The pro's and con's of research hypothesis

We begin with analyzing the pro's and con's of research hypothesis. It is inevitable to generate a research hypothesis in order to really know what is going on. Discussing the design of research hypothesis gives the conductor deep understanding concering the general topic, the relevance of his own study and especially what and how he likes to measure. Thus the researcher has to read other papers to find out what they did and how they would move on. This leads to an important understanding of related work. When designing hypothesis one always has to ask if the correct metrics are applied to measure what needs to be evaluated. Is the System Usability Scale (SUS) really an appropriate metric to find out if a recently developed interface is superior to the common standard? When considering these aspects Hornbaek describes "The Question". Every researcher has to raise a specific question before the experiment and later answer this question when conducting and evaluating it.

Hornbaek also discusses the limitations and potential issues when designing research hypothesis but also provides solutions to avoid it. For instance an experiment that only investigates one hypothesis may exclude interesting observations and only focuses on one specific metric. By designing an experiment with several hypothesis there are several aspects that could be covered and by designing two or three hypothesis one could enhance "the question" with additional evidence. For instance if you design an experiment to evaluate a new menu structure of an interface measuring only the task time may not be sufficient. Back it up with a questionnaire and see if the results point in the same direction. Another problem when developing research hypothesis is that by focusing only on confirming or rejecting hypothesis the researcher may lose vision or foresight about the potential of his experiment.

Self-reported vs objectively measured task performance

We already mentioned task time as a common metric to measure efficiency. We could measure the time from start to end to finish given tasks by the subject. We could build an interface which provides this data directly in the program. In some contexts, we could also use a stopwatch to measure time. But when we objectively measure the task performance by collecting time we still have no clue how the user really felt when interacting with our system. We may have a problem answering "The Question" by simply looking at numeric data with timestamps. We need to collect further information and since measuring time is quantitative data we could mix it with qualitative data. We could ask the user after each task to say a few words about how he felt and which problems occurred. This would be self-reported. We could also ask the user to describe his own performance with a number between 0 (bad) and 10 (good). We could compare this numeric data with the numeric data we collected by logging the experiment. But relying only on self-reported data may not be sufficient either since the answers may differ from one subject to another.

Asking HCI students in a HCI study

When designing a study the selection of suitable subjects is crucial in order to avoid bias and to make sure that the research hypothesis can be answered accordingly. In HCI many experiments are conducted with subjects from the University where the researcher typically works. These subjects are easy to acquire and do not require many motivation (e.g. money) to participate in our experiment. But when comparing two interfaces, a common and well-known interface and a new innovative approach, media informatics students could not be the right subjects to ask. They tend to be more interested in technology and may even have some experience in our experiment's topic. When comparing a smart speaker with no visual interface (e.g. Google Home) with a smart speaker with a screen (e.g. Google Home Hub) media informatics students may already work with an Alexa at home for years. If we ask medicine students for instance, they may not have as much experience with smart speakers. This can be good or bad for conducting our experiment. Good or bad depends on what questions we raised and how we would answer them. One could also argue media informatics students are human people after all and since our study proved that our innovative interface shows a better task performance than the traditional one when using media informatics students maybe the same results would emerge when asking medicine students. But as Hornbaek mentioned repeating the same study with different subjects is not always an exciting way to go.