

## *Assignment 5: Empirical user studies*

**General comments:** Remember the general statements from previous assignments about carefulness and taking proper time. For this one, also consider that there is a need to have face to face sessions. Make a good time planning at the begin and consider some buffer at the end in case there are any unforeseeable problems coming up.

**General rules:** Email your results in a PDF to Wolfgang (huerst@uu.nl). Include all group members in the cc of this email (students not included in the email will get NO grade!) and consider the usual formalities:

- For the email: Use the subject [INFOMSCIP] Assignment  $\langle i \rangle$  group  $\langle j \rangle$
- For the PDF file: Use the name INFOMSCIP-X-Y.pdf where X is the number of the assignment and Y is your group number. Put all group members' names and student IDs at the begin of the PDF.

Deadlines are strict. Each additional day after the deadline will result in a grade deduction of 1.0.

**Important note:** Make sure to read **the whole assignment** carefully before starting it. Step 3 will require you to make some compromises that might have an impact on your decisions in step 1. This is not bad planning from my side, but a general aspect of research: you always have to think ahead and plan your whole approach beforehand to avoid running into problems at later steps.

**Deadline for this assignment:** Send the email to Wolfgang before Thu, Oct 17, 11am.

**General remarks:** This assignment is about *experimental user studies*. You will specify a related research question and/or hypothesis, design the experiment, execute it with a few people, and analyze the data (to some degree). The topic will be related to typing on mobile phones. I assume that at least one member of your group has a smartphone that you can use for these tests.

**To deliver:** The result of this assignment should be delivered as a PDF file. Start this file with a list of your team members. Before handing it in, add a rough estimate of hours spent on this assignment for each team member plus a short description of special tasks that he or she has done (unless you split all tasks equally or did all work together). One sentence is enough for this; no details required unless you deem them necessary for your grading.

**First step:** *Specification of the research.* Before we do research, we need to motivate it, introduce our goals, and specify our research correctly. The latter is usually done by introducing a research question and a hypothesis (but other options exist).

Here, we want to study typing on mobile phones with a touch screen and a so-called *soft keyboard*, i.e., a set of on-screen keys operated via a touch screen. There are plenty of ways to implement such a keyboard and what option you chose will have an impact on typing behavior, performance, and experience. Because the focus here is on doing an empirical study, we are not looking into new creative ways to implement on-screen typing, but want to study existing approaches. For the sake of this assignment, you can also skip the literature study that should usually be done before you start your actual research.

Sit together with your group and discuss possible options. Check first what kind of phones you have available, and see how their keyboards are implemented and what parts of the implementation are interesting and suitable to study with an empirical experiment. This can be on a rather “high level”. For example, different manufacturers might do things a bit different (does it, e.g., make sense to compare Android vs. iOS keyboards?), or offer alternative ways for text input (e.g., Android phones often allow input via the so-called *Swype* technique in addition to normal typing). Or you can study a more “low level” feature. For example, different types of feedback are often provided when typing on a soft keyboard, including changing visuals, vibrations, or sounds when you hit a key. The latter two can often be turned off, so you could evaluate their impact on typing performance. Check what options the device(s) that you have available offer and decide on an interesting and relevant aspect that you want

to study. Also, make sure that you actually *can* study this aspect, i.e., that the results of your experiment will indeed allow you to draw conclusions (which requires you to already think about what you will be measuring and how). And consider pragmatic decisions (will you have devices available for testing, etc.). Note that there are also typing apps out there that allow you to record things such as typing speed. If you know one or are eager to do a little research to find one, you are welcome to use it, too.

Once you decided on a problem to study (i.e., your research aim or general goal), specify it concretely (i.e., phrase a research question and/or a hypothesis). Make sure to consider the characteristics that we discussed in the lecture when phrasing your research question. And keep in mind that it is okay and sometimes even essential that your research question narrows down the scope of your research problem.

**To deliver:** After listing all your team members at the begin of the PDF, address the things mentioned above. Be precise and to the point. No lengthy explanations with issues of minor relevance.

Use the following structure (but slight deviations are acceptable if they make sense and server your purposes):

- Research aim / general goal / research problem: ...

*Write one sentence or phrase to introduce it. Then add some text to shortly explain it; 1-2 sentences could be enough if you phrase it nicely.*

- Research question: ...

*Phrase your question. If necessary, you can add a short text to explain it.*

- Hypothesis: ...

*Not all research needs an hypothesis, but for empirical research it is often helpful (and if you do a statistical analysis it might be required). Thus, phrase one, and add some text explaining it (the text should illustrate why you make this assumption about the outcome of your experiment).*

Feel free to add any other text that you deem relevant, but avoid adding unnecessary information or issues of minor relevance (again, be precise and to the point). It is totally okay, and might even be better, not to use full text but phrasings and a bullet point structure.

*Some related comments and grading criteria:* The major issues and thus criteria for the grading are: Is the problem relevant and interesting? How well is the research question, hypothesis, and/or other

type of specification formulated. Is it specific? Can it be answered, verified, or resolved with an empirical study? Etc. Grading will also depend on how well you are “to the point”, that is, both missing relevant information as well as providing too much irrelevant content could both have a negative impact on your grade.

