Homework Assignment

Experimentation & Evaluation 2022

Experiment 2

Alessandro Gobbetti
Bojan Lazarevski

Contents

Università della Svizzera italiana
Faculty of Informatics
Switzerland

Abstract

In this report we discover whether separating words in the `kebab-case` or `camelCase` improves code readability and by what extent. These two methods are compared and based on statistical and quiz data gained by voluntary participants. At the end, what is measured is the time the participant took to answer each question and whether it was correct or not. We also gather some personal details about the participants to be able to better categorize the results. What we observed is that participants recognize the correct kebab-case separated phrases in a shorter time than camelCase, but the answer accuracy seems to be somewhat similar between the two cases. In the following sections we provide a detailed explanation and discussion of the obtained results.

1 Introduction

Studies have shown that people read more accurately when the words are separated with some separator. In this experiment, we test whether `kebab-case` and `camelCase` methods improve code readability, by what extent, and whether one outperforms the other one. In the `kebab-case` every word in the phrase is separated with the character `-` and all letters are lower-cased. In `camelCase`, the initial word starts with a small letter and the first letter of every new word in the same phrase is capitalized. This experiment is based on gained data by voluntary participants by completing a short quiz. The quiz is separated in two parts. Initially, the participants answer a short query about personal data and coding experiences. This is used to categorize the results from the second part of the quiz. In this second part, the participants are asked several questions. A question consists of a word phrase and the participant has to pick the correctly spelled phrase out of 4 possible given answers in the shortest amount of time. At the end, we collect the data of each participant by saving the picked answers and their corresponding answering time and perform the analysis.

Hypothesis:

The response time and answer accuracy are the same for `kebab-case` and `camelCase` separated phrases.

2 Method

In the following subsections, all independent, dependent and control variables are listed, as well as a procedure on how to replicate the experiment. A complete procedure on how to run the experiment is present in subsection 2.5. It takes usually less than 5 minutes for a participant to finish the entire quiz on their personal laptop on our deployed website. After the quiz is completed, we automatically receive an e-mail with the participant's data and answers in a `JSON` format. The participants are notified that their data will be collected. On the initial starting page on our website there are plots present with some interesting statistics based on the results of the entire control group. Please read the README on GitHub for more information on the code and the ways to replicate the experiment.

2.1 Variables

In our experiment we include independent variables (Table 1), dependent variables (Table 2) , control variables (Table 3) and blocking variables (Table 4). Each one of these variables is used in the experiment in order to capture more combinations of the expected output and try to get as precise as possible to the actual solution.

The independent variables are variables that we as experimenters manipulate. In this experiment we ask the participants to pick the correct phrase out of 4 possible solutions written in `kebab-case` or `camelCase`. A phrase can be contained of 2, 3 or 4 words.

Independent variable	**Levels**
Length of phrase	2 words, 3 words, 4 words
Type of separator	`kebab-case` and `camelCase`

Table 1: Independent variables: what we manipulate.

The dependent variable describes what we measure and in what scale we represent our measurements. In our experiment we measure the answer response time in milliseconds. An answer can either be correct or incorrect.

Dependent variable	**Measurement Scale**
Response time	ms (milliseconds)
Answer accuracy	correct or incorrect

Table 2: Dependent variables: what we measure.

The control variables are the constant variables that ensure that every run of the experiment is executed in a stable environment to prevent bias in the results. In our experiment we ask the participants to attempt the quiz on a laptop/PC device. The questions and the answers ordering are randomized to ensure that a participant cannot learn the questions and answers by heart and compromise the results. All 14 questions and phrases are written in English.

Control variable	**Fixed Value**
Type of device	Laptop
Ordering of questions and answers	Randomized
Phrase language	English
Number of questions	14

Table 3: Control variables: what we keep constant.

The blocking variables show potential sources of variability we measure and that help us partition the experimental units into blocks, but are not part of the hypothesis.

Blocking variable	**Levels**
Gender	Male or Female
Age group	< 18, 18-25, 26-35, 36-45, 46-55, 56-65, >65
Education level	None, Primary, Secondary, High school, Bachelor, Master, PhD
Years of programming experience	None, 1-2, 3-5, 6-10, 11-15, 16-20, >20

Table 4: Blocking variables

In this experiment, we don't measure whether the participants are using a mouse, a trackpad, a touchscreen, or whatever else, since we are not interested in how fast they are able to answer compared to others, but rather the relative time they spend on each question.

