So we do not reject the null hypothesis.

It is quite weird that the correlations are similar, as we expect the renewal midterm and renewal part of the exam to have a larger correlation, which is not the case.

25.3.2 Year 2021

In 2021, the correlation is 0.238 for the ranks on the renewal midterm and the exam, with a p-value of 0.0835. So we do not reject the null hypothesis.

In a similar vein we have that the correlation between the ranks of the renewal midterm and the renewal part of the exam is 0.375, with a p-value of 0.007408. So in this case we do reject the null hypothesis.

Subject:	Kendall tau	p-value	Significant?
Whole exam in 2020	0.252	0.061	No
Branching subject in 2020	0.269	0.060	Yes
Whole exam in 2021	0.238	0.0835	No
Branching subject in 2021	0.154	0.265	No

Table 35: the Kendall tau correlation coefficient between the branching midterm and the subject described in the table.

Subject:	Kendall tau	p-value	Significant?
Whole exam in 2020	0.240	0.0832	Yes
Brownian subject in 2020	0.264	0.0648	Yes
Whole exam in 2021	0.323	0.0174	Yes
Brownian subject in 2021	0.374	0.0065	Yes

Table 36: the Kendall tau correlation coefficient between the Brownian midterm and the subject described in the table.

25.3.3 Conclusion on the renewal midterm

The only correlation that was significantly different than 0 was between the renewal midterm and the renewal part of the exam in 2021. However the correlations are quite small for all 4 analysis, as they are below 0.4.

25.4 Branching processes midterms

In table 35, we can see that there is once again a correlation of around 0.25 between the ranks on the Branching midterm and the exam. Oddly enough the correlation between the specific subject and the midterm is lower than the exam as a whole. We already know that the branching midterm had very high grades with a low standard deviation. This might mean that the rankings on the midterm is not close to the real ranking of the students.

25.5 Brownian midterm

We once again see that there is a correlation between the Brownian midterm and the exam in table 36. In this case the correlation between the midterm and the exam are significantly different from 0.

25.6 Conclusion

We have observed correlations between 0.15 and 0.40 for the ranks of the midterm and exam, which means that the dependency between the ranks on the midterm and exam is not very high.

25.7 Did the group that watched the video retain their midterm rank on the exam?

There might be a difference in learning retention between the students that watched the videos and those that did not. We will compare the group that watched the videos with the group that did not watch the videos. We do need to take the group structure of the midterm in account, so we will only look at the high scoring students in a group, so that we get a one to one relation between the midterm and the exam. From these 31 students we still need to subtract six students that did not watch the videos. This means that we have 11 students left in the group that watched videos and 14 students that did not watch the videos. This means that the power of the Kendall tau test will be small, so we have to interpret the results with caution.

Independence: Yes
Non-discriminatory questions removed: No
One to one relation: Yes
Number of Students used in the comparison: 14-11

Table 37: The assumptions for the tests in section 25.8, 25.9 and 25.10 are given here.

25.8 Renewal subject

25.8.1 Students that did not watch the videos

The Kendall tau correlation coefficient between the rank on the renewal midterm and the whole exam of the group that did not watch the videos is 0.2007, with a p-value of 0.265. So we do not reject the null hypothesis, so the correlation might also be zero.

The Kendall tau of the renewal midterm and the renewal part on the exam is 0.314, with a p-value of 0.132. So we reject the null hypothesis in this case.

25.8.2 Students that watched the videos

The Kendall tau correlation coefficient of the renewal midterm and the whole exam is 0.382 for the group of students that watched the videos, with a p-value of 0.121. So we do not reject the null hypothesis.

The Kendall tau of the renewal midterm and the renewal part on the exam is 0.699, with a p-value of 0.00348. So in this case we reject the null hypothesis.

25.9 Branching subject

In table 38, we can see that the correlations are not significantly different from zero.