**Second step:** *Experiment design.* Now, you need to set up and clearly define the design of your experiment. Identify the variables (including their type, e.g., independent and dependent) and measures used, as well as the following things (copied from one of the lecture slides):

- Subjects: who, number, ...
- Environment
- Equipment / material used
- Other contexts
- Tasks to perform, instructions given, ...
- Within/between-subjects design
- Order (for within-subjects design) and other mappings (for both)

Notice that this list might not be complete and some of these items might not be needed for your test. There might also be things that we discussed in the lecture that are not listed here. And there might be relevant other aspects that we have not discussed yet or will not address at all, because there are too many of them, and some might only apply in special cases, such as your concrete study. See also the related remarks about grading below. Also, you will need to make some compromises in the setup because this is a one-week assignment, not a full-fledged research project (see info in step 3).

**To deliver:** Add the study design to your PDF document. There is no template, but make very sure that all information is there and clearly and nicely represented. I recommend using the layout to structure it in a way that key issues are easy to spot (e.g., don’t write full text but use bullet lists, font style, etc.). I often see reviews of submitted papers where the reviewer complains about missing information that is actually there, but gets just overlooked because it is badly represented. It is your job as author to make sure that the reader gets everything, not the readers’.

*Some related comments and grading criteria:* Clarity and completeness are the most important grading criteria for this step. If you miss

specifying things that have been discussed in the lecture and are of relevance for your experiment, you will get a deduction in your grade. If you miss specifying relevant things that have NOT been discussed yet, you will not get punished, but you can get extra credit if you consider them. Again, feel free to add any text that you deem relevant, but avoid adding unnecessary information or issues of minor relevance. The comments above about “being to the point” and presenting your work accurately apply here as well.

**Third step:** *The actual experiment.* In the second step, you (hopefully) specified how many subject you need for your experiment. Given the time and scope of this assignment, this will very likely not be feasible. But the purpose of this assignment is not to do a full research project but to learn how to do empirical research, so we can make a compromise here. For the sake of this experiment, assume you want to make a dry run with some people before the actual test to see if your setup and everything else works as planned. (Note: I always recommend to do this with at least one to two persons before a real experiment, e.g., in your master thesis, as well). For this assignment, other group members can mutually serve as test subjects. We will use the following pairs of groups: 1&2, 3&4, 5&6, 7&8, 9&10&11. That is, each of these groups should run a pilot test with 2-3 group members from the other one. Because we have an uneven number of groups, group 9 will use participants of group 10, group 10 will use some of group 11, and group 11 will use some of group 9.

Because we use existing implementations, we are also a bit limited in recording the data. For example, the time is too short to implement an automatic tracking and analysis of typing actions. Likewise, it will not be possible to implement automatic time logging. Yet, you will probably be able to track typing errors by either observing the participants or having them type in a document that you can analyze later (but notice that you cannot analyze, e.g., the number of corrections in such a case). You are allowed to make compromises (e.g., it might be okay to measure the timing with a separate stop watch; for some time measurements this might give you a sufficient approximation, for others not). But also think about more creative ways to solve this problem. (E.g., maybe you can mirror the phone screen to a desktop PC that can be monitored better by an observer? Or how about a screen recording that is analyzed later?)

Decide on these limitations, write them down (see comments under “To deliver” in the box below) and then go ahead and do the experiment.

**To deliver:** Complete the study design in your PDF by a paragraph or subsection entitled **Limitations** that introduces the above decisions and compromises that you made. Shortly describe their relevance and how they might impact the result.

*Some related comments and grading criteria:* It is important that you clearly list the potential limitations that your compromises might have on your results. You can illustrate that also by stating how you did it here and how you would do it ideally. For example, something like: “For the sake of the test run, typing speed is measured with a stop watch by the experimenter. In the final experiment, it will be implemented on the phone.” Notice that in this example, you also need to clearly specify how “typing speed” is defined.

**Fourth and final step:** *Analyzing the results.* After finishing your tests, go and analyze the results. Because you will likely not have enough data anyhow and we have not discussed it in the lecture yet, you do not have to do a full statistical analysis (e.g., significance test). Yet, depending on your project and what you measured, you should report, for example, averages, variances, etc. of your data. We did not discuss qualitative research yet (which includes for example observations by the experimenter), but if you want to add related comments, feel free to do so. If done correctly and nicely, you may get extra credit (keep in mind though that stating irrelevant facts might also have negative impacts).

There are no rules about what to analyze, because you specified that yourself already in the preceding steps (if you defined your research question properly and came up with a correct study design). Likewise, there are no requirements for how to represent your results other than that it should make sense, answer your research question, and solve the problem you are studying. Think about what will be the best way to do this. Is it, for example, a table with the data? A diagram? Both? If a diagram, what kind and how to plot the data? And do not forget to provide data about the subjects if you gathered it and if it is relevant.

Finally, do not just represent your results but draw some conclusions from them. First and foremost, use them to answer your research question. Because your study is only a little pre-experiment or pilot study, also discuss the issues that you might change for the final experiment. For example, as flawed as it may be due to the limitations, it could be that your little pre-experiment already identifies a clear trend that suggests to reconsider not just single aspects of the experiment design, but major things - in the extreme case even your original research question and goals.

**To deliver:** Add the results and analysis to your PDF. Structure it in the following way (minor changes are okay if you think they allow you to represent your results and conclusions in a better way):

- Results (*the actual data*)
- Discussion (*the analysis*)
- Conclusion (*this could include comments on how to change the actual experiment because of the conclusions drawn from your pilot study*)

*Some related comments and grading criteria:* As above, clarity, correctness, and suitability of the presentation, discussion, and conclusions that you draw are the major grading criteria. Other aspects, such as presentation style, relevance of the text (everything that is needed, but no irrelevant and unnecessary information), etc. apply as well.