

HCI Testing in Laboratory or Field Settings

Chuan-Hoo Tan^{1(✉)}, Austin Silva², Rich Lee³, Kanliang Wang⁴,
and Fiona Fui-Hoon Nah⁵

¹ Department of Information Systems, National University of Singapore, Singapore, Singapore
tancho@comp.nus.edu.sg

² Sandia National Laboratories, Albuquerque, NM, USA
aussilv@sandia.gov

³ IBM, Taipei, Taiwan
richchihlee@gmail.com

⁴ Department of Management Science and Engineering, Renmin University, Beijing, China
kanliang.wang@gmail.com

⁵ Missouri University of Science and Technology, Rolla, MO, USA
nahf@mst.edu

Abstract. This paper presents perspectives from both academia and practice on how an HCI testing is to be conducted and the deliberations that go into the testing. HCI testing can be conducted in closed-door laboratory or in a field setting. While there is an increased interest in field testing of an HCI artifact, there is always an enduring concern over how to administer a field testing given that the testers will have less control over the course of testing. In this paper, we cover HCI testing deliberation as well as the operational issues of field testing, and conclude the paper with an exemplary case of a large-scale field testing conducted in Taiwan.

Keywords: HCI testing · Laboratory · Field

1 Introduction

HCI testing is an integral part of an HCI design and development process. By having human subjects to physically test and use the artifact, HCI designers are able to validate the usability of an artifact and identify opportunities for further design improvements. Among many considerations, two prominent testing choices that a designer faces are (1) closed-door laboratory testing, or (2) field testing which is carried out in a setting that is more open [1]. While closed-door laboratory testing enables testers to have greater control over the course of testing, field testing narrows the gap between the intended usage of the artifact and its actual usage. Through the latter (i.e., actual usage), the HCI designers are able to better understand how an HCI artifact is utilized and valued by the target audience.

Whether to conduct a testing in laboratory or field setting is often based on the nature of the artifact itself. For example, it is viewed that a laboratory is suitable for testing a user interface and assessing an artifact's usability [2] while field is an appropriate setting for evaluating the acceptance of the artifact and its associated functionalities [3]. Two other considerations include (1) whether the artifact is revolutionary/disruptive in nature

and (2) company culture and propensity toward opening up the artifact for early field-testing, among others.

This panel takes a slightly different perspective by starting with the proposition that field testing is preferred due to the ability to more realistically test an HCI design. Building on this proposition, we ask: what are the operational considerations for field testing?

In the following sections, we will document the different perspectives taken by the panel members, which covers both the perspectives of academicians and practitioners on HCI testing. We conclude the paper with a large-scale field test conducted in Taiwan.

2 An Academic Perspective of HCI Testing

HCI testing often occurs at a later stage of HCI design and development. Hence, how to go about performing testing can be a delayed deliberation task. Fortunately, we are seeing more contemporary approaches such as agile development [4] and the design thinking principle [5], which further advocates the importance of pushing forward testing early in the artifact development. By having experience in conducting empirical studies in laboratory and field settings for research purposes, we are also able to see the challenges and issues confronting the setting to conduct HCI testing.

The first consideration factor for academicians to conduct field testing is the availability of the venue and the facilitating conditions. For example, a prototype that is still fairly distant from completion (i.e., far from ready to be commercialized) will be very difficult to test in the field setting as it could arouse emotional displeasure by the target audience. Also seeking companies' buy-in and facilitation will be challenging too. Thus, academicians often have to revert back to laboratory testing or conduct testing within the university boundary.

The second consideration factor is how "clean" the data is for us to attribute the causal relationship between an HCI artifact and user behavior. The laboratory setting restricts users from extraneous influences that could adversely affect his/her response to a given artifact. However, the field setting allows HCI testers to be vulnerable to environmental and situational conditions that may change during the course of testing. Thus, the research criterion of validity may be challenged.