Viewed the video&Subject	Kendall tau	p-value	Significant?
Did not watch&whole exam	0.201	0.3223	No
Did not watch&branching subject	-0.0894	0.66	No
Watched the videos&Whole exam	0.315	0.183	No
Watched the videos&branching subject	0.0189	0.9372	No

Table 38: The Kendall tau correlation coefficient of the branching midterm with either the whole exam or just the branching part of the exam.

Viewed the video&Subject	Kendall tau	p-value	Significant?
Did not watch&whole exam	0.258	0.2053	No
Did not watch&branching subject	0.307	0.1355	No
Watched the videos&Whole exam	0.367	0.1183	No
Watched the videos&branching subject	0.315	0.183	No

Table 39: The Kendall tau correlation coefficient of the branching midterm with either the whole exam or just the Brownian part of the exam.

25.10 Brownian subject

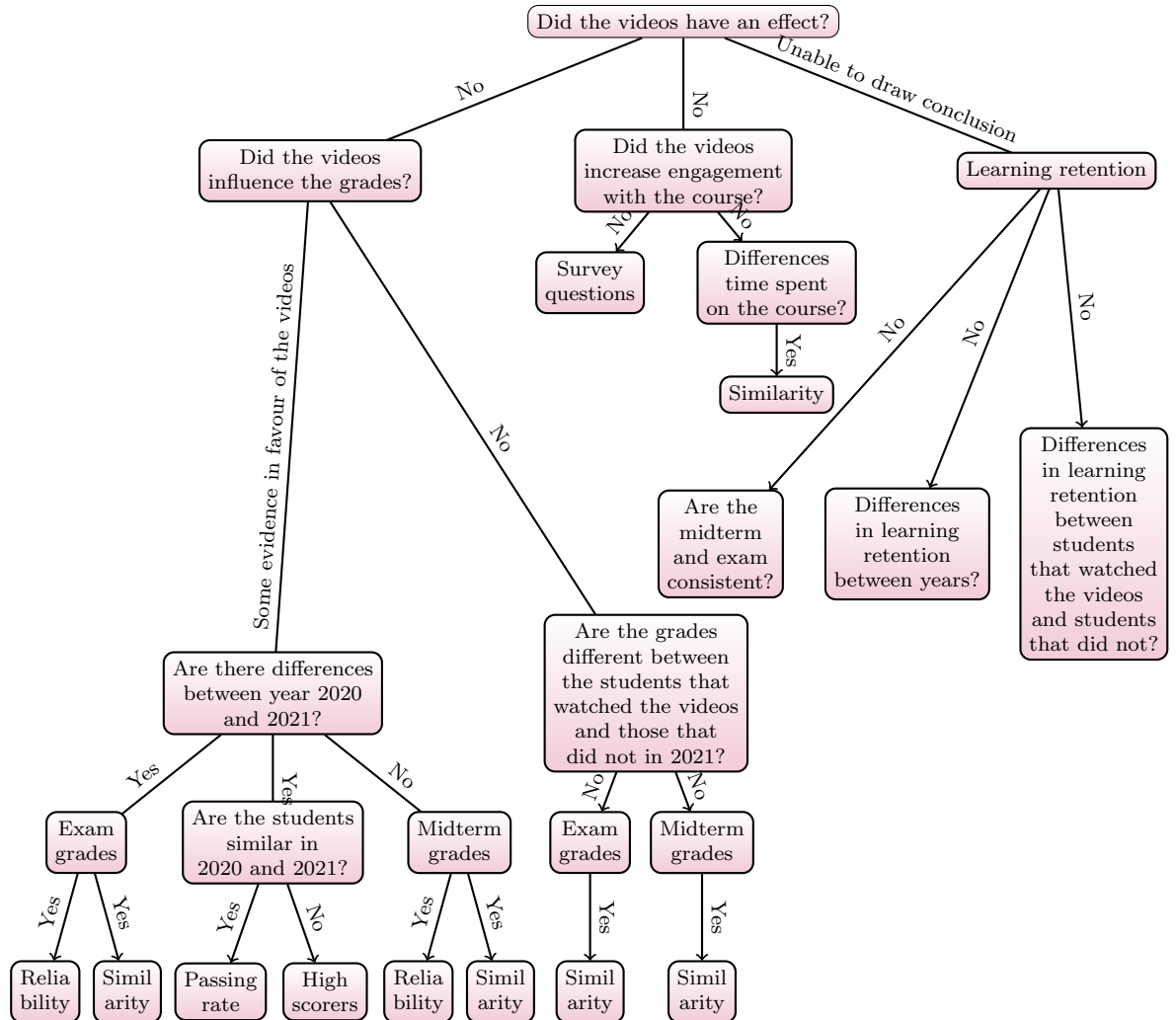
In table 39, we can see that the correlations are around the same strength for both the exam and the specific subject on the exam. Also there does not seem to be a difference between the group that watched the video and those that did not.

