

MFM GROUP

Experimentation & Evaluation



Authors: Mehmet Fatih Tekin, Mike Fiorita, Mustafa Özyürek

camelCase vs kebab-case

Abstract

This study investigates how readability, particularly speed and accuracy, is affected by identifier style camelCase versus kebab-case. The experiment was carried out using a web application that was created with Vue.js. In a recognition exercise, participants had to choose the right identifiers from a list of possibilities in phrases that were presented in both styles. Each participant completed 20 tasks, 10 for each style. Response times and accuracy data were recorded and stored in a CSV file for further analysis. JASP was used for statistical analysis, producing separate t-tests and boxplots to compare the two styles' performance. The findings are intended to provide information on the best identifier conventions for increasing developer productivity and code readability.

Introduction

Readability in software development is essential for efficiently comprehending and managing source code. In programming, composed identifiers names composed of many words are common because they aid in conveying the function, class, and variable purposes. These identifiers' formatting, however, may affect how well a developer reads and understands code.

According to natural language processing research, using explicit separators like spaces or symbols might improve reading comprehension and speed. Though constructed identifiers usually follow certain standards like camelCase (capital letters as word boundaries) and kebab-case (hyphens as separators), it is unclear whether these conclusions hold true for programming languages.

The purpose of this study is to determine whether identifier style affects recognition accuracy and speed. To ascertain which style is easier to understand, we specifically compare camelCase with kebab-case. The research's conclusions can help guide best practices for naming conventions, which will increase developer productivity and promote code readability.

To accomplish this, we ran a controlled experiment in which participants had to choose among a variety of distractions the proper identifiers written in camelCase or kebab-case. We seek to ascertain whether one style provides a quantifiable reading advantage by examining accuracy and response times.

Hypotheses:

Hypotheses 0 identifier: There is no significant difference in reading speed or accuracy between camelCase and kebab-case identifiers.

Main hypotheses identifier: camelCase is faster to read rather than kebab-case.

Hypotheses 0 length: There's no difference in reading speed or accuracy comparing short sentences and long sentences between camelCase and kebab-case.

Main hypotheses length: camelCase is faster to read with both short and long sentences compared to kebab-case.

2. Method

The experiment on identifier style readability can be replicated using the comprehensive instructions in this section. The dependent variables (accuracy and response time) are measured, and the independent variable (identifier style: camelCase vs. kebab-case) is manipulated. The web application created in Vue.js, the tasks completed by participants, and the procedure for gathering data are all described in the experimental setup. To ensure consistent results throughout testing, hardware and software setups are specified, and methods are well-defined to ensure accurate measurements. The findings were saved in CSV files and subjected to statistical analysis to determine how different the two identification types were in terms of readability and how effective they were at understanding code.

2.1 Variables

Independent variable	Levels
Identifier styles	camelCase or kebab-case

Dependent variable	Measurement Scale
Reading speed	Time measured in seconds
Reading accuracy	Correct/Incorrect response

Control variable	Fixed Value
Sentences	Difficulty, number of sentences used
Hardware	MacBook Air, Apple M2, 8 GB RAM, macOS: Sequoia 15.1.1

Blocking variable	Levels
Demographic Information	Age, Gender, Programming Experience, Educational Background, Preferred Identifier Style, Work Experience, Favourite Programming Language

2.2 Design

Type of Study (check one):

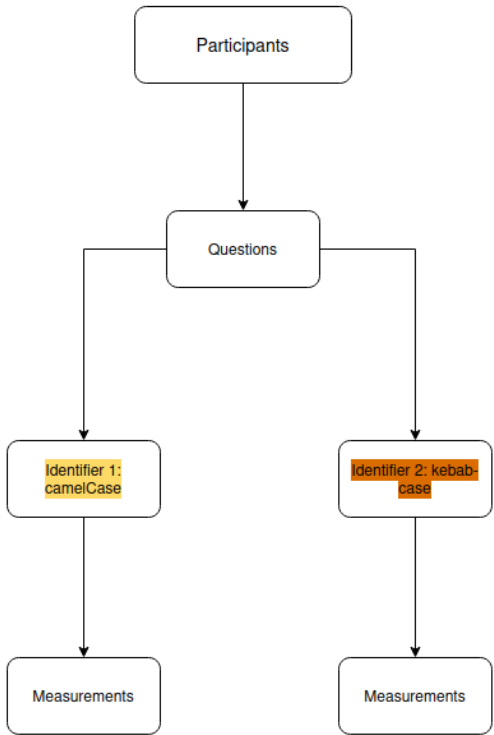
<input type="checkbox"/> Observational Study	<input type="checkbox"/> Quasi-Experiment	<input checked="" type="checkbox"/> Experiment
---	--	---