The third consideration factor is whether the researchers are able to replicate or reconduct the testing with additional parameters should such needs arise. This is an enduring concern in the academic field due to the nature of the journal review cycle. In our discipline, papers could be in the review process for two to three years or even longer. During the review cycle, it is very plausible that reviewers would give comments and suggestions on how testing should have been conducted. In some cases, the researchers have to conduct another round of testing by taking the reviewers' suggestions into consideration, which raises the issue of whether, and to what extent, a field test could be conducted again. If the field testing involves industry partners, conducting additional rounds of testing may not be feasible.

Our view on HCI testing is: the decision to have the testing conducted in the laboratory or field setting may go beyond the artifact itself, as it could be a forced choice in

light of the constraints that academia may face. The next section covers a more detailed discussion on the operational issues relating to field testing.

3 A Practitioner Perspective of HCI Testing

HCI testing deliberations involve the understanding of whether the artifact is ready for testing, the operational considerations, the experimental design, and the analysis and interpretation of the data collected.

Is your artifact ready for field-testing? There is an increased recognition to launch products into the market quickly, i.e., the benefits of short time to market and first-mover advantage. Furthermore, it is always exciting to see new technologies enter the market. However, doing so requires the artifacts to be tested extensively in the laboratory before escalating to field testing. As a rule of thumb, when they are being tested in the field, it is ideal to first have a field pilot to ensure that the system is working as expected and the data is as clean as possible, since the cleanest data that one can ever hope to achieve is typically gathered in a highly controlled laboratory setting. If the technologies do not perform well in the laboratory testing, it is expected that things can only become worse in the field during live testing. You may not be able to assess or know in advance how well the data will be in the field.

How to conduct the testing? To truly understand the strengths and limitations of your developed system or procedure, it may be necessary to observe and assess users as they interact with the system in the actual environment that they would naturally perform. Unfortunately, it is nearly impossible to collect that data seamlessly and without noise. The big question becomes: What are you willing to trade in for the realism of the task in order to get the data you want? First, the researcher must understand the task and if there are elements that could be best computerized and simulated. If there are tasks that get automated from physical and observable tasks, the experiment may drift further away from the true environment. But if the subjects were to perform in the actual environment with the normal ambient distractions and interactions, it may be relevant to the overall hypothesis and only give up some fidelity for the sake of clean data.

How to design the testing? To help assess the importance in needing to go into the field for data collection, one must first decide what the main objective is. If the main objective is feedback on an interface, it may not need to be conducted in the field. However, if the key objective is to measure cognitive load via eye tracking while performing tasks in the operational environment, then there may arise a need to conduct the experiment in the field. Sometimes, researchers will add capabilities and tools on top of an existing study to try to collect more data in order to answer a secondary hypothesis. A study that measures behavioral performance could have an eye-tracking component added to it, but the data may be noisy if it was not designed to be an eye tracking study (i.e., the subject is required by the primary task to move around a lot where the eye tracker may not be able to focus on the subject). The field presents numerous potentials for sources of noise in the data if relying on biometric data such as

eye tracking or EEG devices. However, we also caution that creating lab-like mitigations of these sources of noise (like chinrests) can create even more distractions and complications for the users.

Can you handle the data collected? While field studies may hold more promise for operational validity, there may be an introduction of additional noises. With the introduction of noises, one may now have to collect more data than needed in the laboratory. With this new requirement, you now can start to run into big data problems in which you are potentially collecting gigabytes of data across hundreds of subjects. In order to be able to handle this inflow of data after collection, it is important to develop a well-documented data analysis procedure across the team on how the data will be stored, accessed, and updated. If a researcher wants to test something that could affect the policy and practice of an organization, it is important to ensure that the large dataset can be analyzed and validated with ease. However, if the study is a review of usability of new technology where the study involved only a few subjects, it may not present a big data problem. When out in the field, we suggest performing quick data checks on the new data to ensure that the data is being properly collected, or else the remainder of the data collection may need to be postponed. It would also be helpful to thoroughly inspect the data after the first few experimental sessions before continuing with the rest of the sessions.