25.11 Conclusion about learning retention

When we compared the ranks in years we saw correlations between 0.15 and 0.4, so the ranks on the midterm and exam do not have a high correlation. We did not do a statistical test to see if the correlations of the renewal part of the course were higher, but the values look similar.

When we compared the ranks on the midterm and exam of the group that watched the videos and those that did not, we did find that the students that watched the videos retained their rank on the renewal part of the exam.

26 Recap of the decision tree



27 Conclusion

We found a bit of evidence that the videos improve the grades of the students, but only when we look at differences on the exam between years. The comparison within the year 2021 gave no evidence of the videos influencing the grade. This comparison is stronger because the midterms and exam are the same, however the sample sizes are smaller. So our conclusion is that the videos do not significantly improve the grades of the students. We also did not find strong evidence that the videos engaged the students. With the available data we could not draw a conclusion about learning retention.

28 Discussion

28.1 Do the videos improve the grades of the students?

We do not know when students watched the videos, so comparing the group of students that watched the videos with the students that did not may not be totally fair if some students watched the videos after the renewal midterm, because then they should be included in the group that did not watch the videos for the comparison of the renewal midterm.

We also compared two different years, so the midterm and exam had different questions. We tried to account for this with the reliability analysis, but this is not a perfect measure as some questions are just harder than others. Comparing the years is still useful to see if the course was structured in the same way both years.

During both years the COVID-19 pandemic was ongoing. We assumed that there were no differences between the two years because of the pandemic, however we know that the pandemic influenced study results significantly [3].

We did not perform a power test before comparing the grades. For example a power test for the two sided Mann-Whitney U test with an effect size of 0.5 and the number of students we had, has a power of 0.73, which is lower than the advised 0.8. One possible way to obtain an effect size of 0.5 is if the difference in means between both sets is 1 and the standard deviation of both sets is 2. The grades mostly had smaller differences, so the power of our tests is was low in most cases. This means that we do not reject the null hypothesis, while the alternative hypothesis is true.

28.2 Do the videos affect student engagement with the course?

Students could decide whether they wanted to view the videos or not. There seems to be a selection bias due to this. Students that already spent more time on the course in general also watched the videos. So these students were already more motivated for the course. Oddly enough we did not see this reflected in the grades of these students.

The videos were also quite lengthy with 30 minutes, while this paper recommends 6 minutes per video for optimal engagement, with a steep drop off after 6 minutes [1]. This means that the videos were less impactful than they could have been for student engagement.

28.3 Do the videos affect learning retention

We did not have the right data to draw a conclusion about learning retention. We only have grades for the midterm and exam, between which there is a period of 5 to 6 weeks. In another paper they found that children aged 10, lose 50% of the words they had to memorise in 2 days [4]. Students are of course older and memorising words instead of understanding mathematics is different, but still the effect of the videos and the midterm on the exam might be small due to the time between the two.

29 Further recommendations

Some students do like videos as a means of studying, so making shorter videos to explain new concepts might be a good idea as these videos are more engaging. If the goal is to study the effect of videos on knowledge retention then try to keep the measurements within 2 days to a week at the most, otherwise there will be too much noise in the data.

