Report on Introduction to Cryptography

Aleandro Eccel, Ravinithesh Annapureddy, Maxime Jan

Report as part of the CS-411 Digital Education Project

January 19, 2021

Contents

1 Introduction

In the context of the project for the course CS-411 Digital Education, we designed and empirically tested a short lesson. For our project, we chose to teach an introduction to Cryptography, revolving around the Caesar Cipher. More specifically, this introduction course contains a first part motivating the need for cryptography. It is then followed by the theory and practice of the Caesar Cipher. Finally, the last part introduces the students to the concept of security of a cryptographic algorithm related to a brute-force attack.

While this lesson is typically given to students of the Digital Education's course itself, we had the very special opportunity to test our lesson in a real class setting at the *Collège du Sud* (CSUD) in Bulle. Thus, our target audience consists of first-year high school students, between the ages of fifteen and seventeen, from the Fribourg's districts of Gruyère and Veveyse. While these students come from various *Cycles d'Orientation* (CO) in these districts, we assume all of them have a similar skill level, as they were assigned randomly to one of the 11 classes at the beginning of the semester. The students in these classes study Computer Science for two periods a week starting from this September and have not been introduced to cryptography yet. As part of their curriculum, they have already learned about the basics of data representation and have also started learning to program using Dr. Kohn's TigerJython from ETH. We conducted the experiment in two separate classes. We ensured that the students did not have any prior knowledge about Caesar Cipher by conducting a small survey beforehand. Concerning the developmental and social skills of these classes, the students had been together for 3 months by the time of the experiment while some of them already knew each other from the CO. As one of the experimenters is a teacher at the school, it has also empirically observed that the atmosphere between the students has been pleasant, and everyone seems to get along and communicate well. Therefore we assume that working in groups is not a problem for them.

Despite Covid-19 regulations, the lesson was given offline in one of the CSUD's computer rooms, as schools remain open in the Canton of Fribourg. In this setting, all of the students are seated in front of a Desktop. The lesson lasted 45min, the standard period time at the CSUD.

2 Learning Goals

For our *Introduction to Cryptography* lesson, we defined both procedural and conceptual learning goals.

The first two conceptual learning goals we identified concern the very basics of cryptography and are rooted in the *Understand* level of Bloom's taxonomy. Firstly, we want students to be able to explain the relation between a plain text and its encrypted version via a cryptographic algorithm and a secret key. Secondly, students should also be able to describe what the benefits of using cryptography and sharing ciphertexts instead of plain texts are.

As the cryptographic algorithm studied in our lesson is the Caesar Cipher, we have identified different procedural learning goals so that students would be able to use it. Firstly, students should be able to encode plain texts using the Caesar Cipher and a given key. Conversely, they should also be able to decode such a ciphertext with a given secret key. Furthermore, they should be able to find what the secret key is, given plain text and its encoded version. In Bloom's taxonomy, these learning goals correspond to the *Apply* level.

Our last learning goals are focused on the weaknesses of the Caesar Cipher. The first one is conceptual: students should be able to explain why this cryptographic algorithm is not safe (namely because there are very few secret keys). Then, students should be able to apply a brute-force attack to break an encoded message for which the key is missing. While this latter procedural learning goal is at the *Apply* level of Bloom's taxonomy, the former one is grounded at the *Understand* level.

3 Task Analysis

In this section, we enumerate the learning goals outlined above and describe the task breakdown for each of them.

- Relation between a plain text and its encrypted version via a cryptographic algorithm and a secret key.
  - The plain text is the readable text that you want to transmit.
  - The ciphertext is the encoded version of the plain text.
  - You can transform a plain text into ciphertext by inputting it into a cryptographic algorithm, along with a secret key of your choice.
  - You can transform a ciphertext into plain text by inputting it into a variant of the same cryptographic algorithm, along with the same key used for encryption.
  - For decryption, if you use a different secret key or a different algorithm than the one used for the encryption, you will not get the decoded plain text, but some nonsensical text instead.
- Security benefits of sharing ciphertexts instead of plain texts
  - It is possible for malicious users to intercept your online communications. For example, these users can read the "wifi signals" you are sending.
  - If you share a plain text, anyone listening to your communications will immediately be able to understand your messages and will compromise your privacy.
  - If instead, you agree with your receiver on a cryptographic algorithm and a secret key, you can encrypt your messages and send the ciphertexts. A malicious user intercepting this encoded message will not be able to decrypt it and will not know what to get from the nonsensical text he will be reading.
- Encoding with the Caesar Cipher
  - For each letter of the plain text, apply the following transformation
    1. Starting from the current letter you are looking at, go forward in the alphabet of several steps that is specified by the secret key.
    2. If you arrive at the end of the alphabet, go back to the beginning (i.e. if you have to go forward when you arrive at "Z", continue with "A").
    3. Replace the current letter with the one you have landed on.
- Decoding with the Caesar cipher
  - You can apply almost the same technique as for encoding, except that you have to go backward in the alphabet instead of forward.
  - Conversely, if you have to go one step back when you are at the letter "A", you can go at the end of the alphabet and continue with the letter "Z".
- Limits and weaknesses of the Caesar Cipher
  - When a malicious user knows that you are encoding your messages with Caesar cipher, all he has to do is to try all the 26 possible keys to find the correct one.
  - Trying all possibilities for a secret key is called a brute-force attack.
  - A cryptographic algorithm with a secret key is only as secure as the number of secret keys one can choose.
  - The Caesar Cipher should not be used nowadays and only worked in the past because almost no one knew how it worked.

- Performing a brute-force attack
  - Given a ciphertext without a key, try to decode it by using all possible keys.
  - To not waste too much time doing it, you can, for each possible key, decrypt the first few letters and only continue if it seems plausible that they form the beginning of a French word.

