Cognitively Informed Intelligent Interfaces:

Systems Design and Development

Eshaa M. Alkhalifa
University of Bahrain, Bahrain

Khulood Gaid
Royal University for Women, Bahrain

Information Science
REFERENCE

Managing Director: Lindsay Johnston
Senior Editorial Director: Heather A. Probst
Book Production Manager: Sean Woznicki
Development Manager: Joel Gamon
Development Editor: Hannah Abelbeck
Acquisitions Editor: Erika Gallagher
Typesetter: Lisandro Gonzalez
Cover Design: Nick Newcomer

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com

Library of Congress Cataloging-in-Publication Data

Cognitively informed intelligent interfaces: systems design and development / Eshaa M. Alkhalifa and Khulood Gaid, editors.
p. cm.
Includes bibliographical references and index.
Summary: “This book analyzes well-grounded findings and recent insights on human perception and cognitive abilities and how these findings can and should impact the development and design of applications through the use of intelligent interfaces”-- Provided by publisher.
ISBN 978-1-4666-1628-8 (hardcover) -- ISBN 978-1-4666-1629-5 (ebook) -- ISBN 978-1-4666-1630-1 (print & perpetual access) 1. User interfaces (Computer systems) 2. Artificial intelligence. 3. Cognition. I. Alkhalifa, Eshaa M., 1966- II. Gaid, Khulood, 1989-
QA76.9.U83C58 2012
005.4’37--dc23
2012000166

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

Chapter 15
The Bootstrap Discovery Behaviour Model:
Why Five Users are Not Enough to Test User Experience

Simone Borsci
Brunel University, UK

Stefano Federici
University of Perugia, Italy

Maria Laura Mele
Sapienza University of Rome, Italy

Domenico Polimeno
University of Perugia, Italy

Alessandro Londei
Sapienza University of Rome, Italy

ABSTRACT

The chapter focuses on the Bootstrap statistical technique for assigning measures of accuracy to sample estimates, here adopted for the first time to obtain an effective and efficient interaction evaluation. After introducing and discussing the classic debate on p value (i.e., the discovery detection rate) about estimation problems, the authors present the most used model for the estimation of the number of participants needed for an evaluation test, namely the Return On Investment model (ROI). Since the ROI model endorses a monodimensional and economical perspective in which an evaluation process, composed of only an expert technique, is sufficient to identify all the interaction problems—without distinguishing real problems (i.e., identified both experts and users) and false problems (i.e., identified only by experts)—they propose the new Bootstrap Discovery Behaviour (BDB) estimation model. Findings highlight the BDB as a functional technique favouring practitioners to optimize the number of participants needed for an interaction evaluation. Finally, three experiments show the application of the BDB model to create experimental sample sizes to test user experience of people with and without disabilities.

DOI: 10.4018/978-1-4666-1628-8.ch015

INTRODUCTION

The ROI model, which was proposed in 1993 by Nielsen and Landauer, shows that, generally, the least number of users required for a usability test ranges from three to five. This model is an asymptotic test which allows practitioners to estimate the number of users needed through the following formula:

$$\text{Found}(i)=N\,[1-(1-p)^{i}] \quad (1)$$

In (1), the *N* value corresponds to the total number of problems in the interface, the *p* value is defined by Nielsen and Landauer (1993) as "the probability of finding the average usability problem when running a single average subject test" (i.e., discovery detection rate), and the *i* value corresponds to the number of users. For instance, by applying formula (1), practitioners can estimate whether five users are sufficient for obtaining a reliable assessment and, if not, how many users (*N*) are needed in order to increase the percentage of usability problems. Nielsen, starting from the results obtained by many applications of the ROI model, suggests that the practitioners, in order to test different categories of users, have to divide users into multiple groups composed as follows (Nielsen, 2000):

- 5 subjects of a category if testing 1 group of users;
- 3-4 subjects from each category if testing 2 groups of users;
- 3 users from each category if testing three or more groups of users.

