Applied Ergonomics 74 (2019) 85–96

Contents lists available at ScienceDirect

Applied Ergonomics

journal homepage: www.elsevier.com/locate/apergo

# Extra-laboratorial usability tests: An empirical comparison of remote and classical field testing with lab testing

Juergen Sauer*, Andreas Sonderegger, Klaus Heyden, Jasmin Biller, Julia Klotz, Andreas Uebelbacher

*Department of Psychology, University of Fribourg, Fribourg, Switzerland*

ARTICLE INFO

*Keywords:*
Usability test
Remote testing
Field testing
Synchronous testing
Asynchronous testing

ABSTRACT

The present article examined the effects of using different extra-laboratorial testing procedures in usability testing. Three experiments were conducted using different artefacts (website, computer-simulated mobile phone, fully operational smartphone) to compare different methodological approaches in field testing (synchronous and asynchronous remote testing, classical field testing) to lab-based testing under different operational conditions (dual task demands, poor product usability). Typical outcome variables of usability testing were measured, including task completion time, click rate, perceived usability and workload. Overall, the results showed no differences between field and lab-based testing under favourable operational conditions. However, under difficult operational conditions (i.e. dual task demands, poor product usability) differences between field and lab-based testing emerged (corresponding to small and medium effect sizes). The findings showed a complex pattern of effects, suggesting that there was no general advantage of one testing procedure over another.

## 1. Introduction

### 1.1. Factors of influence in usability testing

Usability testing is a widely used method in the evaluation of consumer products. While its utility is not generally questioned, there are concerns that various factors may have an undue influence on the outcomes of usability testing (Bevan and Macleod, 1994). This refers to factors that are related to the properties of the product being tested, to the characteristics of the user, to the tasks being selected or to the testing environment being chosen (Lewis, 2006).

A prominent question in usability testing has been concerned with choosing the best location for conducting the test. This has generally centred on the question of whether lab or field testing would be the better option (e.g. Kjeldskov and Stage, 2004). This discussion may be considered part of the more general debate in ergonomics and psychology about the pros and cons of experimental research in the lab and in the field (e.g. Anderson et al., 1999; Dipboye and Flanagan, 1979). In the context of usability testing, the costs incurring from tests are also an important issue (e.g. Kaikkonen et al., 2005), with field testing generally being more costly than lab testing. Choosing the most appropriate testing method also needs to consider the influence of factors such as task demands and product properties. For instance, users in lab-based testing environments using a single task scenario were found to be more responsive to product information than if the typical usage scenario involves the simultaneous completion of more than one task (Sauer and Sonderegger, 2011).

These issues are also addressed in the Four-Factor Framework of Contextual Fidelity (Sauer et al., 2010), which proposes a more formal model of the factors identified by Lewis (2006). It proposes four chief factors that are expected to influence the outcomes of usability testing: user characteristics, product properties, task scenarios and testing environment (see Fig. 1). The last factor is the focus of the present study. The factor testing environment relates to physical features (e.g. size of laboratory) but also to social ones (e.g. observer presence). Both may influence the test outcomes, sometimes in rather complex ways. For example, the direction of the impact of observer presence on user performance may depend on the difficulty of the task scenario (i.e. observer presence increases performance for simple tasks and decreases it for difficult ones), as it would be predicted by social facilitation theory (Zajonc, 1965). Such observer effects may also be caused by the presence of experimenters, which might lead to a 'social desirability bias' on the part of the test participant by behaving and responding in a way the experimenter would appreciate.

* Corresponding author. Department of Psychology, University of Fribourg, Rue de Faucigny 2, 1700, Fribourg, Switzerland.
*E-mail address:* juergen.sauer@unifr.ch (J. Sauer).

https://doi.org/10.1016/j.apergo.2018.08.011
Received 16 February 2017; Received in revised form 19 April 2018; Accepted 11 August 2018