4 Lesson Design and Activities

The entire lesson is broken down into the following activities. Please note that following CSUD's lecture time, we designed the lesson to last for 45 minutes and that some activities have two different timings due to the way we constructed our research question.

1. **Caesar wheel exercise (5 min or 0 min)**

   As an introduction to the lesson, we designed a discovery exercise *à la* Piaget to be done individually. In this exercise, students get a Caesar wheel as well as a word. This word has already been partially encoded/decoded, and the students' task is to fill in the blanks using the wheel. This discovery exercise intends to give an alternate representation of the Caesar Cipher algorithm. While the explanations later in the lesson focus on *counting forward or backward in the alphabet*, the wheel gives a graphical representation of a mapping between two alphabets.

2. **Presentation: Why do we need cryptography? (4 min)**

   This activity is a class oral presentation with the support of slides. It is designed to give the students motivations as to why cryptography is a crucial subject and how it affects their daily life.

   It starts with an example of two people exchanging *Whatsapp* messages in plain text. Making a connection with the "data representation" topic they studied earlier in the year, we explain to the students how the signals sent by their cell phones are nothing more than a sequence of 1's and 0's encoded with a table such as the ASCII table. Therefore, they must understand that anyone "in the room" can listen to these signals and can read their conversations.

   We then introduce cryptography as a solution to this problem and explain that two elements are needed: a cryptographic algorithm and a secret key that you have to share in advance with the person you want to talk with. Then, you can encode your plain texts into ciphertexts using these tools. We show examples of plain texts and ciphertexts so that the students understand that ciphertexts are nonsensical texts of which you cannot understand the meaning. We then show how the receiver of your encoded messages can decrypt them using the pre-shared algorithm and secret key.

   At this point, we reiterate the same situation as earlier, except with encoded messages. This time we show that the malicious user listening to the signals of their cell phones can only read the ciphertexts and thus cannot understand their conversation.

3. **Presentation: A cryptographic algorithm, the Caesar cipher (6 min)**

   Diving a little bit into the past, this class presentation explains the fundamentals of the Caesar Cipher. Translating the cell phone situation to one that could have occurred in Ancient Rome, we show why Caesar did not want to send his messengers carrying plain texts anymore and how he came up with a solution.

   While students had a brief exploration of this algorithm in the discovery exercise, we applied Piaget's semi-constructivism and now gave them the required theoretical grounds to perform full encryption and decryption by themselves. We thus explain each step of the algorithm, as shown in the Task Analysis, in detail.

4. **Exercise: Encoding and decoding with the Caesar cipher (8 min or 10 min)**

   After the theory, three types of individual exercises are proposed for knowledge proceduralization.

   - Given a plain text and a key, encrypt it
   - Given a ciphertext and a key, decrypt it
   - Given a plain text and a ciphertext, find the key used for encryption

5. **Presentation: Weaknesses of the Caesar cipher (5 min)**

   After the practice, we go back to Ancient Rome to show how an encrypted message that is stolen from a messenger is still not secured. Indeed, if the thieves know the technique Caesar uses to encode his messages, all they have to do is to try the 26 possible keys to find the correct one. We explain that it is called a brute-force attack and that it is a general attack that can be performed against all cryptographic algorithms that use a secret key. Hence the more possible secret keys there are in the algorithm, the more robust it is against such an attack.

6. **Group exercise: Break the code! (5 min or 8 min)**

   Students can now try to perform a brute-force attack themselves. They are presented with the encoded version of the first sentence of a famous fable from La Fontaine, but without the key to decipher it.

   To not waste too much time trying the 26 possibilities, students are in groups of three. They have access to a chat to coordinate with their group so that they all try different keys. Once they have found the correct one, they can coordinate to decrypt as many words as possible in the allotted time.

7. **Assessing quiz (5 min)**

   As the last exercise, we individually test the procedural skills they have acquired for the Caesar Cipher. For that, we designed a short quiz containing questions similar to the ones asked in activity 4.

8. **Presentation : Debriefing and opening (4 min)**

   This last class activity aims to tie all concepts together and to recap the content of the lecture. We remind them that the Caesar Cipher algorithm is not secure due to the very short number of possible secret keys. However, we tell them that in the upcoming weeks they will learn about other cryptographic algorithms, such as the Vigenère Cipher, that are way more secure.

5 Experimental Design

5.1 Hypotheses

We hypothesize that students who learn by doing a discovery exercise (Cipher wheel) followed by instructions and problem-solving activities will perform both faster and better than students who learn only by instructions and problem-solving activities.

5.2 Hypothesis Testing

The experiment is performed on two classes where one class first performs the Cipher wheel discovery exercises to find the missing letters and then continues with theoretical lectures, problem-solving activities, and a final exercise to encode and decode messages using Caesar Cipher. The second class repeats all the activities performed by the first class except for the Cipher wheel exercises. In the end, we shall compare the time taken by both classes to encode and decode a message in the last exercise along with its correctness to check if our hypotheses stand correct.

5.3 Variables

5.3.1 Independent Variables

The independent variable would be whether a class is undertaking a cipher wheel exercise before attending the lecture. It is widely known that women have been a major part of code breakers during various wars including World War II. Thus, we also measure the effects across the genders by taking it as another independent variable.

5.3.2 Dependent Variables

The time that is taken to encode a message, decode the message, and the correctness of encoding/decoding.

5.3.3 Control Variables

The lecture and the problem-solving activities for both classes remain the same including the total duration of the experiment. To eliminate the effect of prior knowledge on the experiment result, we have conducted a survey 2 weeks before the experiment to assess the knowledge of the participants on cryptography and Caesar Cipher. Based on the results (c.f. Appendix A) we ruled out such possibilities and assumed that none of them had any prior knowledge that would affect the results.