The value "p" (see formula 1) may be considered an index for assessing the effectiveness and efficiency of an Evaluation Method (EM). As some international studies (Lewis, 1994; Nielsen, 2000; Nielsen & Mack, 1994; Virzi, 1990, 1992; Wright & Monk, 1991) have shown, a sample size of five participants is sufficient to find approximately 80% of the usability problems in a system when the individual detection rate (p) is at least .30. The value of 30% was derived through Monte Carlo (MC) resampling of multiple evaluators, and could also be estimated using the full matrix of problems as discovered by independent evaluators (Lewis, 2001).

However, as Nielsen and Landauer (1993, p. 209) underline when discussing their model, the discoverability rate (p) for any given usability test depends on at least seven main factors:

- The properties of the system and its interface;
- The stage of the usability lifecycle;
- The type and quality of the methodology used to conduct the test;
- The specific tasks selected;
- The match between the test and the context of real world usage;
- The representativeness of the test participants;
- The skill of the evaluator.

As Borsci, Londei, and Federici (2011) claim, many studies underline that these factors have an effect on the evaluation of the interaction between system and user that the ROI model is not able to estimate (Caulton, 2001; Hertzum & Jacobsen, 2003; Lewis, 1994, 2006; Schmettow, 2008; Spool & Schroeder, 2001). In this sense, the ROI model cannot guarantee the reliability of the evaluation results obtained by the first five participants.

One of the most relevant problems that Borsci et al. (2011) underline is that the ROI model starts with a "one evaluator" condition and not at zero condition; this means that the characteristics of the system are considered only as the differences between problems found by the first evaluators. Nielsen and Mack (1994) pointed out that the first evaluator (a user or an expert) generally finds 30% of the problems, because these problems are generally the most evident. The subsequent evaluators usually find a smaller percentage of

new problems, simply because the most evident ones have already been detected by the first evaluator. The number of evident problems is determined empirically and it varies because it is dependent on the evaluator's skills, which, as we have already stated, are a factor that this model does not consider.

A serious limitation of the ROI model is that it happened to be working with products for which the value of *p* across evaluators/users was about .3, but as J. R. Lewis (1994) showed, it is possible for the composite value of *p* to be much lower than .3. In order to assess the completeness of a problem-discovery usability study, the practitioner(s) running the study must have some idea of the value of *p*, which differs from study to study as a function of the properties of the system, the interface lifecycle stage, the methodologies selected for the evaluation, and the skill of the evaluators/users, ergo it is not necessarily .3 (30%).

FROM ECONOMIC EVALUATION TO EFFECTIVENESS AND EFFICIENCY OF THE EVALUATION METHODS

Albeit the ROI is a powerful model to obtain an economic index, it does not guarantee the efficiency and the effectiveness of the evaluation. Indeed the assumption of this model is that all problems identified in the evaluation are real problems. This assumption is true just when we consider separately all the problems found in a test.

However, when the problems identified by experts and users are matched, three different kinds of problems can be identified: false problems, which are detected only by the expert analysis; missed real problems, which are problems identified by the users during the interaction that were not detected by the experts; and real problems, which are problems identified by both the user-based and the expert-based analyses.

Our idea is that the problem with the ROI model is not only the limit in the estimation of the *p* value, but we want to claim that the most important deficiency of the ROI model is the economical perspective mediated by it. In fact under this perspective, in which all the found problems are considered as real *per se*, the *p* value estimation problems cannot be solved.

The limits of this perspective can be summarized as follow:

1. *It is not related to a standard definition of effectiveness and efficiency*: Nielsen and Landauer created the ROI model five years before the definition of effectiveness and efficiency provided by the International Standards Organization (ISO) 9241-11 (1998). In this sense, they did not refer to the effectiveness and the efficiency as dimensions of usability (multi-dimensional perspective), but they endorsed an economical perspective (monodimensional perspective): implicitly, in the ROI model, the effectiveness is considered as the amount of problems found by a technique, while the efficiency as the amount of evaluation costs.
2. *It is focused on a quantitative point of view*: The more efficient an EM is the less it costs. The costs of an EM are mainly calculated on the number of the participants in the evaluations, because more participants require more time for the evaluation, money for the participants' fees, etc. Therefore, an EM is more efficient when it employs few participants. Moreover, an EM is considered effective when with the smallest number of participants it finds the largest number of problems (i.e., more than 80% of the probable problems in the interface). Nevertheless, as we already claimed, the model does not consider which kind of problems the EMs detected.