Number of Factors (check one):

<input type="checkbox"/> Single-Factor Design	<input checked="" type="checkbox"/> Multi-Factor Design	<input type="checkbox"/> Other
--	--	---------------------------------------

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

<input type="checkbox"/> Between Group Design (independent measures)	<input checked="" type="checkbox"/> Within Subject Design (repeated measures)	<input type="checkbox"/> Other
--	---	---------------------------------------



2.3 Participants

Participants were taken from a group of people that included people with mostly some programming experience, but only a small number that had none. The age range was in the early to mid 20's, with only one being over that. There was no requirement for specific expertise in camelCase or kebab-case, which allowed for a mix of familiarity levels. To examine possible effects on performance, demographic information such as age, years of programming experience, and preferred coding style was gathered. As for control group

2.4 Apparatus and Materials

The experiment was done by using a web application developed through the VUE framework and Javascript. Its main feature was the recording of the 20 answers from the participants, divided into 10 camelCase and 10 kebab-case experiment tasks, which were then saved into a .csv file, containing all the information needed to perform data analysis through JASP, mainly the time taken for the participant to click on the answer and whether or not it was the correct one. The application also contained a page instruction to explain how the test worked, as well as demographics form to gather data for programming/work experience, age and identifier/programming language preferences.

2.5 Procedure

The procedure was divided into 4 parts:

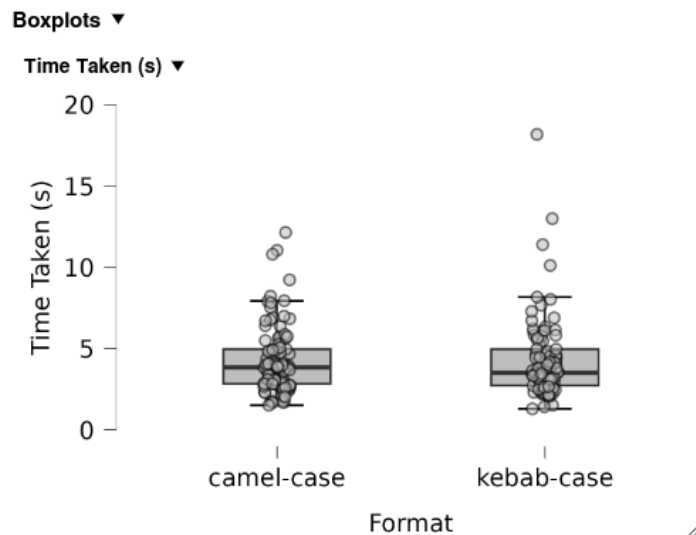
1. **Introduction and Consent:** Participants were welcomed, provided with instructions, and gave consent to participate.
2. **Demographics Collection:** Participants completed a short form capturing key demographic and professional data.
3. **Task Execution:** Participants were shown a target sentence followed by a set of identifiers. Their task was to identify the correct match as quickly and accurately as possible. Each participant completed 20 trials, with camelCase and kebab-case tasks randomized to minimize learning effects.
4. **Data Collection:** Response times and accuracy for each task were automatically recorded and saved for analysis.

3. Results

3.1 Visual Overview

Boxplots for all the data

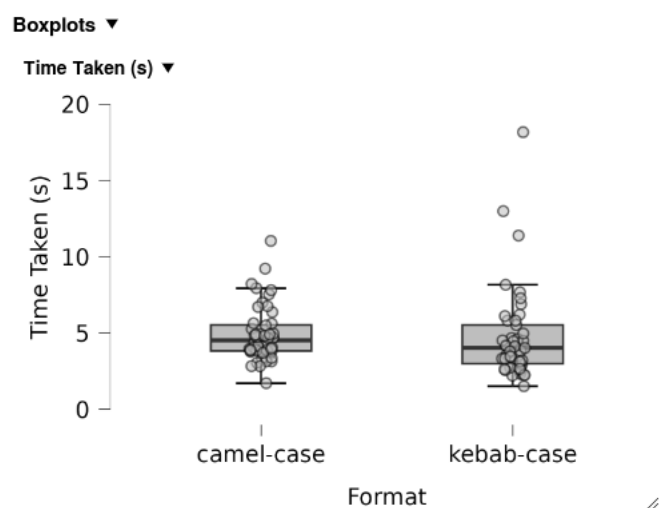
Across all data, including short and long words, the boxplot displays the Time Taken (s) to identify identifiers in camel-case and kebab-case. No discernible difference in response times is seen by the medians for the two techniques, which are extremely comparable. Similar performance variability is suggested by the comparable interquartile range (IQR).