6 Frog Graph Design

To give our lesson while answering our research question, we developed the following two orchestration graphs (Figures 1 and 2) on Frog. The name of the activities are written in French to not confuse the students, but they translate exactly to those described in section 4.

As it can be observed, these are very similar. The only difference is that the second one does not contain the discovery exercise at the beginning and that more time has been allocated for other exercises to compensate.

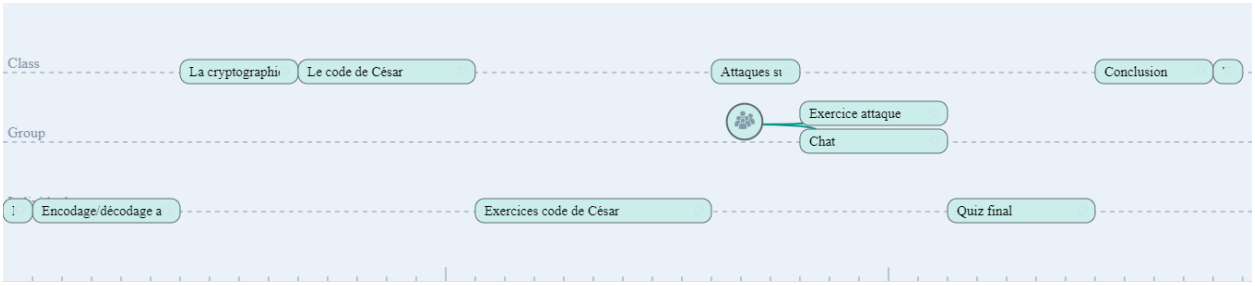

Figure 1: Frog Lesson Design for experiment with a discovery Exercise.

Please note that on the graph, the first very small activity is simply a welcoming message with a text input asking for the name of the student. Similarly, the last small activity is a text wishing the students a good week.

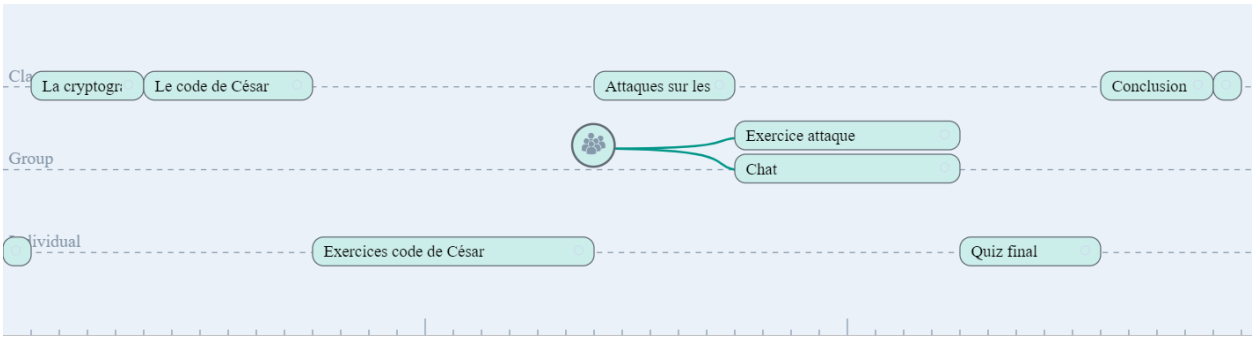

Figure 2: Frog Lesson Design for experiment without a discovery Exercise.

Each of the core activities has been implemented in the following way:

1. **Caesar wheel exercise (5min or 0min)**

   This activity is implemented with multiple-choice questions. For each question, we pasted a Caesar wheel in a different position and wrote partial encoding of a word using it. The students are then presented with 3 different possibilities of complete encoding and have to choose the correct one.

2. **Presentation: Why do we need cryptography? (4min)**

   As this is a class presentation, this activity is only a small text asking to look in the teacher's direction. The slides used for all of the presentations are attached to this report's submission.

3. **Presentation: a cryptographic algorithm, the Caesar cipher (6 min)**

   Idem as for activity 2.

4. **Exercise: Encoding and Decoding with the Caesar cipher (8 min or 10 min)**

   This activity was implemented using a quiz with open textual or numerical answers. For encoding and decoding questions, the prompt gave a plain or ciphertext along with its key, and the students could write the encrypted or decrypted version in the text field. For the type of questions where students had to find the key, the prompt gives both plain and ciphertext and the students can write their answer in a numerical input.

5. **Presentation: Weaknesses of Caesar cipher (5 min)**

   Idem as for activity 2.

6. **Group exercise: Break the code! (5 min or 8 min)**

   As this is a group activity, we use a social operator to form the groups. This operator creates groups of three students. When groups of exactly three students could not be made, we chose to increase the size of some groups so that everyone would get a chance at finding the correct key and at decrypting some of the code.

   This exercise is implemented using two parallel activities. Firstly, a chat is used for communication and coordination within teams, to be the most efficient in the brute-force attack and the decryption. Secondly, a quiz with a single textual question, similar to the ones of activity 4, is used for the collaborative decryption.

7. **Assessing quiz (5min)**

   This quiz is implemented analogously to activity 4, using a quiz with textual and numerical answers.

8. **Presentation : Debriefing and opening (4min)**

   Idem as for activity 2.

7 Participants, Data and Analysis

7.1 Participants

In line with the experimental design, the experiment was conducted with two separate classes. As we present the experimental analysis, we refer to the students as participants.

Discovery Exercise	Gender		Total
	Female	Male	
Yes	14	10	24
No	9	7	16
Total	23	17	40

Table 1: Statistics for Particpants.

A total of 40 participants were part of the experiment. As shown in Table 1, 24 participants completed the discovery exercise while 16 participants directly moved to exercise 2. Of the 40 participants, there were 23 men and 17 women.

