Empirical Validity sec:research-methodology:empirical-validity

In sec:research-methodology:philosophical-stances, we state that this study primarily adopts a critical theorist stance. Critical theorists assess research quality by the utility of the knowledge gained Easterbrook:2007ws. Lau:1999vs established criteria on validating information systems research specifically for action research unifying four dimensions of the research (conceptual foundation, study design, research process, and role expectations) against 22 varying criteria. We also partially adopt constructivism as we attempt to model the developer mindset rich in qualitative data, to which eight strategies identified by Creswell:2017vn cover.

We identify possible threats to internal- (study design), external- (generality of results), and construct-validity (theoretical understanding) in the following sections, and describe how we mitigate these threats.

Threats to Internal Validity

Hawthorne Effect

Observational field techniques involving participants often run a risk producing misaligned results from laboratory versus environmental (practical) conditions. This is commonly known as the Hawthorne effect Robbins:2014tr,Draper:vb and careful consideration of this effect must be reflected when designing our controlled observation (ssec:research-methodology:experiments:2). We aim to carefully explain the purpose and protocol to research participants, encouraging them to act as much as possible as in their practical conditions. We also encourage the 'think-aloud' to participants protocol to reinforce this. By highlighting the Hawthorne effect to them, we anticipate that participants will be aware of the condition, and therefore avoid doing things that do not reflect real-world action.

Misleading Statements in Interviews

Similarly, threats to the interview survey instrument exist where participants do not often report differences in behaviour from what they actually do in practice Singer:2007tu. We anticipate that conducting interviews in a semi-structured manner may assist in following up with unexpected statements (as opposed to structured interviews) and additionally corroborate findings using Jick:1979el's concurrent triangulation method Jick:1979el to verify potentially misleading statements from participants with questionnaire results and observation findings.

Participant Observation Accuracy

Conducting participant observations is a skill that requires training. While every effort will be instilled to ensure all relevant observations are noted, it is impossible for a single observer to note every possible interaction that occurs in all observations made. Therefore, to validate the consistency of data collected, we may require rater agreement exercises Judd:1991ug and we will likely use a form of recording device (with participant consent) to ensure all information is transcribed correctly in the interview.

Unintended Interviewee Bias

Interviewers should introduce the research by which participants are involved in by describing an expiation of the research being conducted. However, the amount of information described may impact the bias instilled on the interviewee. For example, if the participant does not understand the goals of the study or feel that they are of the 'right target', then it is likely that they may choose not to be involved in the study or give misleading answers. On the other hand, if interviewees are told too much information, then they may filter responses and leave out vital data that the interviewer may be interested in. To mitigate this, varying levels of information will be 'tested' against colleagues to determine the right level of how much information is divulged at the beginning of the interview.

Poor Questionnaire Responses

Unless significant inducements are offered, Singer:2007tu report that a consistent response rate of 5% has been found in software engineering questionnaires distributed and in information systems the median response rates for surveys are 60% Baruch:1999vf. We observe that low response rates may adversely effect the findings of our survey, typically as software engineers find little time to do them. Tackling this issue can be resolved by carefully designing succinct, unambiguous and well-worded questions that we will verify against our colleagues and within the pilot study in a2i2, wherein any adjustments made from the pilot study due to unexpected poor quality of the questionnaire will be reported and explained. We also adopt research conducted in the field of questionnaire design, such as ensuring all scales are worded with labels Krosnick:1999wt or using a summating rating scale Spector:1992uj to address a specific topic of interest if people are to make mistakes in their response or answer in different ways at different times. Similar effects