Some response times for both methods are over 10 seconds, and one extreme outlier for kebab-case was around 20 seconds. Participants generally responded similarly regardless of identification type or phrase length, as evidenced by the majority of reaction times clustering between 3 and 6 seconds. Although there aren't any noticeable differences overall, rigorous statistical study would verify whether any differences are significant.

Boxplots for long sentences

The Time Taken (s) to identify identifiers in camel-case and kebab-case for large texts only is shown in the boxplot. Given how comparable the medians are for the two forms, there isn't much of a difference in response times for longer phrases. Indicating constant variability in both contexts, the interquartile ranges (IQRs) stay similar.



Kebab-case outliers are more prevalent, though, with one extreme reaction time approaching 20 seconds, while camel-case outliers don't go over 10 seconds. Even so, the majority of response times for both approaches fall between 4 and 7 seconds. Formal statistical testing is necessary to validate the data, which generally indicates that identifier style may not have a significant impact on performance for lengthy sentences.

Boxplots for short sentences

For short sentences, the boxplot displays the Time Taken (s) to find identifiers in camel-case and kebab-case.

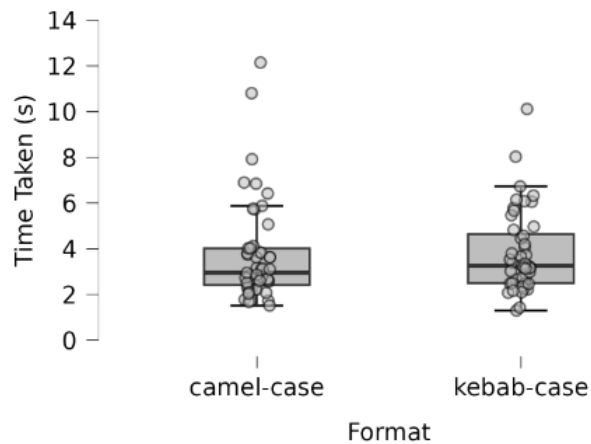
The medians for the two approaches seem to be very similar, suggesting that response times are not that different. Additionally, the interquartile range (IQR) is comparable, with kebab-case showing a little wider spread, indicating modest variability.

Both styles contain outliers, but they are less frequent and less

severe than those found in boxplots for lengthy sentences. For brief words, the majority of response times fall between 2 and 4 seconds, indicating faster performance overall. In conclusion, the two forms function quite similarly for short sentences, despite the fact that reaction times for kebab-case are somewhat more varied.

Boxplots

Time Taken (s)



3.2 Descriptive Statistics

Descriptive Statistics for all data

As a preface, we gathered 200 results from 10 subjects, but only 95% of them were used, since we do not consider incorrect answers in our statistical calculations. Of the incorrect answers, 3.5% were from the kebab-case, while the remaining 1.5% camelCase. This can be taken as a pre-analysis sign that camelCase is slightly easier to read than kebab-case.

Looking at the descriptive data, we've come to find out that the results are evenly split in terms of identifier performances and in only 2 cases the preferred identifier didn't match with the mean obtained.

Putting all the results together, we obtain that the median of camelCase is 3.850s, while for kebab-case 3.520s, which is slightly lower. The mean times are also close, with camelCase (4.315 s) slightly higher than kebab-case (4.282 s). However, the difference in means is very minimal, suggesting no significant advantage for either style on average.

Standard deviation is smaller for camelCase, suggesting that participants' performance was more consistent with camelCase compared to kebab-case, which shows greater variability. The minimum and maximum times reveal a wider range for kebab-case (1.300 s to 18.180 s) compared to camelCase (1.520 s to 12.150 s). This variability could be influenced by outliers or differences in participants' familiarity with kebab-case.

In conclusion, the data obtained indicates that both cases have similar average and median processing times, with camelCase showing slightly more consistent performance across participants. The null hypothesis is at this moment confirmed, contradicting our main assumption.