7.2 Data

During the process of the experiment, the following data is collected.

1. The number of correct answers in the quiz that was conducted at the end of the class.
2. The time taken to answer the quiz that was conducted at the end of the class.

7.3 Analysis

In this experiment, 6 types of analysis were conducted. The first two are the comparisons of quiz scores and time taken to finish the quiz between the groups that have taken the discovery exercise and did not. The third and fourth are the comparison of the same score and time across genders. The last two are the comparisons of score and time across genders and taking discovery exercises at the same time. As the first four analysis requires us to compare the independent categorical variable having 2 categories with a dependent continuous variable, we shall use a t-test or alternative non-parametric test, in case the assumptions are not met. For the last two comparisons, as we have two independent categorical variables having 2 categories each with a dependent continuous variable, we shall use a two-way ANOVA test.

7.3.1 Effect of discovery exercise on quiz score

The quiz conducted at the end of the class contained 3 questions as mentioned above. Each correct answer acquires one point for the participant and an unattempted or wrong answer would give no points. The participants can be divided into two groups, where one group has taken the discovery exercise before the lecture and the other group did not take it. The histogram of scores is shown in Figure 3. The mean score across all the participants was 2.775 while the majority of them had a score of 3. The mean score in the group that had the discovery exercise is 2.88 (24 participants) whereas it is 2.62 in the other group (16 participants).

As the first step to test if the two groups have significant differences in the scores, we test if the data follows the assumptions of being normal and having equal variance. We start by visual inspection and produce a histogram and Quantile plot. We can see from Figure 4, that the data is not normally distributed as neither the values do not fit a bell-shaped curve nor lie on a straight sloped curve in the quantile plot.

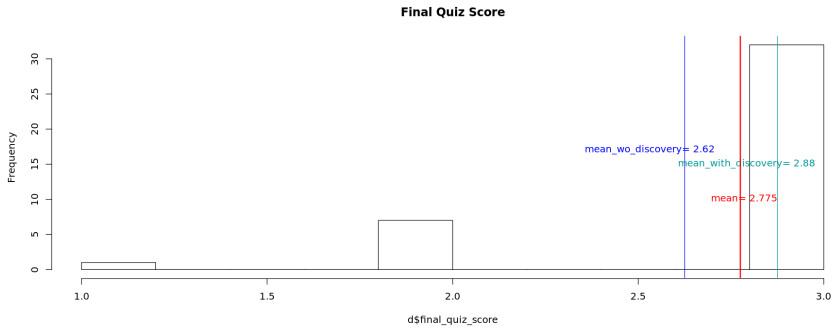

Figure 3: Distribution of Quiz Scores separated across Discovery Exercise

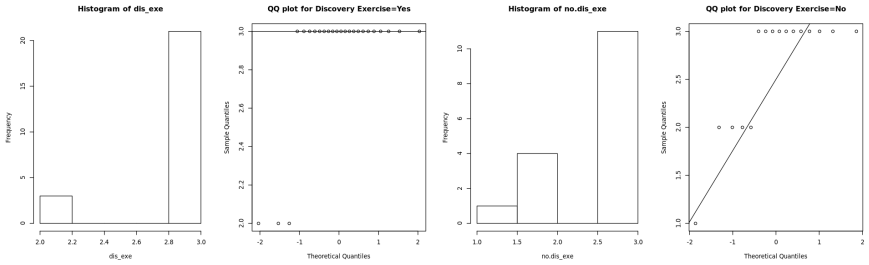

Figure 4: Histogram and Quantile plots of Quiz Scores separated by Discovery Exercise

With the benefit of the doubt, we further conduct the Shapiro-Wilk test. From the result, we see that for both the participants who took and did not take the discovery exercise the p-value is less than .05 and conclude that samples are not normally distributed.

To continue checking on the assumptions, we now test for equality of variance. Looking at the box plot (Figure 5), the score for participants who did not take the discovery exercise has a larger variance than the other group.

Further to confirm, we conduct a Bartlett test. From the result, we see that p-value = $0.01 < 0.05$ and therefore we reject the null hypothesis, and conclude that the variance of the quiz scores is not the same in the groups that took and did not take the discovery exercise.

As the assumptions are not met, we test the assumptions after the transformations such as inverse and square root, and find that the assumptions are not met even after transformation. Thus, we resort to using the non-parametric alternative to the t-test i.e. Wilcoxon-Mann-Whitney test. The Wilcoxon-Mann-Whitney test does not indicate that the quiz score was greater for those who took the discovery exercise (Mdn=3) than those who did not take it(Mdn=3), W=154, p = 0.1.

Thus, in conclusion, we can say that the discovery exercise of caesar disk did not bring statistically significant different scores in the group that had taken it before the lecture.

7.3.2 Effect of discovery exercise on quiz time

Continuing with the two groups that did and did not do a discovery exercise before the lecture, here we compare the time taken by those groups to complete the quiz. From Figure 6, we can see that the average time taken to complete the quiz is 125.8 seconds. Looking across the groups, the average time taken by the

