

# User Study Report

Human Computer Interaction, ETH Zürich, Fall 2022

Group 13

**Luca Pinggera**  
lucapin@ethz.ch

**Mattias Schenk**  
mschenk@ethz.ch

**Nikhil Sethukumar**  
nsethukumar@ethz.ch

**Ziruo Xiao**  
xiaozi@ethz.ch

## 1 Introduction

The idea behind the project is that a beginner to programming could use it to grasp some of the basic concepts used in coding. In order to do this, the beginner is guided in using an interface which allows them to put together a program which drives a game. In this case we have chosen the simple game “Asteroids” where the player moves a rocket sideways and shoots at asteroids raining down at the ground.

The main interface of the project allows coding this game, and once coded, the user can play the game. We have two prototypes for this interface. A block based coding platform is shown, where blocks are to be dragged and dropped into corresponding blanks to make the game run. The whole game is divided into chunks visible on the right sidebar, and the left sidebar allows choosing of blocks. The middle area is where the code and blanks are visible.

## 2 Study Design

### 2.1 Apparatus

The study was conducted on a laptop using a mouse as the input modality.

### 2.2 Independent variable

- **Interface Layout:** Prototype A has all blocks available displayed on the left sidebar. Prototype B has only the blocks of the correct type for a specific blank displayed on the blank clicked.
- **Block fill-in modality:** Users fill in blocks to blank in Prototype A by directly dragging the block from the left sidebar and dropping it in the blank. In Prototype B, users have to first click on the blank in the middle panel, and drag and drop blocks of the corresponding type that appear on the right.

### 2.3 Dependent variable

- **User-perceived ease of use:** User’s feeling of how easy to use the two interfaces. Measured in three dimensions: how intuitive the interface is, how readable the interface is, and how easy is it to find the needed blocks.
- **Task completion time:** Time taken for users to finish the same task with the two interfaces.
- **Number of clicks the user does:** Number of clicks user does in order to complete the same task with the two interfaces.

### 2.4 Hypotheses

- **Hypothesis 1:** There will be no effect of the block selection interface on the users perceived ease of use.
- **Hypothesis 2:** There will be no effect of the block selection interface on the task completion time.
- **Hypothesis 3:** There will be no effect of the block selection interface on the number of clicks the user does.

### 2.5 Experimental procedure

We first formulate a protocol for the whole experiment process. Each of our group members recruits participants from people he/she knows and conducts the experiment with them following the protocol:

- **Initialize:** Set up hardware and software for testing and recording. Decide on which prototype the participant should start with based on their test ID.
- **Set up participant:** Introduce them to the purpose and procedure of the experiment. Let them fill in the first questionnaire with some personal information.
- **Experiment on the first prototype:** Present and introduce the participant to the interface. Introduce the training and testing task: complete a

specific frame for each task following the paper template. Ask for and answer any possible doubts then process the training task. After that, set up timer, record the total clicks and start the testing task. After finishing the task correctly, record the time and clicks, and ask the participant to fill in the second questionnaire with experience with this prototype.

- **Experiment on the second prototype:** following procedure same as the first prototype.
- **Ask general questions and End:** Ask general questions orally and do records, say thanks and the test is done.