2.2 Design

Characteristics of the experimental design:
Type of Study: **Experiment**
Number of Factors: **Multi Factorial Design**
Between vs. Within: **Within Subject Design**

In this experiment we use 2 independent variables, therefore it is a multi-factorial design that allows us to analyze how they interact with each other. There are two dependent variables. The experiment is performed in an artificial environment where participants are invited to answer randomized quiz questions. Every

participant is tested on both `kebab-case` and `camelCase` phrases and the response time and accuracy of each answer is measured and collected. Since all users are exposed to the same user interface and run the quiz on same conditions, the experiment is in a within subject design.

2.3 Participants

The participants are all voluntary participants who we invited by message or email to participate in our study. In the first part of the quiz, we collect some of their personal data in order to be able to categorize the results from the second part of the quiz and generate the plots based on those blocking variables. We gather information such as gender, educational level to see if higher education builds up a habit of reading words easily. For somewhat similar reasoning we also ask the participants how many hours they spend in front of the computer and how many years of programming experience they have. Lastly we ask if the participant prefers kebab or camel case method of coding and then compare with their results to actually see if that is the case. The general data we obtained from participants is displayed on Figure 1. Once the participant finishes the quiz, we automatically receive a `JSON` file with their results. The participant can also decide to exit or cancel the quiz at any point of time, for which we do not receive any results as they are incomplete or invalid. In order for the participants to get the hang of the quiz and understand what kind of study we are performing, before the quiz starts there is a short description on how to complete it, as well as two tutorial questions for which we do not collect the data.

2.4 Apparatus and Materials

The users are asked to perform the quiz on a laptop or PC device and focus in a quiet environment. The time the user takes to answer the questions is measured in milliseconds from the moment the possible answers are displayed to the moment the user clicks on the correct answer. The correctness of the answer is also determined.
The data collected is stored in a `JSON` object and sent to us at the end of the quiz using automatic email.

2.5 Procedure

The moment the participants enter on the deployed page where the quiz is, they encounter a text description. This description is meant to be read by every participant. It first describes the goals of the experiment and then explains the topic of the quiz and how to perform and answer the questions on it. Once the participant has read the description, they perform a trial (tutorial) run where they are asked to answer on 2 questions to actually understand how the system and the quiz work. The data of the tutorial questions is not stored. Once the user has finished the tutorial, the actual quiz starts. The quiz is separated in two parts. In the first part, the participant answers personal and educational questions and has unlimited time to do so. Once they finish this, they are taken to the most important part of the quiz. There are 14 questions that test whether the participant finds kebab or camel case more suitable. A question is composed of a phrase. The user has 3 seconds to memorize the phrase and then has unlimited time to pick the correctly-spelled one in either kebab or camel case (depending on the question). Even though the participant has unlimited time, he is advised to answer as quick as possible as that is the goal of the experiment. This procedure is done for all 14 questions of the quiz. Once the user finishes the quiz, they are given some interesting statistics, such as answer accuracy and average response time along with a visual representation of each answer individually

with its corresponding response time. After every successful and complete run of the quiz, the data of each participant is sent to us in a JSON format by email.

3 Results

In this section, we present the results of the experiment. For a more interactive experience and more plots, please visit our website! Moreover some of the discussion here is based on the plots displayed on the website (that are not displayed on this report), so it is highly recommended to look at those plots while reading the report. Most of those plots are concerned about the blocking variables and plots that separate the obtained results based on participants' characteristics (age, educational or programming background, gender, etc.).

3.1 Visual Overview

Figure 1: Visual overview on the population.