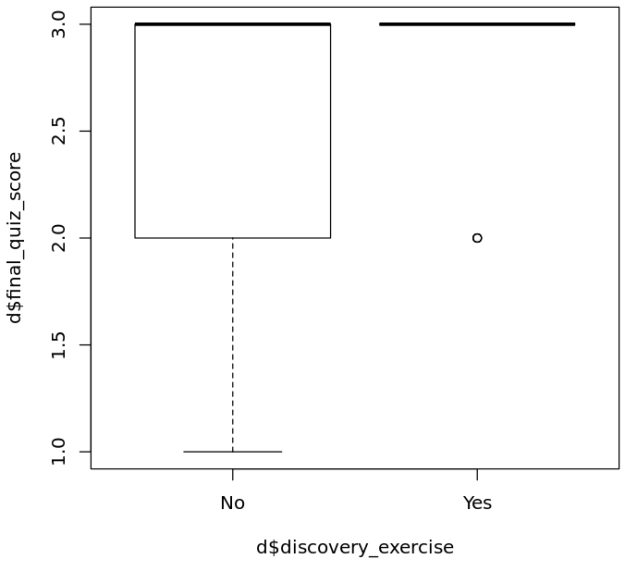

Figure 5: Box plots of Quiz Scores separated by Discovery Exercise

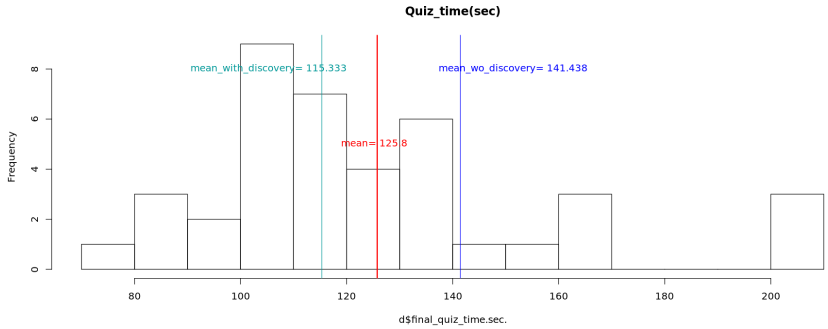

Figure 6: Distribution of Quiz Time separated by Discovery Exercise

group who did a discovery exercise is 115.3 seconds whereas the other group has a mean of 141.438 seconds.

Similar to the previous section, we first check if the assumptions are satisfied or not. From visual inspection of Figure 7, we see that the time for the group that did the discovery exercise is normally distributed and other groups' time is not.

Further, the Shapiro-Wilk test reveals that for time with Discovery exercise is normally distributed ($p > .05$) whereas in the group without discovery exercise, we can reject the Null hypothesis ($p < .05$) and conclude that samples are not normally distributed.

Using the box plot (Figure 8) and the Bartlett test, we see that p-value $= 0.08 > 0.05$ and therefore we cannot reject the null hypothesis, and can conclude that the variance of the quiz time is the same in both the groups.

Although the variances are the same, the normality assumptions are not met even after transformation using a square root or inverse. We once again resort to a non-parametric test. The Wilcoxon-Mann-Whitney test result indicates that the quiz time was smaller for those who took the discovery exercise (Mdn=109) than those who did not take it(Mdn=124.5), W=284, $p = 0.01$.

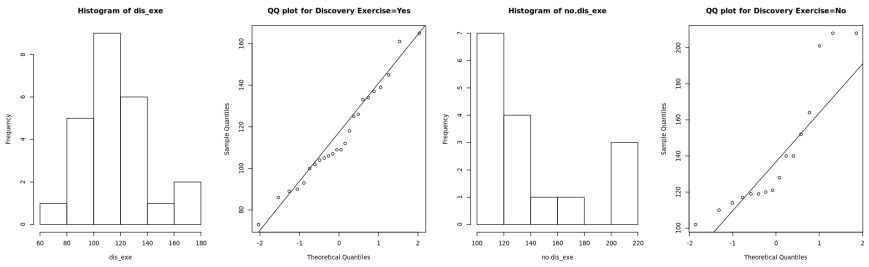

Figure 7: Histogram and Quantile plots of Quiz Time separated by Discovery Exercise

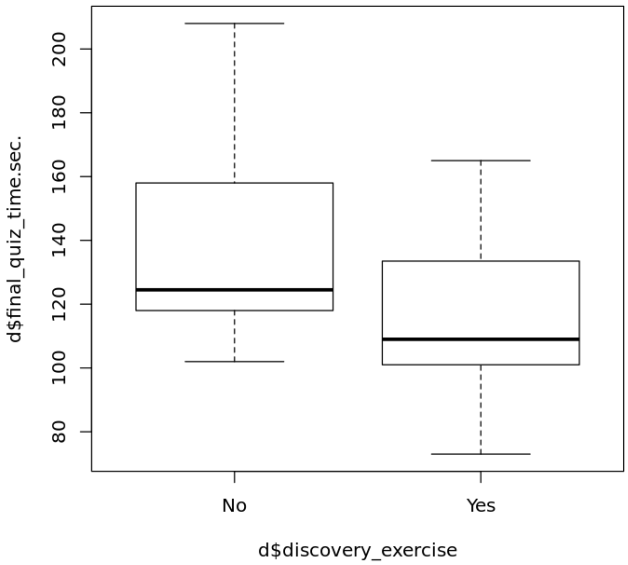

Figure 8: Box plots of Quiz time separated by Discovery Exercise

Hence we conclude that the discovery exercise of caesar disk did have a statistically significant effect on the time taken to answer the quiz and the group that took the discovery exercise took less time compared to the other group.

7.3.3 Effect of gender on quiz score

Similarly to 7.3.1, here we compare the quiz scores across the gender among all the participants. From Figure 9, we can observe that the average score of females is 2.83 (23 participants) and the average score for the males is 2.71 (17 participants).

The histogram, quantile plots (Figure 10) and the Shapiro-Wilk normality test indicate that the samples are not normally distributed ($p < 0.05$).

However, from the box plot of the scores (Figure 11) and the Bartlett test we see that the variance of scores in both the genders is the same ($p > 0.05$).

As the assumption of normality is not met, even after the transformation of the samples, we once again use the Wilcoxon-Mann-Whitney test. The test result does not indicate that the Quiz score was greater for females (Mdn=3) than for males (Mdn=3), W=210, $p = 0.6$.