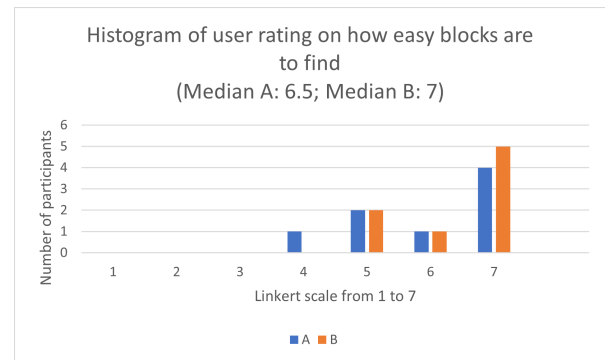
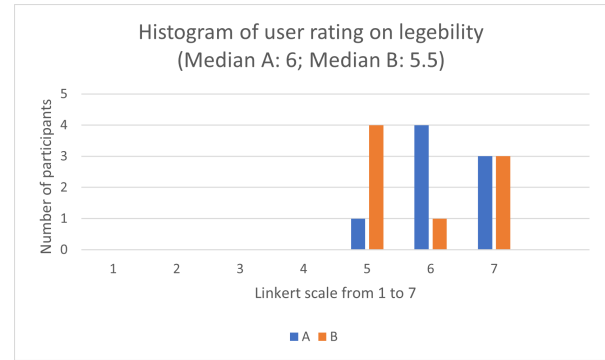
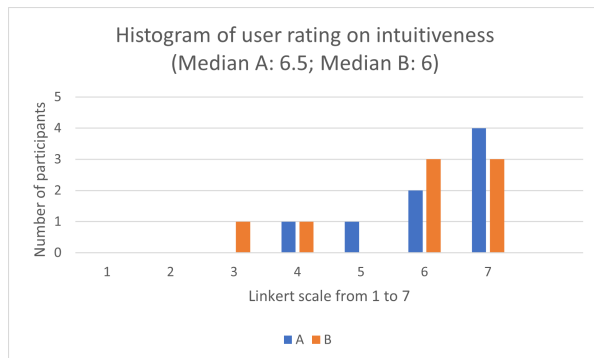
## 2.6 Participants

Gender, age, and previous experience with programming are recorded of participants. 87.5% of the participants are male and 12.5% are female. As for the age, the mean is 29(years) and SD is 10.23(years). The mean and SD for previous experience with programming, ranking 1 to 7 in integer from none to professional, is 3.5 and 2.14.

## 3 Results

### 3.1 Hypothesis 1: There will be no effect of the block selection interface on the users perceived ease of use.

To measure the ease of use we asked our participants in the questionnaire three questions regarding the intuitiveness and the legibility of the different interfaces and how easy it was to find desired blocks. The user had was able to select form a 1 to 7 Linkert scale and the histograms of the users' feedback can be found below in Figures 1 to 3.



#### 3.1.1 Reporting and Analysis

To compare the effect of the interface on user-perceived ease of use, we conducted a non-parametric Wilcoxon signed rank test on our ordinal data. Because the data is ordinal we don't perform the Shapiro-Wilk test and Levene's test.

With interface A, participants rated the intuitiveness with a median of 6.5, the legibility with a median of 6, and the easiness to find desired blocks with a median of 6.5. With interface B, participants rated the intuitiveness with a median of 6, the legibility with a median of 5.5, and the easiness to find desired blocks with a median of 7.

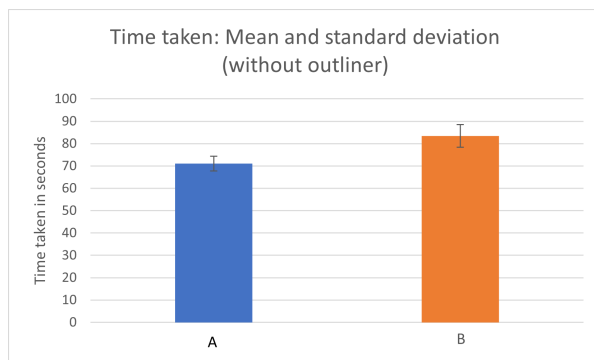
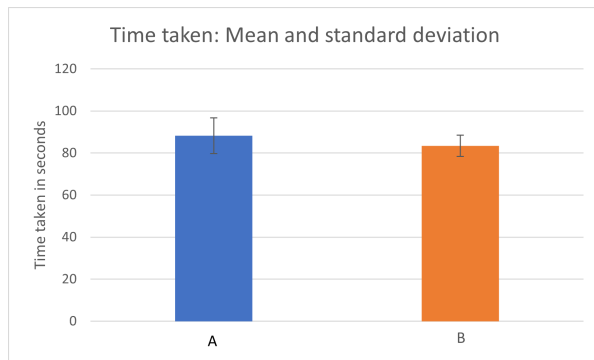
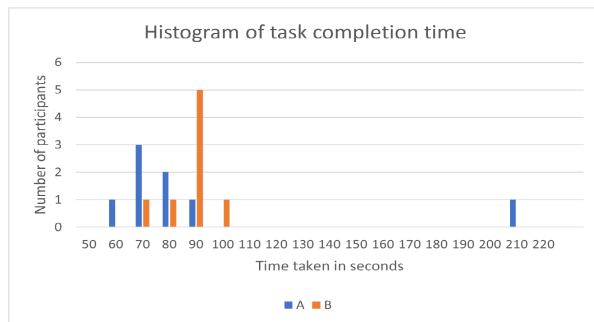
The mean difference between the two groups was not statistically significant; With a t value of 4.5 and p value of 0.41 for intuitiveness, a t value of 2.0 and p value of 0.26 for legibility, and a t value of 8.0 and a p value of 0.58 for the easiness to find desired blocks.