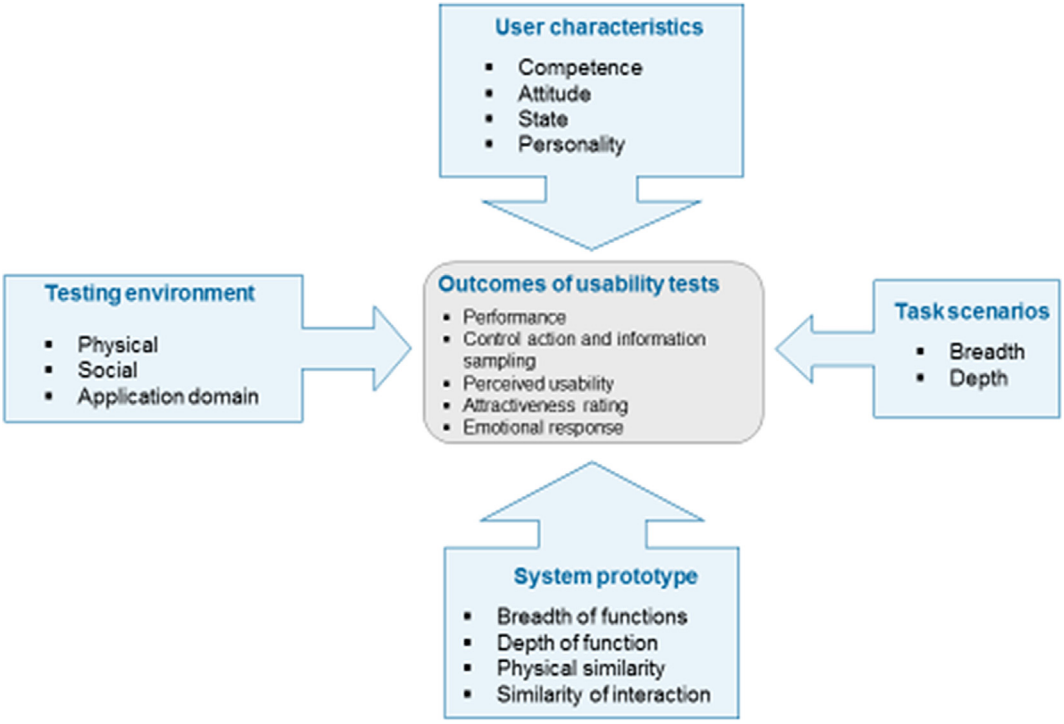


**Fig. 1.** Four-factor framework of contextual fidelity (adapted from Sauer et al., 2010).

### 1.2. Methodological issues in usability testing

#### 1.2.1. Testing outside the lab

In usability testing, one may distinguish between different extra-laboratorial approaches to testing: synchronous remote testing, asynchronous remote testing, and classical field testing. They differ from lab-based testing in different ways (see Table 1). For example, remote usability testing may be considered as a particular form of field testing because the test administrator and the user are in different locations during the product evaluation process (Dumas and Fox, 2012). A remote test may be conducted while test administrator and user can communicate via technical means in real time (e.g. telephone, chat) or with a time gap (e.g. email). These two forms are referred to as synchronous and asynchronous remote testing, respectively (e.g. Alghamdi et al., 2013). The testing methods listed in Table 1 differ from each other on a number of criteria, including experimental control, presence of test administrator, presence of environmental distractors, and chosen location. The rating of the criteria was made by the authors of the present article to provide a coarse assessment of the differences between testing methods. The ratings were based on the typical set-up in each testing method. The ratings have revealed considerable differences between the three forms of extra-laboratorial testing on most criteria (e.g. experimental control). However, the results also showed similarities between the lab setting and certain forms of extra-laboratorial testing (e.g. presence of test administrator). In summary, the advantages of lab-based testing are high levels of control over a range of factors such as variations in noise or interruptions during task completion. The different extra-laboratorial testing methods have their advantages in the form of higher levels of ecological validity. Participants in extra-laboratorial testing may complete the tasks in environments, in which they are typically done (e.g. public domain, at home). Furthermore, the two remote testing methods are characterised by lower experimenter presence, which can represent an advantage in certain set-ups (e.g. reduced tendency to fulfil social expectations of the experimenter).

#### 1.2.2. Remote testing

Remote usability testing provides a number of advantages over lab-based testing (e.g. Albert et al., 2010). From a practitioner's point of view, the advantages of remote testing over traditional lab-based testing refer to budgetary savings (e.g. travelling expenses for test participants, renting lab space, access to larger sample size) but also to methodological benefits (e.g. remote testing allows for the evaluation of a system with culturally diverse users). A further advantage would be that testing would be possible in a familiar and less artificial environment (e.g. if home-based testing is required). However, the downside of remote testing is reduced experimental control (e.g. users may be distracted during task completion).

A number of studies have examined the effects of remote usability testing (e.g. Andrzejczak and Liu, 2010; Madathil and Greenstein, 2011). Some work found differences between remote usability testing and lab-based testing, for example, in the form of longer task completion times (e.g. Andrzejczak and Liu, 2010; McFadden et al., 2002; Thompson et al., 2004) and lower task completion rates in remote testing (e.g. Ames and Brush, 2003). However, other work did not find any differences between testing methods (e.g. Madathil and Greenstein,

**Table 1**
Similarity of lab-based testing to different forms of extra-laboratorial testing.

| | Experimental control | Presence of test administrator | Control of environmental distractors | Chosen location |
|---|---|---|---|---|
| Synchronous remote testing | low | medium | medium | user-selected |
| Asynchronous remote testing | very low | weak | low | user-selected |
| Classical field testing | medium | strong | very low | experimenter-selected |
| Lab testing | high | strong | high | experimenter-selected |