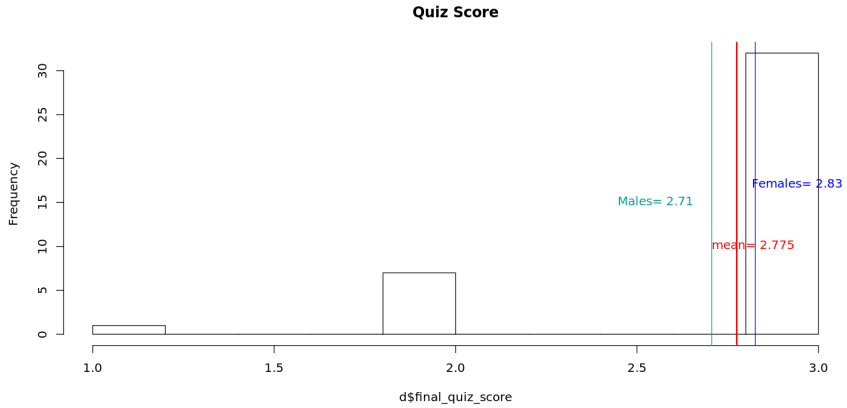

Figure 9: Distribution of Quiz Scores across genders

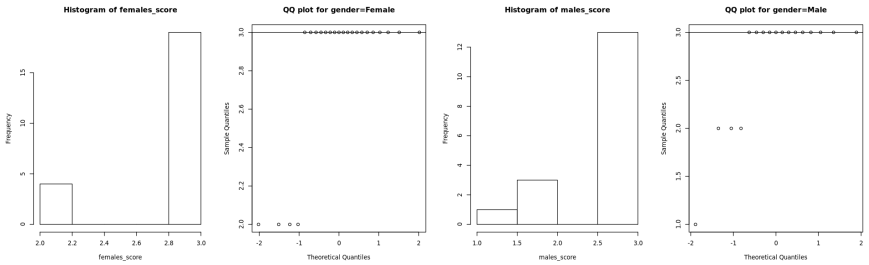

Figure 10: Histogram and Quantile plots of Quiz Scores separated by Genders

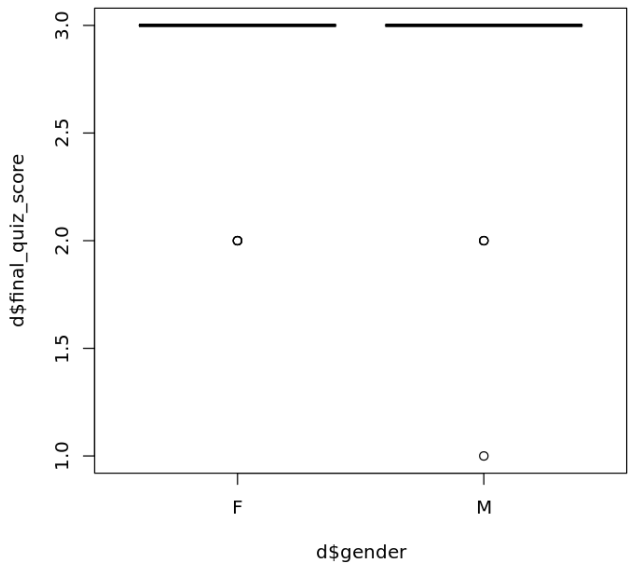

Figure 11: Box plots of Quiz Score separated by Genders

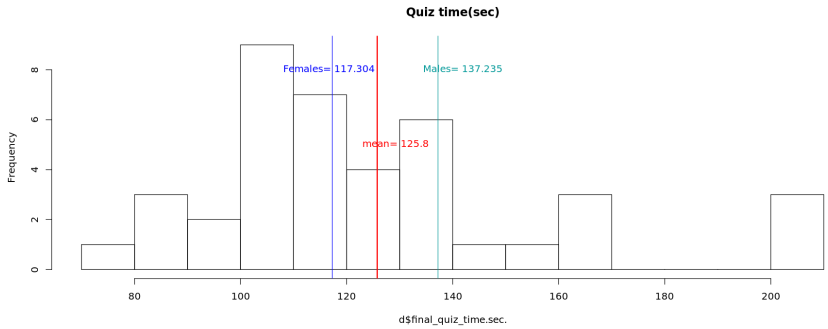

Figure 12: Distribution of Quiz Time separated by Gender

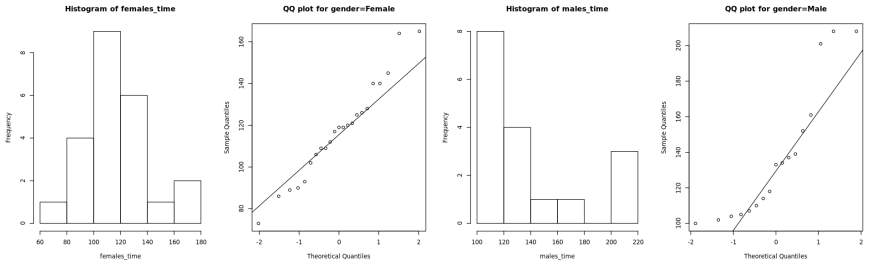

Figure 13: Histogram and Quantile plots of Quiz Time separated by Genders

Thus we conclude that there is no statistically significant difference in scores among the female and male participants in the experiment.

7.3.4 Effect of gender on quiz time

This analysis is similar to the analysis in 7.3.2 except that the quiz time is compared between female and male participants. The average time taken by the female participants is 117.3 seconds and for the males is 137.2 seconds (Figure 12). We test if this difference of 20 seconds is significant or not.