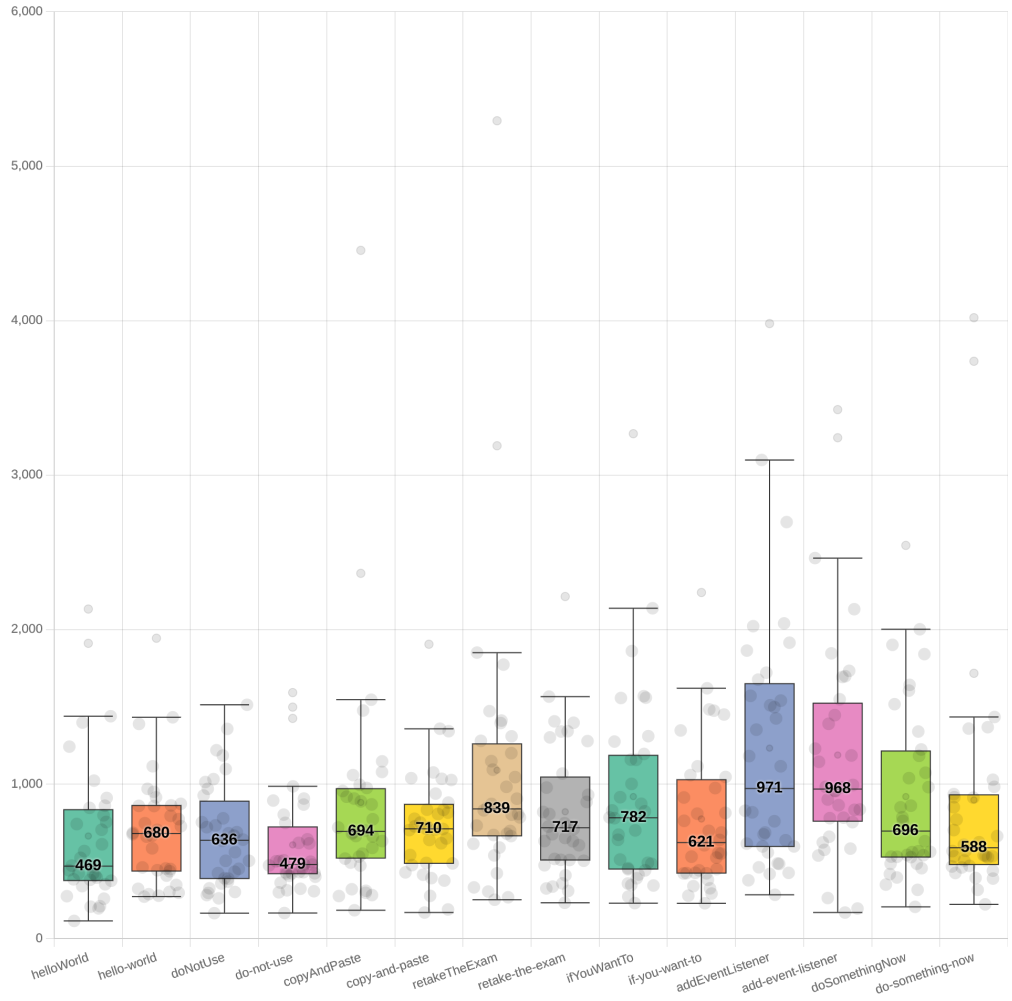

Figure 2: Times for each question.

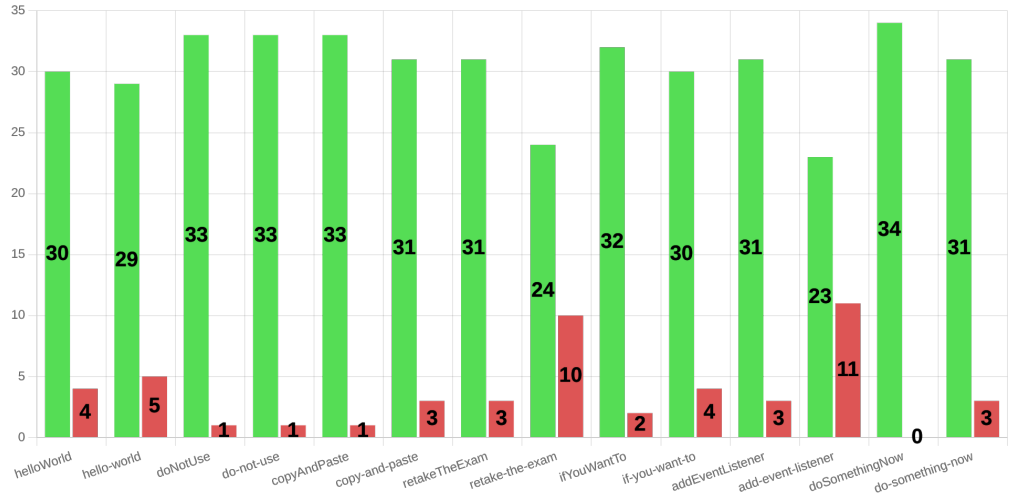

Figure 3: Accuracy for each question.

