How Do API Documentation and Static Typing Affect API Usability?
Experiment

- Five hours max duration!
- 20-30 participants
- Documentation: plain text (?)
- Language: Dart (?)
Design

- 2x2 design
  - IVs: type system, documentation
  - DV: development time
Results

The experiment has been performed at the University of Duisburg–Essen and the University of Koblenz–Landau. In total, 25 students were used as participants in the analysis (two subjects were removed from the data, one subject dropped out of the experiment). Each student was in the fifth semester or above and had finished basic programming courses at the corresponding university. More than half of the students were bachelor students, the other ones were Master students in computer science (or related studies). All students were males between the ages of 22–30.

The data was collected in multiple sessions, depending on student schedules. Students were randomly assigned to one of the four groups in order to reduce the potential for accidental bias, the reasons for which are explained thoughtfully by Vogt (see [32]).

Figure 1 describes the names of the groups and the number of subjects. Almost all groups have the same number of subjects (each has six subjects) with the exception of group 1 which has seven.

Group 1 has both documentation and static type system, group 2 had no documentation and static typing. The other groups had the same configuration of documentation with dynamic typing.

Table 1: Groups: groupNumber (numberOfSubjects)

<table>
<thead>
<tr>
<th></th>
<th>Documentation</th>
<th>Static TS</th>
<th>Dynamic TS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Yes</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>No</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Yes</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>No</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>

4.1 Empirical Data

Table 2 shows the raw anonymized data for the experiment: The first row shows each participant number, the second the experimental group. Continuing rows show development time (DT), file switches (FS), number of switches to documentation (DS), time spent reading documentation (DRT) and coding time (CT).

Three participants (8, 20, and 23) were unable to complete the development task within our time constraints. All of the participants that were unable to complete the tasks happened to be in the group without documentation. Two of these (20 and 23) used the dynamically typed language, while one other (13) was given the statically typed language.

In order to ease the reading of the raw measurements, Figure 3 illustrates the development times via boxplots, which reveals the following possible trends, which will be evaluated more carefully with standard statistical assessments:

1. The raw median amount of time taken to complete the task appears to increase from groups 1–4.
2. The raw difference between static and dynamic typing, with documentation (groups 1 and 3), appears to be large.
3. The raw difference between static typing without documentation, and dynamic typing with documentation, does not appear to be large.
4. The raw differences between static typing without documentation and dynamic typing either with or without documentation, appear to be rather small, although this is potentially mitigated by the fact that 2 of six participants in group 4 did not complete the task.

Although boxplots are a common way to visualize the results of the measurements, boxplots might be considered problematic in this experiment, because the number of subjects for each groups is relatively small. Hence, Figure 4 illustrates the development times by showing each individual measurement (and without showing any descriptive statistics). Again, the raw measurements seem to suggest that there is the tendency that the absence of static types and the absence of documentation increased the measured development times.

Figure 3: Boxplot for raw development time measurements

Figure 4: Raw development time measurements

4.2 RQ1: Documentation Usage

There are many ways in which an analysis of our first research question can be quantified, although which is the "right" one is not...
Discussion

• Why do people insist on using dynamically typed languages?
• Can we show they're better (in some cases)?
  • If so, what might the implications be?
• Gradual typing to the rescue?
External Validity

• Dart?
Type Systems Research Exercise

• First, identify hypotheses: what are the usability implications of types?

• Then: how might we test them?