The histogram, quantile plots (Figure 13) and the normality test indicate that the samples are not normally distributed (P value less than 0.05 for males).

From the box plot of the scores (Figure 14) and the Bartlett test, we see that the variance of scores in both the genders is the same (P equals to 0.05).

In an attempt to normalize the samples, we apply the inverse transformation. The Shapiro-Wilk normality test indicates that the inverse of the samples is normally distributed (p-value greater than 0.05). The Bartlett test indicates that the variance of inverted time in both the genders is the same (p-value is greater than 0.05 and we cannot reject the null hypothesis).

As the assumptions of normality and equality of variance are met, we now perform a t-test on the transformed variable. There was no significant difference for inverse quiz time, $t[38] = 1.9$, $p = .06$, despite females ($M = 0.0088$) having higher inverse quiz time than females ($M = 0.0077$).

Thus, in conclusion, we can say that the difference of time (precisely, the frequency of answering the questions) across genders is not significant. A non-parametric Wilcoxon rank-sum test also does not indicate

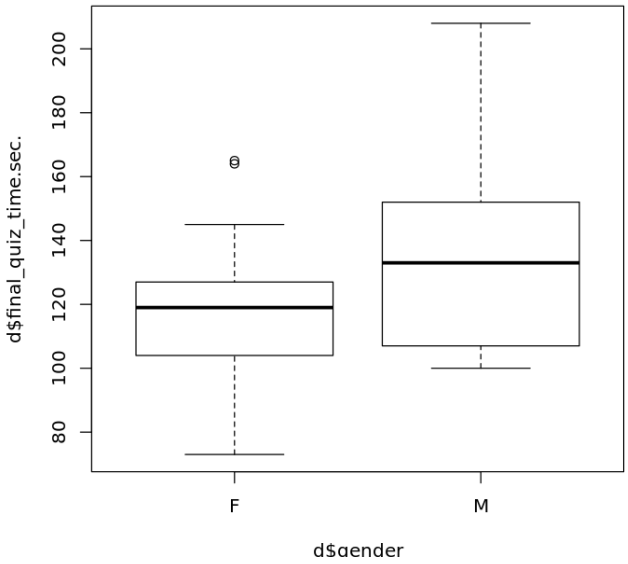

Figure 14: Box plots of Quiz Time separated by Genders

that the quiz time is greater in males (Mdn=133) than the females (Mdn=133), W=148, p = 0.2.

7.3.5 Effect of gender and discovery exercise on quiz score and quiz time

In the previous analysis, we have seen the effect of discovery exercise and gender on quiz time and quiz score separately. However, in this section, we would like to see this effect together. To this end, we try using a two-way ANOVA test on the gender and discovery exercise. Plotting the bootstrapped mean of quiz time and quiz score across gender and separated by discovery exercise is shown in Figures 15 and 16.

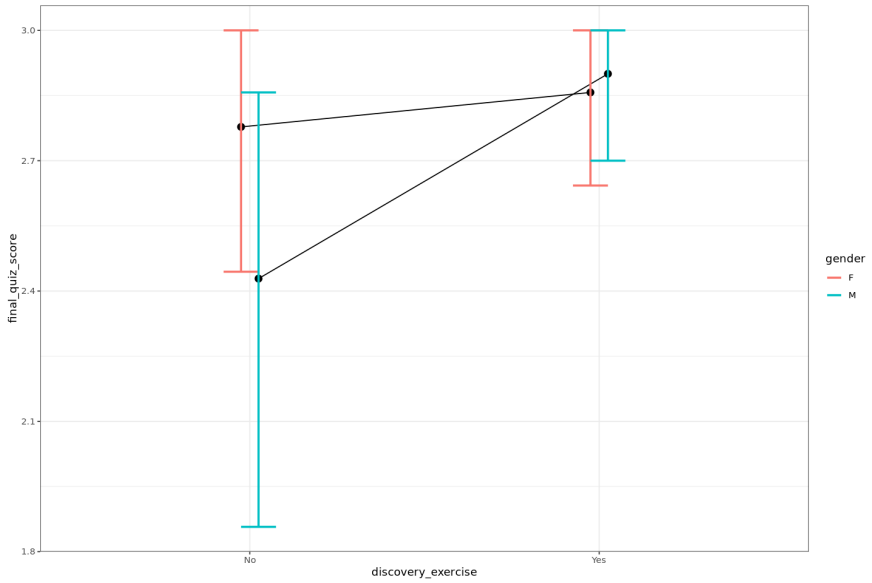

Figure 15: Bootstrapped Quiz scores of female and male groups depending on discovery exercise

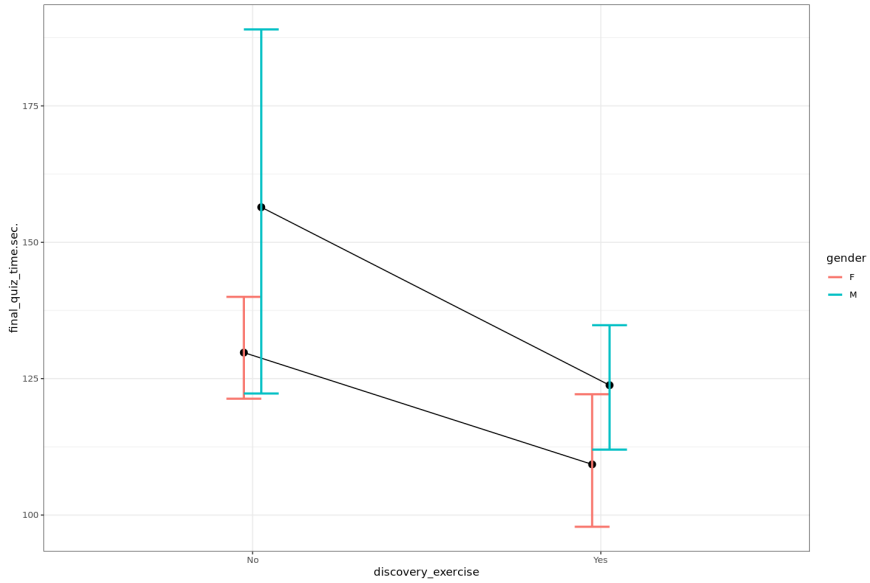

Figure 16: Bootstrapped Quiz time of female and male groups depending on discovery exercise

From the plots, we can see that the quiz score has increased and quiz time has decreased for both genders in the case of taking the discovery exercise.