In order to avoid the first limit of the ROI model perspective, we try to provide a definition of

evaluation technique effectiveness and efficiency that accomplishes the ISO standard.

The ISO 9241-11(1998) standard defines the effectiveness and the efficiency of a system as follows:

- *Effectiveness* is the accuracy and completeness with which users achieve specified goals;
- *Efficiency* is the amount of resources expended in relation to the accuracy and completeness with which users achieve goals.

Following this statement, and applying it to an EM technique, we wonder: how can we define the effectiveness and the efficiency of an EM? We propose these following definitions of effectiveness and efficiency by applying in the context of EMs the aforementioned content of the ISO-9241-11.

- *Effectiveness of an EM*: Considering the effectiveness as "the accuracy and completeness with which a user achieves specified goals," we define the Effectiveness of an EM as the ability to estimate which real problems are present in the evaluated system (i.e., which problems do not allow an effective interaction);
- *Efficiency of an EM*: Considering the efficiency as "the amount of resources expended in relation to the accuracy and completeness with which users achieve goals," we define the Efficiency of an EM as the amount of all the costs of the evaluation meant as the time spent by user and/or expert.

Since the aim of the evaluation process is the identification of real problems, the effectiveness of a technique should be related to the quantity of "real" problems found (and not just to the number of all problems); while the efficiency should be linked to the cost of the evaluation—i.e., the number of participants and the time spent for the analysis.

As a second step, in order to overcome the quantitative point of view of the ROI model perspective, according to our definition of efficiency and effectiveness of an EM, we provide a new perspective on the identified problem in evaluation by distinguishing the quantity and the quality of their nature.

Nielsen and Landauer (1993) consider heuristic evaluations (i.e., expert-based analysis) to be always more powerful than any user-based evaluation: in fact, heuristic evaluation provides a large amount of problems' identification (i.e., effectiveness according to ROI) with a lower cost than the users' tests (i.e., efficiency according to ROI). Now, it is clear that, while we could endorse the idea that a high efficiency equals a low cost evaluation, we cannot endorse the idea that a high effectiveness equals a high amount of problems found, without distinguishing between real and false problems. In our opinion, an effective evaluation is that one that detects the highest number of just "real" problems.

Using a metaphor to explain the difference between the ROI model and our perspective, we can say that:

- *Following the ROI model*: A fisher (i.e., the evaluator), in order to have an effective and efficient fishing process, needs to use the largest fishing net possible (i.e., the EM); in this way, in fact, s/he would be able to obtain a low-cost process (efficiency) and a high number of fishes caught (effectiveness).
- *Following our idea*: A fisher, in order to have an effective and efficient fishing process, not only needs to use a kind of fishing net (i.e., EM) able to guarantee a low-cost process (efficiency), but s/he also needs to catch a certain kind of fishes: i) fishes that can be sold as edible and ii) fishes that can

be fished without breaking the law. Our idea is that an effective fishing net is not the largest one but the one able to catch those fishes that the fisher will not have to throw overboard. Out of our example: an effective EM must find the highest number of "real" problems, minimizing the identification of "not real" ones.

Summarizing, in this section we have divided the interaction problems into false and reals. The real ones are the problems identified during the interaction by both experts and users, while the false problems are identified only by experts. We have stressed that the ROI model does not take into consideration our previous distinction, because it endorses a monodimensional and economical perspective in which an evaluation process, composed of only an expert technique, is sufficient to identify all the interaction problems. As we discussed above, today the ROI model perspective is overcome by a multidimensional perspective (ISO 9241-11, 1998) in which the evaluation process aims at identifying problems by matching results from experts' and users' techniques (i.e., computing only the real problems). In order to extend both the ROI perspective and its mathematical model, we applied a bootstrap statistical technique for assigning measures of accuracy to sample estimates. By following this aim we create an alternative model to the ROI, based on the probabilistic behaviour in the evaluation, the Bootstrap Discovery Behaviour (BDB) model. The BDB, by considering more factors in the *p* value estimation and endorsing a multidimensional perspective of the evaluation, may be applied for defining the sample size of both the users and the experts needed to identify the real problems of interaction by a matching of all experienced problems.