These results indicate that the perceived ease of use for the users did not differ much between the two interfaces and we therefore fail to reject the null hypothesis. The qualitative feedback does also not allow any conclusion about which prototype seems to be preferred.

### 3.2 Hypothesis 2: There will be no effect of the block selection interface on the task completion time.

For each participant the time to complete the task was measured and was plotted in the histogram below.

Just as in the dataset for the clicks used for interface A there is also an outlier which heavily skews the mean and standard deviation of interface A. The participant made a detour from the given task and started to test the interface for bugs. The plots below of the mean and standard deviation demonstrate the skew inflicted by the outlier. By applying the Shapiro-Wilk test it can be seen that the outlier causes the distribution to not be normally distributed. Therefore, for the rest of this statistical analysis the outlier will be excluded from the dataset.



#### 3.2.1 Reporting and Analysis

To compare the effect of the interface on completion time, we conducted a paired samples t-test. Through the Shapiro-wilk test and Levene's test we can assume that our dataset is normal distributed and homoscedastic.

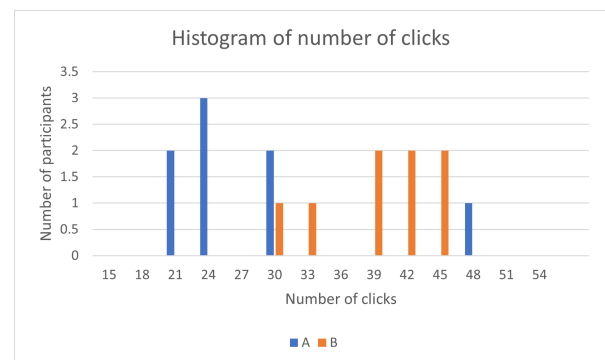
With interface A, participants on average used 71 seconds (SD = 8.94 seconds) to complete the task. With interface B, participants on average used 83 seconds (SD = 9.23 seconds) to complete the task.

The mean difference between the two groups was statistically significant; with a t value of -3.04 and a p value of 0.018.

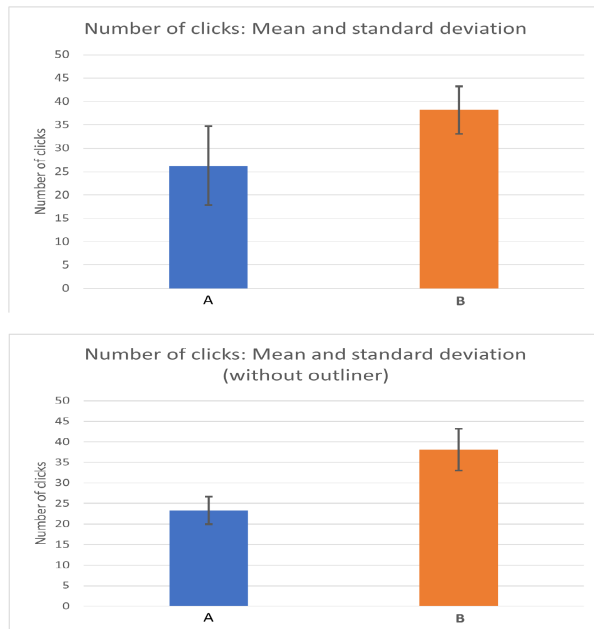
These results indicate that working with interface A is faster than with interface B and the null hypothesis can therefore be rejected  $H_0$ .