However, testing the data for normality gives us the data is not normal even after performing transformations such as inverse and log and we can cannot statistically conclude anything about the interaction effects. Note that in case of Figure 16 the lines are more or less parallel indicating no interactions.

Thus in conclusion we would like to say that although there is a difference in quiz score and quiz time across gender and discovery exercise we cannot conclude the result is statistically significant or not as we are unable to perform any test on the data.

From the previous result, we had that there were no statistical differences in the quiz score depending on whether the participant was a male or female or whether they took the discovery exercise or not.

8 Conclusions and Reflections

Overall, as part of the course project for Digital Education, we have created an experiment to introduce cryptography to first year highschool students. While we had originally planned to present both Caesar and Vigenère Cipher in this lesson, the time constraints forced us to drop the Vigenère Cipher and we and created a first Orchestration Graph (OG) as shown in Figure 17.

The essential learning objectives of the lesson are to define plain text and ciphertext, encode and decode a message using the Caesar cipher and perform a brute-force attack to decode a ciphertext encoded with the Caesar cipher and describe its complexity.

Moving on to the Frog platform to design the experiments, we quickly encountered some technical constraints in implementing the OG shown in Figure 17. The major constraint for us was not able to programmatically verify the answers of the students without developing a new module for FROG. Thus, we moved from text-based answers in the discovery exercise to multiple-choice questions to get instant scores. However, for the remaining practice exercises and quiz, we decided to continue with textual answers and not show

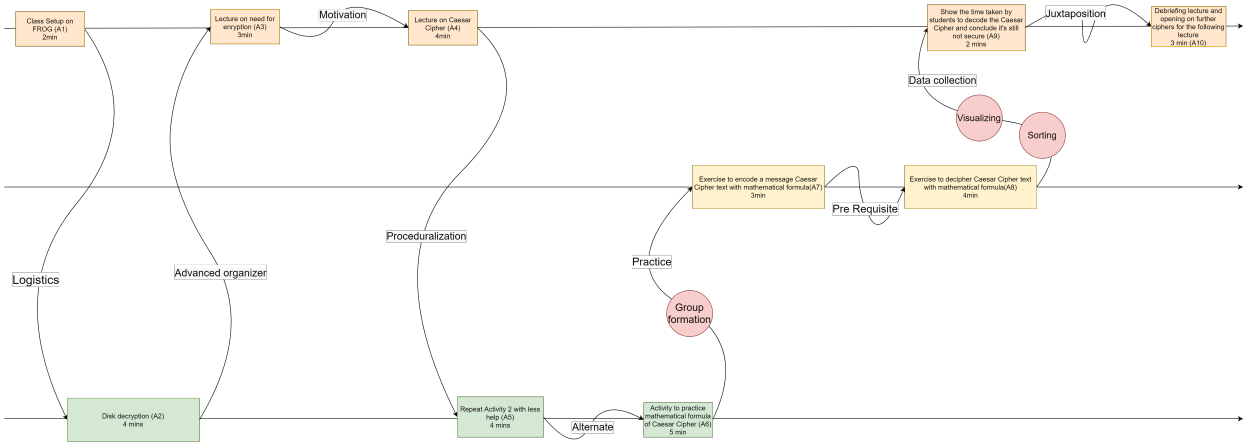

Figure 17: Initial Orchestration Graph

the performance results immediately in the debrief session but complete post-analysis of the answers. This resulted in the new OG as described in Section 6.

The hypothesis we wanted to test with this experiment was if students who learn by doing a discovery exercise followed by instructions and problem-solving activities will solve exercises faster and perform better than students who learn only by instructions and problem-solving activities. At the same time, we wanted to observe if female students would perform better than the males in a cryptography exam.

On experimenting with two classes of 40 students in total, we concluded that irrespective of gender or if a student has done a discovery exercise before the instruction there was no significant difference in the scores in the test. However, the amount of time taken by the students who took the discovery exercise was significantly lesser than those who did not take it. There were no significant differences in the time taken to solve the quiz across the genders. Prima facie we cannot conclude if there is or there is not an interaction effect between the variables or not (as we were unable to perform two way ANOVA test due to the data not being normal). However, we see that there is a difference in males for quiz score and decrease in time for both the gender who took the discovery exercise.

The reasons for these conclusions could be multiple and some of them even unknown. However, having only three questions in the test may be the first problem and thus most of the data is around the maximum score. Also, both the groups are not perfectly balanced both in terms of size. There could also be students who are inherently good at studies or who are interested/not interested in the topic or the teacher.

In reflection, we learned a lot on how to design a class and verify a hypothesis through experiments. However, if given the chance to modify our lesson, we would include more questions in the final quiz and also perform analysis by comparing the balanced datasets either before or after the experiment.

Appendix A Pre-test survey

This survey contained 3 questions. The second and third questions only had to be answered if the previous ones were answered positively.

- Class A : 7 out of 25 students stated they had already heard about the Caesar Cipher. Out of these 7 students, only one of them declared knowing how it worked, but did not answer correctly to the encryption question
- Class B : No student stated having heard of the Caesar Cipher.