Available online at www.sciencedirect.com
SCIENCE DIRECT®
Interacting with Computers 17 (2005) 773–786
Interacting with Computers
www.elsevier.com/locate/intcom

Cross-user analysis: Benefits of skill level comparison in usability testing

Laura Faulkner*, David Wick[1]

Applied Research Laboratories, The University of Texas at Austin, P.O. Box 8029, Austin, TX 78713-8029, USA

Received 15 April 2004; revised 19 April 2005; accepted 23 April 2005
Available online 5 July 2005

Abstract

This study presents a cross-user usability test approach and analysis technique that extends beyond merely identifying the existence of a usability problem to introducing an empirical basis for identifying the type of usability problem that exists. For experimental purposes, 60 users were tested with three levels of user-competency determined by experience in using: (1) computers, and (2) the tested application. Applying the Tukey honestly significant difference (HSD) test to each test element provided statistical comparison between different experience levels. Analysis results between experience levels suggested which levels encountered usability problems. The authors demonstrate that statistical calculations of cross-user data can render empirical support for categorizing usability problems.

Keywords: Usability testing; Users; HCI methodology; Usability research; Empirical method

1. Introduction

Often in traditional usability testing, novice testing focuses on learnability while expert testing focuses on optimal use. This difference in focus means different tasks are used and different tests are performed, making them unsuitable for comparison across user levels. This split between novice and expert user testing is placed in historical context by Jakob Nielsen (2000), who traces several shifts in the focus of usability testing. For example, in

* Corresponding author. Tel.: +1 512 835 3328; fax: +1 512 835 3100.
E-mail addresses: laura@arlut.utexas.edu (L. Faulkner), dwick@arlut.utexas.edu (D. Wick).
[1] Tel.: +1 512 835 3646.

0953-5438/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.intcom.2005.04.004

the 1970's, when computers were the domain of an elite group, usability was focused on these experts. However, the widespread use of computers by average people in 1980s led to a shift toward novice users, with primary focus on learnability. Referring to the change in focus as a pendulum, Nielsen describes two further shifts one in toward late 1980s to the early 1990s the performance of expert users, and another in the early 1990s arising from web usability issues and focusing on novice users. Nielsen concludes that the current focus on novice testing will start to swing back towards expert testing, and asserts the importance of both forms of testing in providing substantial evidence for usability.

The separate testing of either novice or expert, but not both on the same tests, is supported by examining descriptions of usability tests on various products. One such description discussed the existence of expert users, and implied having received feedback from those experts, but performed testing only on novice users. Further, the description highlighted this marked view of separation by describing information-gathering approaches to fulfill the need 'to understand how each group approaches' the application, with 'each group' specifically being composed of either novices or experts (Benson, 2001; see also Straub, 2003 and Fu et al., 2002). Further evidence of this split is found by examining usability test guidelines and recommendations from popular usability handbooks, such as Barnum (2002), Hackos and Redish (1998), Nielsen (1993), and Rubin (1994), which confirm the usual approach of testing novices and experts for different reasons. In the case of determining whether a product is sufficiently simple and intuitive for beginner use, or its level of learnability, testing novice users is essential; however, data from all levels of users would be needed to gain a picture of its full range of usability.

These time-honored approaches continue to render valuable usability information. As the field matures, however, it is appropriate for its methods to evolve in sophistication and accuracy. Hackos and Redish (1998) do provide a deeper level of granularity to the field by identifying four levels of users, namely, 'Novices', 'Advanced Beginners', 'Competent Performers', and 'Expert Performers', with sensitive, detailed descriptions of each. This provides more highly developed representations of user populations; however, in their descriptions, they continue to maintain the disparate nature of information that can be gained from each. In another extension of usability approaches, researchers from the Hong Kong University of Science and Technology tested novice users against experienced users, having created the 'experienced' users by providing them pre-test training on application tested. The research design and comparative results suggest that testing and analyzing the performance of novice users against experienced users in the same test provides an additional layer of information that testing the two separately does not provide (Goonetilleke et al., 2001).

To further extend the depth and maturity of usability approaches, the authors sought to develop and perform identical usability tests with users of different experience levels, from novice to expert. Such a method allows for comparison between user levels revealing an additional layer of information relevant to the identification of usability problems, as evidenced in the work discussed below.

The potentially critical importance of such information was highlighted by an occurrence in the medical field as reported by Leveson and Turner (1993), in a discussion of Therac-25, a nuclear accelerator designed to administer radiation treatments to cancer