### 3.3 Hypothesis 3: There will be no effect of the block selection interface on the number of clicks the user does.

The number of clicks was collected by the program while the participant was executing the given tasks and the used clicks for interface A and B were plotted in the histogram below.



In the dataset for interface A there is one outlier which heavily skews the mean and standard deviation of interface A. The participant made a detour from the given task and started to test the interface for bugs. The plots below of the mean and standard deviation demonstrate the skew inflicted by the outlier. By applying the Shapiro-Wilk test it can be seen that the outlier causes the distribution to not be normally distributed. Therefore, for the rest of this statistical analysis the outlier will be excluded from the dataset.



### 3.3.1 Reporting and Analysis

To compare the effect of the interface on the amount of clicks, we conducted a paired samples t-test. Through the Shapiro-wilk test and Levene's test we can assume that our dataset is normal distributed and homoscedastic.

With interface A, participants on average used 23 clicks (SD = 3.3 clicks) to complete the task. With interface B, participants on average used 38 clicks (SD = 5.0 clicks) to complete the task.

The mean difference between the two groups was statistically significant; with a t value of -5.63 and a p value of 0.0008.

These results indicate that interface A uses considerably less clicks than interface B and the null hypothesis can therefore be rejected  $H_0$ .

## 4 Discussion

As we fail to reject hypothesis 0, users' subjective feeling on ease of use with the two panels does not differ much on their first try. But when looking into the quantitative analysis of the objective time consumed and clicks required, prototype B costs more in both dimensions. This suggests that having one more step to click the blank first to help reduce choices of blanks does not make users feel easier to use while costing more time and clicks.

As for the general question part, we got two comments related to the different experiences with the two interfaces, suggesting B was easier to use but A helps one better understand what's going on in regards to the code, and

one can probably work faster with A than with B as soon as he/she is used to the interface. We may imply from these comments that A may provide better learning and interactive experience in the long run.

We could therefore make a preliminary conclusion that interface A is better than B. And we agree that studies involving more methods and data shall be applied for further analysis.

## 5 Limitations

Due to time and the number of participant restrictions, we only had tests with a fixed goal (implementation based on a given template). As an interface aimed at a long-term study tool, long-term user experience with the interface would be critical to the evaluation. Bias in experience with users in different programming backgrounds would also be an important dimension to refer to.

As for the setting for studying hypothesis 1, questions on user-perceived likelihood to make mistakes/performance is not as expected could be added, as it's mentioned a few times in the general feedback but we could hardly analyze it without quantitative data. The other questions may not be well designed and described enough.

## 6 Future work

For hypothesis 1, our data suggest that user-perceived ease of use is relatively similar to the two interfaces, which may also suggest that our questions could be more selective and mindful. More specific hypothesis and study settings could be made in order to collect the data we exactly need. Moreover, we could only make quite preliminary analysis and conclusions with the variables studied. For the overall study design, we should try to think more as a whole and refer to more study methods. In this way can we do further quantitative analysis between variables and draw more findings.

## 7 Conclusion

In this user study, we studied differences in user experience between two interfaces of a programming learning tool. We made three hypotheses and rejected two, figuring out that Prototype B which needs a click on the blank before providing fewer choices of blocks would cost more time and clicks for the same task than A while the user-perceived ease of use is quite similar. We made some preliminary conclusions, discussed the limitation, and suggest what further work could be done.