The BDB Model

The term "bootstrapping," defined by Efron (1979), is an allusion to the expression "pulling oneself up by one's bootstraps"—in this case, using the sample data as a population from which repeated samples are drawn (Fox, 2002). The present bootstrapping approach moves from the assumption that discovering new problems should be the main goal of both users' and experts' evaluations as well as expressed in Formula (1) by Nielsen and Landauer (1993).

Given a generic problem *x*, the probability that a subject will find *x* is *p(x)*. If two subjects (experts or users) navigate the same interface, the probability that *at least* one of them will detect the problem *x* is:

$$p\,(x_1 \bigvee x_2) \tag{2}$$

In (2), where x_1 and x_2 represent the problem x detected by subjects 1 and 2, OR is the logic operator. According to De Morgan's law (Goodstein, 1963), (2) is equivalent to:

$$p\,[\neg\,(\neg\, x_1 \bigwedge \neg\, x_2)] \tag{3}$$

Equation (3) expresses the probability of "the degree to which it is false that none of the subjects find anything" (the logic operator for negation). So (3) can be rewritten as:

$$p\,(\neg\, x) = 1 - p(x)$$

Since the probabilities of different subjects finding a specific problem are mutually independent, Equation (3) can be written as:

$$p\,[\neg\,(\neg\, x_1 \bigwedge \neg\, x_2)\,] = 1 - [1-p(x_1)\,]*[1-p(x_2)\,] \tag{4}$$

Following Caulton's homogeneity assumption (2001) that all subjects have the same probability (p) of finding the problem x, then (4) can also be expressed as:

$$p\,(x_1 \bigvee x_2) = 1 - (1-p)^2 \quad (5)$$

Of course, we can extend this case to a generic number of evaluators L:

$$p\,(x_1 \bigvee x_2 \bigvee \ldots x_L) = 1 - (1-p)^L \quad (6)$$

Equation 6 expresses the probability that, in a sample composed of L evaluators, at least one of them will identify the problem x.

According to Nielsen and Landauer (1993), given N problems in an interface, the probability of any problem being detected by any evaluator can be considered constant ($p(x) = p$). Then, the mean number of problems detected by L evaluators is:

$$F(L) = N\,[1 - (1-p)^L] \quad (7)$$

Leading to the same model presented by Nielsen (Equation 1), in (7), in order to estimate $p(x)$ we adopted the bootstrap model, avoiding estimation merely based on the addition of detected problems. This kind of estimation could in fact be invalidated by the small size of the analysed samples or by the differences in the subjects' probabilities of problem detections.

As its first feature, BDB model is able to take into account the probabilistic individual differences in problem identification. The second feature of the BDB is that it considers the evaluated interfaces as an object per se. The interfaces are considered different not in terms of the number of problems found by the first evaluator (evaluation condition), but different as objects (zero condition) estimating the probabilistic number of evident problems that all the evaluators can detect by testing the interface. The third feature of the BDB model is that, in order to calculate the number of evaluators needed for the evaluation, it considers the representativeness of the sample (as regards the population of all the possible evaluation behaviours of the participants).

Our idea is that the BDB should be able to grant a more reliable estimation of the probability of identifying a problem than the ROI model, particularly when a practitioner has to carry out a User eXperience (UX) evaluation test with a mixed sample of disabled and not disabled users.

