Study Information

Hypotheses

1. Does the baseline have an influence on the results of the brain activation analysis?
2. Can we use one of the four baselines as general baseline to be used in program-comprehension studies?
3. What differences and similarities in reading behavior and brain activation occur between the five different tasks? Are these findings in line with the personal experience of the participants?
4. Can we predict the response time and brain activation during program comprehension using the data collected during the four other tasks?

Design Plan

Study type

Experiment - A researcher randomly assigns treatments to study subjects, this includes field or lab experiments. This is also known as an intervention experiment and includes randomized controlled trials.

Blinding

No blinding is involved in this study.

Is there any additional blinding in this study?

No

Study design

This study is a within-subject design. Each participant will see all program comprehension tasks and all four baselines. We plan to use 28 program comprehension tasks and 7 baseline tasks each.

No files selected

Randomization

We randomize the ordering of the baselines between participants to not always have the same baselines at the first trials.
Example:
first participant: baseline reading -> baseline math -> baseline relax -> baseline complex thinking
second participant: baseline relax -> baseline math -> baseline complex thinking -> baseline reading

Furthermore, we randomize the program-comprehension tasks to smooth the learning effects during the study.

We use the python random method to randomize tasks read from .csv sheets, which determine the presentation ordering in the study.

Sampling Plan

Existing Data

Registration prior to creation of data

Explanation of existing data

No response

Data collection procedures

Participants will be recruited through advertisements at the university and via mailing lists. Participants will be paid 10€ for agreeing to participate. We aim to have 20 participants overall. Participants must be at least 18 years old and have basic knowledge in English and programming to solve the tasks.

No files selected

Sample size

20 (human) participants

Sample size rationale

This number is in line with multiple other studies at our lab and in our field.

Stopping rule

We will stop if there are no new participants signing up for two weeks.

Variables

Manipulated variables

We let the participants solve 4 different baselines and the program comprehension tasks. We then evaluate the program comprehension task with respect to its preceding baseline. Everybody sees all baselines, during the analysis we split them into groups.

No files selected

Measured variables

We measure the response time, the correctness of answers to the tasks, the brain activation using an EEG and the eye movement using an eye tracker. Additionally, we assess general demographic data as well as programming experience prior to the study and participants' subjective opinions on the tasks afterwards.

No files selected

Indices

As we analyze the behavioral data, we will calculate the mean and the standard derivation of the response times.

No files selected

Analysis Plan

Statistical models

For statistical tests, we plan to use ANOVA, t-test ,and cohen's d, if the requirements are fulfilled. If not then we will use Kruskal-Wallis, Mann-Whitney U, and cliff's delta.

We will calculate the mental work load based on the EEG data by computing the theta/alpha power ratio, and we apply relative power analysis per band and per channel along the alpha, beta, gamma, and theta bands to create topoplots. For this analysis we will use the four different baselines and then test, if the results are different based on which baseline was used.

No files selected

Transformations

We use these methods to transform the EEG data:
We remove unwanted noise from the EEG data, particularly motion artifacts, using Hamming-filtered finite impulse response (FIR) filters.
Then, we removed power line noise using a notch filter with a lower cutoff frequency of 49 and an upper cutoff frequency of 51 hertz.
Additionally, for more robust performance in subsequent analysis, high-frequency noise were removed using a bandpass filter in the range of 4 to 200 Hertz.

Inference criteria

We will use the standard $p<.05$ criteria.

Data exclusion

We exclude all trials where the response option "Next" was selected, as well as the data points lying, where the time to completion lies outside 1.5 times the interquartile range.

Missing data

As long as at least 50% of the data are available, the subject will be included in the analysis. If this occurs more often, we will have to invite more participants to compensate the overall missing data.

Exploratory analysis

We take a qualitative look at the different baselines to make a recommendation which baseline could be used in further program comprehension experiments. We want this baseline to be as similar as possible to program comprehension in terms of brain activity.

Furthermore, we also look qualitative at the differences and similarities between all the used tasks to gain insight on how these tasks might be core skills for program comprehension. We look at: brain activation, eye movement and behavioral data. Here, we also use statistical tests (Kruskal-Wallis and Mann-Whitney U) to test for differences.

Additionally, we test if response time or brain activation of the programming tasks can be predicted based on the respective data from the baseline tasks using ElasticNet.

Other

Other

We use the same setup and the evaluation methods from the replication package of:

Norman Peitek, Annabelle Bergum, Maurice Rekrut, Jonas Mucke, Matthias Nadig, Chris Parnin, Janet Siegmund, and Sven Apel. Correlates of Programmer Efficacy and their Link to Experience: A Combined EEG and Eye-Tracking Study. In Proceedings of the ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE). ACM, November 2022. Acceptance rate: 22% (99 / 449). To appear.