Your Title Here

Experiment 2, Experimentation & Evaluation 2024

# Abstract

Short (120-130 words) summary of your entire report. Give the reader a quick idea of what you did and what the main findings were (if you prepare this report ahead of time, leave out the findings until after you finish the analysis).

# 1. Introduction

The focus of our experiment is to investigate whether people read identifiers faster when written in camelCase or kebab-case

The motivation behind this study is to determine whether using a specific separator in composed identifiers can speedup code reading.

Your final paragraph of the introduction should outline your proposed experiment.

|  |
| --- |
| **Hypotheses:** |
| Null Hypothesis (H0): There is no statistically significant difference in reading speed when using camelCase compared to kebab-case for composed identifiers. |

# 2. Method

The following subsections provide all the essential details required to replicate the experiment accurately.

## 2.1 Variables

The independent variables (i.e., the variables manipulated during the experiment) are the following: the style of the composed identifiers and the length of the identifiers.

|  |  |
| --- | --- |
| **Independent variable** | **Levels** |
| Composed identifiers style | CamelCase or kebab-case. |
| Identifier length | Short: 2 words  Medium: 3 words  Long: 4 words. |

The dependent variable (i.e., what is measured in the experiment) is the time taken to select the matching identifier.

|  |  |
| --- | --- |
| **Dependent variable** | **Measurement Scale** |
| time taken to select the matching identifier | Ratio scale (in ms) |

The control variables (i.e., what is kept constant during the experiment) are the following:

the number of tasks per participant, the number of distractor identifiers per task and the experimental environment.

|  |  |
| --- | --- |
| **2.1 VariablesControl variable** | **Fixed Value** |
| Tasks | 20 |
| Distractor identifiers per task | 3 |
| Environment of the experiment | Web application |

The blocking variables (i.e., measured potential sources of variability used to partition the experimental units into blocks, but are not part of the hypothesis) are the following:

the participants' programming experience and age.

|  |  |
| --- | --- |
| **Blocking variable** | **Levels** |
| Programming experience | High: intermidiate or advanced  Low: beginner |
| Age | High: 40 years or older  Low: under 40 years |

## 2.2 Design

Check off the characteristics of your experimental design:

**Type of Study** (check one):

|  |  |  |
| --- | --- | --- |
| ⃞ **Observational Study** | ⃞ **Quasi-Experiment** | ⃞ **Experiment** |

**Number of Factors** (check one):

|  |  |  |
| --- | --- | --- |
| ⃞ **Single-Factor Design** | ⃞ **Multi-Factor Design** | ⃞ Other |

**Between vs. Within** (check one): [for human subject studies]

|  |  |  |
| --- | --- | --- |
| ⃞ **Between Group Design** (independent measures) | ⃞ **Within Subject Design** (repeated measures) | ⃞ Other |

The experiment we designed is neither an Observational Study nor a Quasi-Experiment, as we are not looking at a phenomenon in a systematic and scientifically rigorous way in its environment, and we have complete control over manipulation of the independent variables. Therefore, it is an Experiment.

Moreover, the study employs a Multi-Factorial Design, as it includes more than one independent variable, and a Within Subject Design, since each participant is exposed to

all experimental conditions.

## 2.3 Participants

The participants in the experiment consisted of 30 individuals (15 males and 15 females) with an average age of 32 years (range: 20–50 years) and diverse professional backgrounds.

The participants were categorized into groups based on two key criteria: programming experience and age. Regarding programming experience, participants were divided into two groups: those with low experience (no prior programming experience) and those with high experience (two or more years of programming experience). Similarly, participants were also categorized by age into two groups: low age (under 40 years old) and high age (40 years or older).

Recruitment was primarily conducted through direct contact with personal connections, including classmates, parents, and other students without expertise in informatics.

As the study followed a Within-Subject Design, all participants were exposed to all experimental conditions and completed the entire set of tasks using the tool.

Describe who will take / took part in your experiment. Provide descriptive/summative statistics of their gender, age, professional backgrounds, and any other characteristics that may be relevant to your experiment. Also explain how you will recruit / recruited them (volunteers recruited through email, classmates who were asked to do this, etc) and how you will allocate / allocated them into the different study conditions, i.e., control group vs experimental group(s).

## 2.4 Apparatus and Materials

The experiment utilized several key tools and components.

The programming environment was based on React version 18.3.1, a JavaScript library designed for building user interfaces. Papa Parse (version 5.4.1) was employed to facilitate the conversion of data into CSV format for analysis.