The Mixed Sample User eXperience Evaluation: BDB Model Application

The new concept of UX enlarged the role played by users within the interaction evaluation process. Indeed, as Garret underlines, the UX of the system "is about how it works on the outside, where a person comes into contact with it and has to work with it" (Garrett, 2003). At the same time, the ISO 9241-210 defines UX as "a person's perceptions and responses that result from the use or anticipated use of a product, system or service. [...] User experience is a consequence of the presentation, functionality, system performance, interactive behaviour, and assistive capabilities of an interactive system, both hardware and software. [...] It is also a consequence of the user's prior experiences, attitudes, skills, habits and personality" (2010). Following this definition, the UX concept results in an extent of the usability itself, by taking into consideration the users' experiences, attitudes, skills, and personality. According to this new framework, the analysis of the interaction of disabled users becomes a priority in order to allow practitioners to compose a mixed panel of users that guarantees the reliability of the set of data obtained during the interface evaluation process.

Although the literature on the interaction between disabled users and technology is very wide; indeed, the studies on HCI rarely take into account the UX of persons with intellectual disabilities (Luckasson, et al., 2002; Schalock & Luckasson, 2004), and, when it happens, they mostly focus on either analysing accessibility issues (Bohman

& Anderson, 2005) or describing how to improve design of the technology (Fairweather & Trewin, 2010). Moreover, these kind of studies are mostly centered on identifying the advantages of new communication technologies for persons with intellectual disabilities (Feng, Lazar, Kumin, & Ozok, 2008, 2010).

Following the distinction made by Hartson and colleagues (Hartson, Andre, & Williges, 2001), have identified two approaches endorsed in HCI evaluation studies: the first one is the summative evaluation approach by which the evaluation of the interface is conducted for assess the efficacy of the final design or to compare competing design alternatives in terms of usability; the second approach is the formative evaluation one by which the evaluation is focused on usability problems that need to be solved during the prototype design stage before a final design can be accepted for release.

We use this distinction for classify the literature on the interaction between disabled users and technology. We classify under the 'Summative Oriented' (SO) approach all the studies which aim at improving the system accessibility and analysing disabled users' skills and behaviour while performing a product. These studies (Bohman & Anderson, 2005; Fairweather & Trewin, 2010; Feng, et al., 2008, 2010) endorse an approach that considers the disabled users only as 'customers' of products, instead of users in interaction.

Conversely, we classify the studies focussed on a widening disabled users' participation to systems' creation by a User-Centred Design (UCD) perspective (Norman, 1988) under the 'Participative and Formative Oriented' (PFO) approach (Federici & Borsci, 2010; Federici, Borsci, & Mele, 2010; Federici, Borsci, & Stamerra, 2010; Federici, et al., 2005; Feng, et al., 2008, 2010; Lewis, 2005).

The main purpose of both SO and PFO approaches is to promote and diffuse the *Design for All*, according to the Stephanidis' definition: "Universal Design in information technology and telecommunications products should not be conceived as an effort to advance a single solution for everybody, but as a user-centered approach to providing products that can automatically address the possible range of human abilities, skills, requirements and preferences" (Stephanidis, 2001).

Although both the SO and PFO approaches have the aim to analyse the "match" between the technologies and the users' needs, by following the Design for All philosophy, the SO results as the most used approach by the researchers, even though it reduces the evaluation of interaction to the mere analysis of the system features.

On the other hand the PFO approach—which endorses the motto 'Nothing about us, without us' (Charlton, 1998), that we may reinterpret here as 'nothing is *for all*, without us'—aims to improve the UX of interaction by a UCD process of design and re-design which allow to extend the system's features taking into account the needs of all kind of users. While the SO approach not fully accomplishes the Design for All's goals, because it not considers the users' needs, the PFO approach aims to overcome the SO assessment process by involving disabled users ever since the first phases of both the design and the evaluation processes. In this sense, by following the PFO perspective, the only way to spread the Design for All philosophy is to develop methods and techniques that equally involve disabled and not disabled users in each step of the assessment process, starting from the recruitment of subjects until the final product evaluation. Indeed, one of the most important problems in the UX studies is not only how to test disabled users in order to analyse their interaction experience, but also how to test them by collecting data that can be compared to not-disabled users.

In the next sections after the presentation of the BDB model, we focused on the sample size of participants who must be involved into the evaluation process of a prototype during its development. Specifically, in order to identify the number needed to get the less expensive and the most efficient mixed sample of users for discovering at least the 80% of usability problems, a

usability evaluation has been carried out involving users with intellectual disabilities, blind and not-disabled users. We present three experimental application of the BDB model, in order to show how large a mixed sample has to be created by a practitioner to obtain a reliable (i.e., efficient and effective) UX evaluation.