If we use grades to look at the effect of an intervention. Then first perform a power test before the analysis, as the amount of students for the course might be too low.

References

- [1] Philip J Guo and Rob Rubin. “How Video Production Affects Student Engagement : An Empirical Study of MOOC Videos”. In: March (2014). DOI: 10.1145/2556325.2566239.
- [2] A. Hart. *Mann-Whitney test is not just a test of medians: Differences in spread can be important*. Aug. 2001. DOI: 10.1136/bmj.323.7309.391.
- [3] Megan Kuhfeld and Karyn Lewis. “COVID-19-impacted School Years”. In: 22 (2022).
- [4] Thomas J Shuell and Geoffrey Keppel. “Learning ability and retention”. English. In: *Journal of Educational Psychology* 61.1 (Feb. 1970), pp. 59–65. ISSN: 0022-0663, 0022-0663. DOI: <https://doi.org/10.1037/h0028756>. URL: <https://www.proquest.com/scholarly-journals/learning-ability-retention/docview/614271607/se-2?accountid=27128%20http://tue.on.worldcat.org/atoztitles/link?sid=ProQ:&issn=00220663&volume=61&issue=1&title=Journal+of+Educational+Psychology&spage=59&date=1970-0>.

A Scores per quintile for the discriminatory questions on the exams

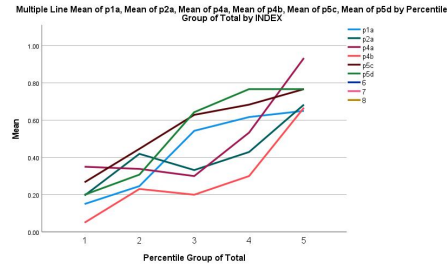


Figure 34: A part of the good questions of the exam in 2020.

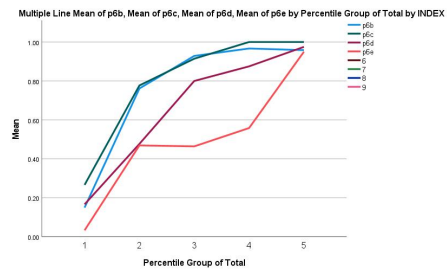


Figure 35: A part of the good questions of the exam in 2020.

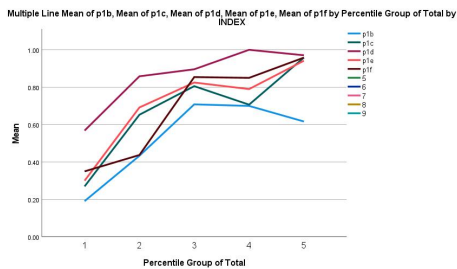


Figure 36: A part of the good questions of the exam in 2021.

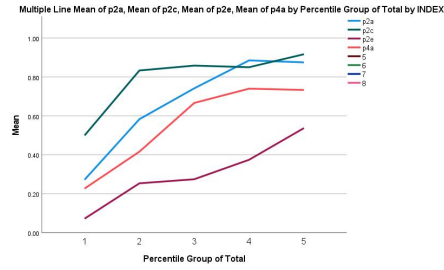


Figure 37: A part of the good questions of the exam in 2021.

A.1 Histograms for the exams with questions that had a low or high average dropped

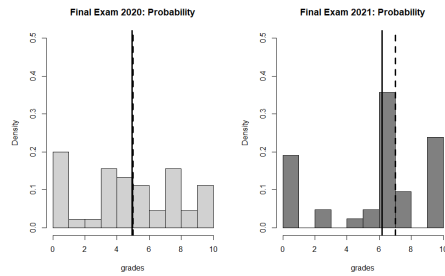


Figure 38: A histogram showing the grades of the probability part of the exam in both years, while some questions were removed from the analysis. The black line is the mean grade and the dotted line is the median grade.