Can you manage your team? There are opportunities for a research team to specialize in or master one of the domains – lab or field. With a larger team, you can appoint a master and apprentice model in which one team member is the main operator supported by a secondary operator who is approximately equal in terms of knowledge in case the main person is unavailable or cannot conduct the experiment. When conducting testing in the field, it is imperative that the person collecting the data is not only knowledgeable in the data collection process, software, and objectives, but is also aware of the technical aspects of the equipment in case something needs to be modified in the software or hardware. If a problem arises and the data collection proctor is not familiar with the system, it could result in having to cancel the remaining session to return back to the lab for changes. Such cases could result in a total loss of data or collection of the wrong data.

4 A Field Testing Example

To understand HCI testing, it is best to understand the complexity of conducting HCI testing in the field. In what follows, I present an example of how a large-scale field testing was conducted in Taiwan. The methodology is the scenario-driven testing approach [6], which is applied to a complicated and heavy-stressed service system — the Taipei Massive Rapid Transit (Subway) [see Fig. 1]. The testing objective is to validate the quality of design and the potential weakest links of the service disruption to determine if the service systems will meet the users' expectations.



Fig. 1. Taipei mass rapid transit

Several key successful factors are:

1. The explicit functionalities, such as sensing the Smart Cards, debiting the toll, etc.;
2. The implicit expectations, such as the response time of the sensing devices, the time period for the sensing device to take the next customer;
3. The reliability, the Mean Time between Failure (MTBF) and the Mean Time to Repair (MTTR) from the Fatigue Analysis report; and
4. The exceptional procedures (in case of service failure), the seamless manual process or the compensating procedures.

It is essential to establish a clearly defined Quality Function Diagram (QFD) created by the stakeholders and the design team (Fig. 2) before beginning the validation process.

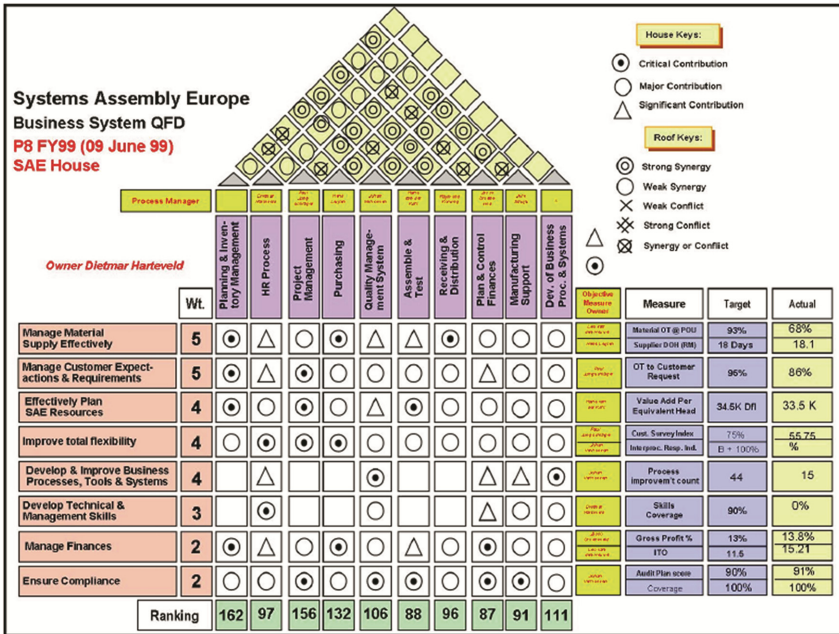


Fig. 2. An example of QFD for transportation

The diagram presents the specific criterion related to the testing details, and serves as the basis for the Work Breakdown Structures (WBS) in the testing project planning.