Descriptive Statistics ▼

	Time Taken (s)	
	camel-case	kebab-case
Median	3.850	3.520
Mean	4.315	4.282
Std. Deviation	2.098	2.515
Minimum	1.520	1.300
Maximum	12.150	18.180

Descriptive Statistics for long sentences

As a preface, we gathered 200 results from 10 subjects, but only 95% of them were used, since we do not consider incorrect answers in our statistical calculations. For long sentences, the median response time for camelCase is 4.525s, while for kebab-case it is slightly lower at 4.030s. Similarly, the mean times are close, with camelCase at

4.903s and kebab-case at 4.776s. The difference in means remains minimal, suggesting no significant advantage for either style in long sentences.

Descriptive Statistics ▼

	Time Taken (s)	
	camel-case	kebab-case
Median	4.525	4.030
Mean	4.903	4.776
Std. Deviation	1.823	3.068
Minimum	1.720	1.520
Maximum	11.050	18.180

The standard deviation is notably smaller for camelCase (1.823) compared to kebab-case (3.068), indicating that participants' performance was more consistent with camelCase. Kebab-case, on the other hand, shows greater variability, which could be attributed to outliers or participants' differing familiarity with this style.

The minimum and maximum values also reveal a wider range for kebab-case (1.520s to 18.180s) compared to camelCase (1.720s to 11.050s). This higher variability in kebab-case suggests that while some participants performed quickly, others struggled more with long sentences.

In conclusion, for long sentences, camelCase exhibits more consistent performance, while kebab-case shows greater variability despite slightly faster medians.

Descriptive Statistics for short sentences

For short sentences, the median response time for camelCase is 2.950s, whereas for kebab-case it is slightly higher at 3.255s.

The mean times are also close, with camelCase at 3.739s and kebab-case at 3.819s, further indicating no major difference between the two styles.

Descriptive Statistics ▼

	Time Taken (s)	
	camel-case	kebab-case
Median	2.950	3.255
Mean	3.739	3.819
Std. Deviation	2.206	1.765
Minimum	1.520	1.300
Maximum	12.150	10.120

The standard deviation, however, is smaller for kebab-case (1.765) compared to camelCase (2.206), suggesting that performance for short sentences was more consistent with kebab-case. This contrasts with the findings for long sentences.

The minimum and maximum response times for kebab-case (1.300s to 10.120s) are slightly narrower compared to camelCase (1.520s to 12.150s), implying that camelCase included a few slower responses that increased its overall range.

In conclusion, for short sentences, response times for both styles are very close, with kebab-case showing slightly more consistency and camelCase exhibiting a broader range of performance.

3.3 Inferential Statistics

The phrase "inferential statistics" refers to the computations used to ascertain whether observed data differences are statistically significant or most likely the result of chance. The separate Samples T-test was employed for this research since it is appropriate for comparing the means of two separate groups.

Inferential Statistics for all data

With a t-statistic of 0.099, the results reveal a very small difference between the means of kebab-case and camelCase. The null hypothesis cannot be rejected

because the p-value of 0.921 (higher than 0.05) indicates that this difference is not statistically significant. Furthermore, identifier style has a minor effect on performance time for all phrases combined, as indicated by Cohen's d of 0.014, which is extremely modest.

Independent Samples T-Test

	t	df	p	Cohen's d	SE Cohen's d
Time Taken (s)	0.099	188	0.921	0.014	0.145

Note. Student's t-test.

Inferential Statistics for long sentences

We compared the camelCase and kebab-case response times for lengthy texts using the Independent Samples T-test.

The resultant t-statistic, which shows a slight variation in means, is 0.246. This difference is not statistically significant, as indicated by the p-value of 0.807 (higher than 0.05), which suggests that there is no concrete proof that one identifier style is faster than the other for lengthy sentences.

Independent Samples T-Test ▼

	t	df	p	Cohen's d	SE Cohen's d
Time Taken (s)	0.246	91	0.807	0.051	0.208

Note. Student's t-test.

There is little practical difference between the two techniques, as seen by the extremely modest effect size suggested by the Cohen's d of 0.051. Even if the descriptive statistics show that there is more variety in kebab-case performance, the t-test verifies that this variation does not lead to a statistically significant difference.

Inferential Statistics for short sentences

Response times for short sentences were compared between camelCase and kebab-case using the Independent Samples T-test.

The findings indicate a little performance difference, with a t-statistic of -0.197. According to the p-value of 0.844, which is more than 0.05, this difference is not statistically significant, indicating that neither style clearly has an advantage for short sentences.