EXPERIMENTAL APPLICATION OF THE BDB MODEL

We compare the BDB and the ROI model discovery rate in a UX test conducted with three experimental groups of users—not-disabled, blind and with Down Syndrome (DS)—in order to estimate the number of users (with and without disability) needed to compose sample that can allow to discover the number of problem as lager as possible (i.e., efficacy) at the minimum costs (i.e., efficiency). Three experimental sessions are here presented, concerning the analysis of three different web interfaces. Each experiment involved a control group composed by not disabled and disabled users, with either visual or intellectual disabilities.

Methods and Tools

We compared the number of users needed in order to identify the 80% of problems by applying both the ROI model and the BDB model (for the source code of the BDB model, see Appendix 1). The analysis was carried out through the IBM® PAWS Statistics18 software and the Matlab software.

Apparatus

For each experiment the apparatus used during the experimental setting was set up as follows: A internet connection ADSL 4 MB; an internet Explorer 8 browser; a PC AMD Athlon 64 (3,200 MHz); a Philips 190S LCD 19" monitor; a Jaws® screen reader version 12 (for both the experiment 1 and 2); a CamStudio version 20 screen recorder; two amplifiers; an audio recorder Digital Zoom h2; a Nikon L2 digital camera; a Stopwatch; and a desk bell.

Experimental Websites

In all the three experiments each user were involved in four scenario-based analyses, which have been created to allow users to recruit information with at least four actions (i.e. mouse click). The tests of the web interfaces for all the experiments were conducted between April and June 2009.

- The experiment 1 consisted on the evaluation of the Italian National Social Service website (www.serviziocivile.it) conducted by a sample of blind users and a sample of sighted users.
- The experiment 2 consisted on the evaluation of a prototype of a sonified visual search engines WhatsOnWeb (WoW)—created at the University of Perugia by Department of Electronic and Information Engineering (DIEI) (Di Giacomo, Didimo, Grilli, & Liotta, 2007; Di Giacomo, Didimo, Grilli, Liotta, & Palladino, 2008) with the collaboration of the CognitiveLab group (www.cognitivelab.it) (Mele, Federici, Borsci, & Liotta, 2010)—conducted by a sample of blind and sighted users. Differently from other common search engines, e.g. Google or Yahoo, WoW has been implemented by using sophisticated graph visualisation algorithms on semantically clustered data: in this way, the indexed information is conveyed by means of a visuospatial data representation allowing to overcome the efficiency limitations of the top-down linear output given by the most common search engines, which generally use a flat representation of the indexed dataset (Federici, Borsci, & Stamerra, 2010).

- The experiment 3 consisted on the evaluation of the website of the Public Transportation of Rome (http://www.atac.roma.it/) conducted by a sample of users with Down Syndrome and a sample of users without any intellectual disability.

Measures

- *Verbal protocols:* In each of the three experiments the Partial Concurrent Thinking Aloud (PCTA) verbal protocol has been used to analyse the interaction of disabled users. The PCTA is a new but consolidated technique (Federici, Borsci, & Mele, 2010; Federici, Borsci, Mele, & Stamerra, 2010; Borsci, Kurosu, Federici, & Mele, 2011) that respects the properties of classic verbal protocols and at the same time overcomes the structural interferences and the limits of the concurrent and retrospective protocols when used during the screen reader navigation. Furthermore, the classic Thinking Aloud (TA) (Ericsson & Simon, 1980) method was used with not-disabled subjects in all the three experiments.
- *Discovery rate measures:* We apply the BDB model with 5000 bootstrap steps. Actually, the BDB approach allows the behaviour of the whole population, the representativeness of the sample data (i.e. the problems found expressed by *p* value) and the different properties of the interface to be taken into account. Since it respects the assumption of the ROI and the results obtained by a Montecarlo resampling, this model opens the possibility of considering both the properties of the interface and the representativeness of data, granting to practitioners a representative evaluation of the interface.
- *The Mini Mental State Exam (MMSE)*: is used for the clinical and neuropsychological evaluation (Folstein, Folstein, & McHugh, 1975) with persons with intellectual disabilities to analyse the degree of disability. The test is composed by 30 items with a scoring point from 0 to 30. The evaluation of the degree of intellectual disability follows these criteria: 1) a Scoring between 25 to 30 points indicates any disability; 2) a Scoring between 21 to 24 points indicates a mild disability degree; 3) a Scoring between 10 to 20 point indicates a moderate disability degree; iv) a scoring less than 9 point indicates an high disability degree. In experiment 3 we used the MMSE to analyse the users skills (e.g. memory, attention and, language comprehension) which are usually associated with a human-computer interaction task.