Time measurements were handled using the Date.now() function, a built-in JavaScript function that provides millisecond-level precision.

The execution platform for the experiment was a browser environment, specifically Google Chrome version 131.0.6778.139. Data storage and management were facilitated through React's Context API, which ensured consistent capture and availability of experimental results for export as CSV files.

Describe in sufficient detail any relevant “props” that you used in your experiment. This could be the computer you used (exact model and specification), the software used (URL, version numbers), the way you measured, e.g., time (A stopwatch? A background process on the computer that got automatically triggered?). Omit needless detail (e.g., think whether details like the size of the table the laptop was placed on, or the hard disk size, might have affected your results or not).

## 2.5 Procedure

The experiment is conducted using a web-based tool designed to guide a participant through a series of tasks. Each session involves a single participant, with the total duration varying based on the time the participant requires to complete all tasks.

The session begins with the participant accessing a welcome page on the tool. This page provides clear instructions about the experiment's goal and an overview of the tasks the participant is expected to perform. The participant is informed that their task is to identify matching identifiers from a list and that their performance, in terms of time, is recorded.

After reading the instructions, the participant completes a short form to provide demographic information, including age, gender, and years of programming experience.

Following the demographic data collection, the participant proceeds to the task execution phase. The participant is presented with a total of 20 tasks, evenly and randomly distributed between camelCase and kebab-case identifiers. For each task, a short sentence consisting of two to four words, such as "move south," is displayed. Afterward, a list of four identifiers is shown, one of which corresponds to the given sentence while the other three serve as distractors. For example, after displaying "move south," the identifiers might include "move-source," "move-south" (the correct answer), "more-south," and "mover-sound." The participant is required to select the correct identifier as quickly and accurately as possible.

The tool automatically records the time taken by the participant to make each correct selection, measured in milliseconds. Once all 20 tasks are completed, the tool displays a completion message to the participant. The results, including task performance data and demographic information, are stored in a CSV file for later analysis.

Describe how you used your props and the participants to perform your actual experiment, i.e., how you actually carried out a single experimental run. What was done to the participants? What did they have to do? How long did each session take (unless this is an actual dependent variable)? If you did not have participants, explain, e.g., what software was started by whom in what order.

# 3. Results

## 3.1 Visual Overview

This subsection presents a visual summary of the experimental results. To ensure clarity and insight, the data is organized into a series of box plots that highlight the reading performance for camelCase and kebab-case across the entire population and all collected samples. By including the plot for the entire population, the visualizations not only highlight the influence of blocking variables (programming experience and age) but also provide a comprehensive view of the overall trends and outcomes of the experiment.

The figures are as follows:

* Figure 1: Comparison of camelCase and kebab-case performance across all participants.
* Figure 2: Comparison of camelCase and kebab-case performance among participants with high programming experience.
* Figure 3: Comparison of camelCase and kebab-case performance among participants with low programming experience.
* Figure 4: Comparison of camelCase and kebab-case performance among participants under 40 years of age (low age).
* Figure 5: Comparison of camelCase and kebab-case performance among participants aged 40 and above (high age ).

// TODO It can be very interesting to add also histograms to show the distribution of the data

Provide an insightful overview of the data you collected. This requires some engineering from your part, to find a good degree of summarization: On one end of the spectrum, you don't summarize, and report hundreds of raw measurement values in a block of text. On the other end of the spectrum, you report a single number (like a mean value). Both approaches are bad.

Instead, use appropriate visual summaries (such as **scatter plots**, **histograms**, **box plots**, or **empirical cumulative distribution functions**) to show the distribution of your data. If you have a very small number of measurement values, then report all of them in a **well organized table** (where rows and/or columns correspond to different levels of different factors).

## 3.2 Descriptive Statistics

The following section provides a detailed overview of the descriptive statistics computed for the experiment. These statistics, calculated for both the entire population and the individual samples, include the minimum, first quartile (Q1), median, third quartile (Q3), maximum, mean, and standard deviation.

The minimum is the smallest value in a data set. The first quartile (also known as Q1) is the value below which 25% of the data falls. The median is the middle value of the data (when it is ordered from smallest to largest). The third quartile (also known as Q3) is the below which 75% of the data falls, and the maximum is the highest value in the set.

The mean, or average, is calculated by summing all the values in the data set and dividing by the total number of observations, offering an overall indication of the data's center. Finally, the standard deviation measures the spread or dispersion of the data around the mean, indicating how much the individual values deviate from the average.

The computed descriptive statistics are reported below:

Table 1: Descriptive statistics for camelCase and kebab-case performance across all participants.