3.2 Descriptive Statistics

In this section we present some descriptive statistics of the results. For each group or condition we summarize the set of measured values with a "five-number summary": minimum, first quartile, median, third quartile, and maximum.
The minimum and the maximum values are the smallest and the largest time in nanoseconds it took the users to answer that question. The first quartile is the value that divides the first 25% of the data from the remaining 75%. The median is the value that divides the data in half. The third quartile is the value that divides the last 25% of the data from the remaining 75%.
In Table 5 we present the statistics results for the answers of the users.

Correct answer	kebab-camel	Min	Q1	Median	Mean	Q3	Max
helloWorld	camelCase	114	375.5	468.5	663.618	834	213
hello-wold	kebab-case	271	436.75	679.5	703.647	861.25	1944
doNotUse	camelCase	164	388.5	636	664.735	889	1513
do-not-use	kebab-case	165	419.75	479	606.088	722.25	1592
copyAndPaste	camelCase	183	520.25	693.5	880.912	970	4455
copy-and-paste	kebab-case	168	486.75	710	738.059	868.5	1905
retakeTheExam	camelCase	251	664.75	839	1089.471	1260	5294
retake-the-exam	kebab-case	231	508.25	717	820.618	1045.25	2214
ifYouWantTo	camelCase	229	449.75	781.5	918.794	1184.75	3268
if-you-want-to	kebab-case	228	423.5	621	773.765	1028	2240
addEventListener	camelCase	283	596	971	1232.353	1649.75	3981
add-event-listener	kebab-case	168	759.5	967.5	1187.735	1523.25	3424
doSomethingNow	camelCase	205	527.5	695.5	918.559	1213.75	2545
do-something-now	kebab-case	221	478.75	587.5	896.971	930	4019

Table 5: Descriptive statistics of the response times of the participants.

3.3 Inferential Statistics

In **??** we describe information gained when performing paired sample t-test on the mean of time between kebab-case and camelCase. Additionally, in Table 6 we perform the same test but on answer accuracy instead. Each sample consists of 7 measurements.

Measure 1	Measure 2	t	p
mean time kebab-case	mean time camelCase	2.3774	0.05496
mean accuracy kebab-case	mean accuracy camelCase	2.3711	0.05543

Table 6: Paired samples t-test

4 Discussion

4.1 Population

As can be seen from Figure 1 the test has been conducted on a population of 34 participants. The age distribution is mostly between 18 and 30 years old, but we managed to get a few participants for each age group. The educational background is very high, and most of the participants have a university degree, this is not really a proper representation of the world population, but it is still good enough for the purposes of this

experiment. The majority of the participants are non-programmers, but we still have a lot of programmers (also with several years of experience). Another interesting thing to look at is how many hours people spend on the computer per day, as can be seen from Figure 1 we got a gaussian distribution between 0 and more than 10 hours per day, with the exception of the 5-6 hours slot that is way less populated due to the fact that we don't have enough participants.
Overall, the population distribution spans a wide range of ages, education, programming experience, and PC usage.

4.2 Compare Hypothesis to Results

In the previous section we have shown the results obtained from the experiment. In this section, we will discuss the results and compare them to the hypothesis. Our initial H_0 hypothesis says that the response time and answer accuracy will be the same for both kebab-case and camelCase phrases. However, by analyzing the data and visual representations of it in the previous section, we can claim that the hypothesis doesn't hold and reject it. On Figure 2 we have a box plot that represents the average response time for each pair of questions. Based on the plots, we can see that for each pair, the kebab-case phrase box is positioned lower on the y-axis. This means that participants generally discover the correct answer faster when it is written in kebab-case rather than camelCase. On the same figure, we also represent the median and outlier data points which are not taken into consideration. A more detailed representation of the same figure can be found as a descriptive statistic on Table 5 where we also show information on minimum, maximum, median, mean, Q1 and Q3 for both kebab and camel cases. On Figure 3, a plot is shown visualizing the answer accuracy per question. The green bar represents the correct responses and the red bar the incorrect ones. Opposing to the response time analysis, we now notice that phrases written in camelCase are more readable and participants tend to recognize the correct answer more frequently. Why we believe this is the case is discussed in subsection 4.4. On the website we also show plots that are fixed more on the blocking variables and the differences in results between each group. For example, we see that people with no experience in coding have slower response time or that men have a faster response time than women participants. We do not discuss the reasoning behind these observations as it is beyond the scope of this course and computer science in general.