The testing plan includes:

1. The Integration Testing—this can be done in the lab to record the results of the performance; a service modeling and simulation tool (Fig. 3) will be helpful to find the potential weakest links of the service system through the Stress Impact Analysis (Fig. 4) according to the performance result records;
2. The Field Testing—this is synonymous to Trial Run, based on the predefined scenarios, choosing few test sites, having the actors physically experiencing the system and observing how various payloads impact on it; and
3. The Sustaining Testing—keep monitoring and re-coding the performance results against the stimuli from the real service sites (Fig. 5), a Quality Attribute perspective, use these data as the evidence of continuous improvement.

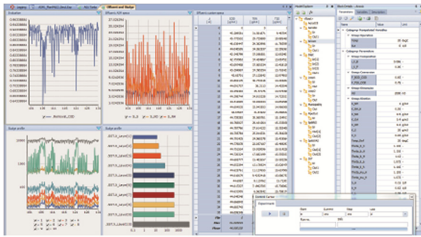


Fig. 3. An example of service modeling & simulation tool

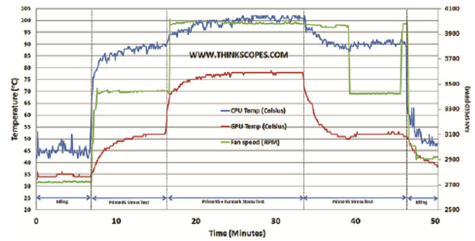


Fig. 4. An example of service system stress impact analysis

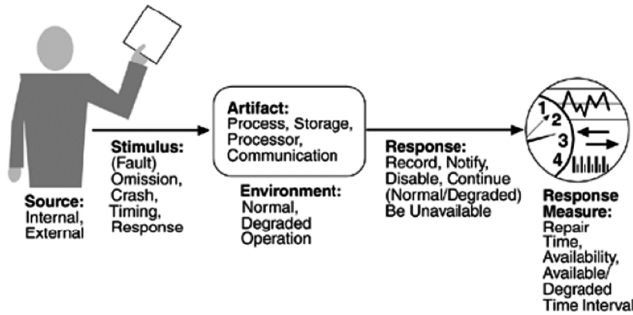


Fig. 5. Fundamental quality attribute model

Based on the aforementioned background information, the suggested panel outlines will also cover the following topics and the associated issues:

1. Service Scenario Design,
2. Service Reliability,
3. Service Testing Types and Procedures,
4. Service System Design,

5. Testing Project Management,
6. Testing Environment, and
7. Simulation and Dynamic Analysis.

5 Concluding Comments

HCI testing continues to remain an integral part of the technology design and development process. We hope through this article and the panel discussion, we are able to generate more insightful and novel ideas of how to go about planning and conducting an HCI testing expedition. There are trade-offs in conducting laboratory and field testing, and both are essential before rolling out a commercial product. The developers or technology-based firms may utilize laboratory testing at earlier stages of product development and conduct more field testing at the later stages. Various iterations of laboratory and field testing may be necessary to ensure that an artifact fulfills its functionality in a realistic setting.

Recognizing the value of both laboratory and field testing in the design and development process of a technology is a key contributor to its commercial success.

References

1. Gerber, A.S., Green, D.P.: *Field Experiments: Design, Analysis, and Interpretation*, 1st edn. W.W. Norton & Company, New York (2012)
2. Hallahan, K.: Improving public relations web sites through usability research. *Public Relat. Rev.* **27**(2), 223–239 (2001)
3. Kaasinen, E.: User Acceptance of location-aware mobile guides based on seven field studies. *Behav. Inf. Technol.* **24**(1), 37–49 (2005)
4. Highsmith, J., Cockburn, A.: Agile software development: the business of innovation. *Computer* **34**(9), 120–127 (2001)
5. Zimmerman, J., Forlizzi, J., Evenson, S.: Research through design as a method for interaction design research in HCI. In: *CHI 2007 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 493–502 (2007)
6. Sivashanmugam, K., Lin, D., Palanisamy, S.: Scenario driven testing. In: *8th International Conference on Information Technology: New Generations (ITNG)*, pp. 299–303 (2011)