* Table 2: Descriptive statistics for camelCase and kebab-case performance among participants with high programming experience.
* Table 3: Descriptive statistics for camelCase and kebab-case performance among participants with low programming experience.
* Table 4: Descriptive statistics for camelCase and kebab-case performance among participants under 40 years of age (low age).
* Table 5: Descriptive statistics for camelCase and kebab-case performance among participants aged 40 and above (high age).

For each group or condition, summarize the set of measured values with a "five-number summary": **minimum**, **first quartile**, **median**, **third quartile**, and **maximum** (note: these are the statistics underlying a box plot).

Moreover, report the **mean** and **standard deviation** (note: for data that is not normally distributed, e.g., for multi-modal data, these two statistics may be less meaningful).

Make sure you explain – in your words – what these statistics mean “in plain English”, but don’t yet interpret them (this is for the Discussion section).

## 3.3 Inferential Statistics

The following section presents a detailed overview of the inferential statistics computed using the Independent Samples T-Test . These statistics, calculated exclusively for the entire population, include the t-statistic (t), degrees of freedom (df), and p-value (p). In addition to the computed inferential statistics, the descriptive measures Cohen’s d and the standard error of Cohen’s d (SE Cohen’s d) are also included as associated descriptive statistics.

The t-statistic (t) measures the size of the difference between the two groups (camelCase and kebab-case) relative to the variation in the data. A larger absolute value of t suggests a greater difference between the groups. The degrees of freedom (df) indicate the amount of information available for estimating the population parameters, which is influenced by the sample size. The p-value (p) represents the probability of observing the current data under the assumption that the null hypothesis is true.

Cohen’s d quantifies the size of the difference between the groups in terms of standard deviations, offering an indication of the practical significance of the difference. The standard error of Cohen’s d (SE Cohen’s d) reflects the precision of this effect size estimate, with smaller values indicating greater confidence in the measure.

The computed inferential statistics, along with the associated descriptive measures Cohen’s d and SE Cohen’s d, are presented below:

Table 1: Inferential statistics for camelCase and kebab-case performance across all participants.

If applicable, you then follow these up with inferential statistics – i.e., the **results of statistical tests** that you did in order to decide whether there were any “real” (i.e., not by chance) differences between the conditions/groups. You should also explain what statistical test you used, and, if not immediately obvious, why.

Make sure you explain – in your words – what these statistics mean “in plain English”, but don’t yet interpret them (this is for the Discussion section).

# 4. Discussion

## 4.1 Compare Hypothesis to Results

The hypothesis states that there is no statistically significant difference in reading speed between camelCase and kebab-case for composed identifiers. The results of the inferential analysis on the entire population support the null hypothesis, as the p-value exceeds 0.05, confirming that there is no significant difference in reading times between the two styles. This suggests that the choice of text format does not have a systematic impact on reading performance across the entire population.

Moreover, the low value of Cohen's d (Cohen’s d = -0.113) further reinforces the conclusion that any observed differences in reading performance within the population are not meaningful.

However, the descriptive statistics provided additional context. The mean and median reading times for CamelCase were consistently higher than those for kebab-case across the entire population and all the samples. In addition, CamelCase displayed slightly greater variability in reading times, as indicated by its larger standard deviation and interquartile range compared to kebab-case.

Based on these results, the findings from the inferential analysis supports the hypothesis that reading performance doesn’t differ significantly between the two text styles. Although descriptive statistics suggest a potential disadvantage for CamelCase, however these trends were not strong enough to reach statistical significance at the population level.

Provide a brief restatement of the main results from the previous section, and if (or if not) these support your research hypothesis.

If there is a discrepancy between your hypothesis and the results of your experiment, speculate about why you were unable to find evidence to support your hypothesis.

## 4.2 Limitations and Threats to Validity

Acknowledge any limitations and threats to validity of your study, and how seriously these affect your results. How could these be remedied in future work?

## 4.3 Conclusions

End with the main conclusions that can be drawn from your study.

Appendix

# A. Materials

Any documents you used for your informed consent (information sheets, consent) or as part of your apparatus (e.g., manual, hand-out), please include them here.

# B. Reproduction Package (or: Raw Data)

Before, during, and after the experiment you collected all kinds of data. Don't ever throw such data away! Any plots, tables, summaries, and statistics provided in this report should be recreatable from the raw data you have.

If you only collected a small amount of data, put it in this Appendix right here.

If you collected data in forms that are better kept in separate files, then zip up those files, and submit them as a "reproduction package" supporting this report.