Independent Samples T-Test

	t	df	p	Cohen's d	SE Cohen's d
Time Taken (s)	-0.197	95	0.844	-0.040	0.203

Note. Student's t-test.

With a Cohen's d of -0.040, the negative impact size is extremely modest, underscoring the fact that there is no discernible difference between kebab-case and camel-case for short sentences. The statistical test verifies that, despite the existence of little variations in variability, these variations lack statistical significance.

4. Discussion

4.1 Compare Hypothesis to Results

The purpose of our study was to determine whether camelCase and kebab-case IDs differ significantly in terms of readability as indicated by response time. The null hypothesis was not disproven, according to the statistical findings. In particular, the t-tests revealed no statistically significant difference across all tasks (all phrases, lengthy sentences, and short words), and the mean and median times for both identifier forms were extremely similar.

The effect sizes (Cohen's d) were consistently very tiny, indicating that these changes are insignificant, even if kebab-case occasionally showed slightly faster median and mean times. It's interesting to note that camelCase performed more consistently, as shown by the lower standard deviations, especially for lengthy texts.

Its perceived stability and familiarity are further supported by the fact that a greater proportion of participants selected camelCase as their preferred identifier. However, there were no quantifiable variations in speed or accuracy as a result of this subjective preference.

There could be a number of reasons for the difference between the results and our original premise, which was that one style would have a quantifiable advantage. For example, participants may have performed similarly if they were equally familiar with both types. Furthermore, it's possible that the activities' brief duration and low level of difficulty prevented them from highlighting minute variations in reading.

4.2 Limitations and Threats to Validity

Several limitations and threats to validity must be acknowledged:

1. **Sample Size:** Only 10 persons participated in the study, despite the fact that we gathered 200 data points. The results may not be as broadly applicable to a larger developer population due to the very small sample size.
2. **Task Complexity:** Finding identifiers in brief, discrete sentences was one of the goals. Readability disparities may be exacerbated in real-world programming situations where identifiers are incorporated into longer code blocks and subject to contextual influences.
3. **Participant Familiarity:** It's possible that participants' past experiences with kebab-case and camel-case affected how well they performed. Its apparent constancy in spite of the small time differences might be explained if the majority of participants were more accustomed to camelCase.
4. **Outliers:** Although outliers were taken into account by statistical analysis, participants' sporadic mistakes or misreading may have contributed to the higher variability in kebab-case performance, which distorted the findings.

In order to overcome these constraints in subsequent research, we advise:

- To increase the generalizability of the results, a larger and more varied group of volunteers should be recruited.
- Creating projects that better mimic programming scenarios by incorporating longer, more realistic code portions.
- Using pre-experiment tests to account for participants' familiarity with each identification style.

4.3 Conclusions

Our research concludes that there is no appreciable difference between camelCase and kebab-case identifiers in terms of speed or correctness. Although the median and mean response times for kebab-case were marginally faster, these differences were statistically inconsequential and came with more variability. On the other hand, people subjectively preferred camelCase, which performed more consistently.

These results imply that, in straightforward situations, the decision between camelCase and kebab-case might not have a substantial effect on readability. However, developers may find CamelCase to be a more dependable option due to its familiarity and stability, especially for tasks that call for consistency. To confirm and build on these findings, more extensive research with bigger sample sizes and intricate code settings is required.

Appendix

You can find all the appendix inside the GitHub repository link:

<https://github.com/fatihtkin/Exp-Eval-Project2>

A. Materials

All materials used in this study are available through the following GitHub repository link.

The materials include:

- **Participant Instructions and Sentences with Distractors (PDF):** This document contains the task instructions, including explanations of camelCase and kebab-case identifiers, as well as the list of example sentences used in the experiment with distractor options for both identifier styles.

B. Reproduction Package (or: Raw Data)

The raw data and reproduction package for this experiment, including all collected data, are available through the following GitHub repository link.

The reproduction package includes:

- **Raw Data:** The data collected from participants, including response times, accuracy, and demographic information, formatted in CSV files for analysis.
- **JASP Files:** The JASP files containing the statistical analysis performed on the data.
- **Screenshots of JASP Results:** Visual outputs from JASP, including descriptive statistics, t-test results, and boxplots used to analyze the data.

By accessing the repository, users can replicate the experiment, analyze the raw data, and review the results from the statistical analysis.