4.3 Limitations and Threats to Validity

- **Environment**: it is very important that participants perform the quiz in a suitable environment. It is very easy to get distracted by events that surround the participant. Despite the fact that we ask from the participants to stay focused when doing the quiz, it is important that they perform the quiz in a distraction-free environment. Distracting environment can affect the response time and answer accuracy.

- **Number of participants and their background**: It is important to gather data from a large number of participants with different educational or programming backgrounds in order to be able to categorize the results in a correct manner. Since we have a few blocking variables, it would be ideal to have enough participants in each block in order to extract meaningful data from it. This means that we must have large number of participants that fall into one of these groups. In our case, we limit the number of participants to a few dozens of people, thus limiting the amount of blocks we can analyze.

- **Question number and variety**: In our quiz, we stick to 14 questions for which we think that we will gather enough data to carry out the experiment. However, in order to properly analyze the outcome of the experiment, we believe that it is necessary to have more questions with phrases of different lengths or phrase patterns designed by lexicology experts. If this is the case, we might also be able to conclude the precise case *why* one method is better than the other one, instead of just observing the data.
- **Device type**: we are asking for the participants to do the quiz on a PC or laptop. Our frontend is not yet adapted to smaller devices such as a smartphone.

4.4 Conclusions

The main conclusion we achieved from this experiment is that participants tend to respond faster when dealing with kebab-case phrases. We believe this is the case because words are separated by - and is more natural to the human eye to read phrases where each word is physically separated with some character. However, people usually do not tend to over-focus on the correct spelling on the word because they generally recognize what word it is even if it is spelled incorrectly. There is a famous experiment and analysis from Cambridge university about this phenomena and a very famous quote that captures the essence of it:

> *Aoccdrnig to a rscheearch at Cmabrigde Uinervtisy, it deosn't mttaer in waht oredr the ltteers in a wrod are, the olny iprmoetnt tihng is taht the frist and lsat ltteer be at the rghit pclae. The rset can be a toatl mses and you can sitll raed it wouthit porbelm. Tihs is bcuseae the huamn mnid deos not raed ervey lteter by istlef, but the wrod as a wlohe.*

This is possible to read as long as the words in the sentences are physically separated. In camelCase instead, we have higher answer accuracy but slower response time. We believe this is the case because the phrase is represented as a single word and the participant needs to properly focus in order to recognize the correct word, generally reading character by character, instead of observing the entire word like the kebab-case. This means that the participant is more focused and selects the right answer more times as they read the phrase character by character looking for an error. For example, if we represent the same quote in camelCase, we notice that it is immediately harder to read the text and takes us more time to do so. But, we pay more attention to what we read and can easily spot the mistakes in the text:

> *aoccdrnigToARscheearchAtCmabrigdeUinervtisyItDeosntMttaerInWahtOredrTheLtteersInAWrodAreTheOlny IprmoetntTihngIsTahTheFristAndLsatLtteerBeAtTheRghitPclaeTheRsetCanBeAToatlmsesAndYouCanSitll RaedItWouthitPorbelmTihsIsBcuseaeTheHuamnMnidDeosNotRaedErveyLteterByIstlefButTheWrodAsAWlohe.*

Because of this we believe that when we want a faster response time or easier readability, it is better to go with kebab-case. But, if we want precision and higher accuracy we should stick with camelCase. Most of the time people still pick one method over another based on personal (and design) preferences.

It is still possible to take part in this experiment as a voluntary participant and complete the quiz. You can do so at the following page. It will help us to go further into the topic and make a more precise analysis of the obtained data.

Appendix

4.5 Materials

- How to Design and Report Experiments: Andy; Hole, Graham J. Field - Book
- iCorsi slides
- Chart.js - A JavaScript library for drawing interactive charts.
- EmailJS - A JavaScript library for sending emails directly from the code. Used to send quiz responses from participants.
- GitHub Pages - where our page with the quiz is deployed.

4.6 Reproduction Package (or: Raw Data)

The reproduction package is included and can be found in the following GitHub repository:
`https://github.com/Alessandro-Gobbetti/Experimentation-and-Evaluation/Assignment2`
In the repository, there is also the `JSON` file that contains all raw data of the participants' answers named `data.json`.
To replicate the experiment carefully read the README.