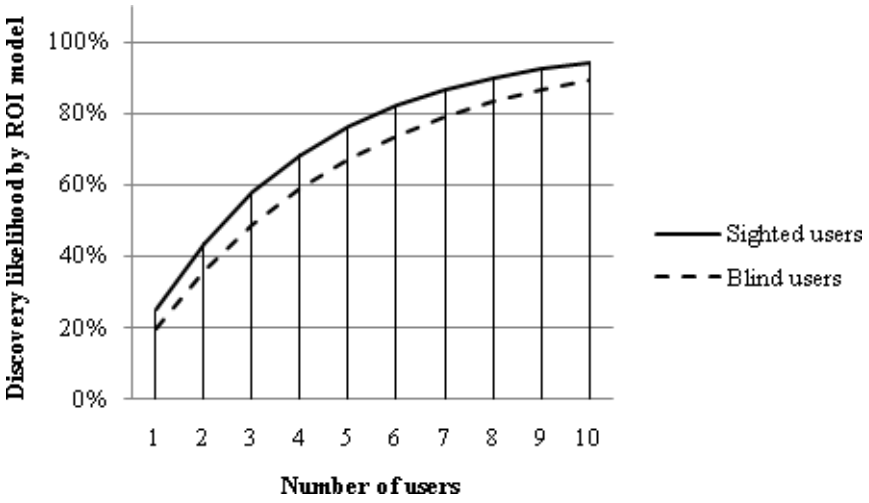
Procedure

The experiments 1 and 2 share the same experimental procedure: *1*) After 20 minutes of free navigation as training *2*) blind users started the PCTA session whereas sighted users started the TA session by following the evaluation tasks presented as scenarios. The evaluation coordinator reported all the problems identified in both the PCTA and the TA session, checking and integrating the report by means of the video recordings of the verbalizations and mouse actions made by each users.

The experiment 3 follows the same procedure adding the MMSE test before the first step.

For all the experiments all the participants are volunteers of different institutions. The not disabled participants are students of the Sapienza University of Rome, while the participants with visual disabilities are members of different institutes for blind people of Rome, and the participants with Down Syndrome are members of the Italian Association of Persons with Down Syndrome (AIPD).

Figure 1. Number of users needed for found more of the 80% of problems. For the ROI model 6 sighted users and 8 blind users are needed for create a mixed panel.



Results

Experiment 1: Blind Users' Interactions with Websites

In the experiment 1 a total of 22 interaction problems have been found: 15 problems were identified by both sighted and blind users, 4 problems were identified only by sighted users and 3 problems were identified only by blind users.

Participants:

- Control group: 6 users (3 male, 3 female, mean age = 22.7) were involved in the TA analysis of the target website.
- Experimental group: 6 blind users (3 male, 3 female, mean age = 27.3) were involved in the PCTA analysis of the target website.

Discovery rate results: According to the *p* value estimated by the ROI model 6 sighted users (*p* =.25) and 8 blind users (*p* =.2) are needed to find more than the 80% of problems (see Figure 1 and Table 1). While the *p* value, estimated by the BDB model, shows that practitioners have to add 4 more sighted users – 10 users for this category (p =.15)—and 3 mode blind users—11 users for this category (p=.14)—to obtain a reliable evaluation (see Figure 2 and Table 2).

Experiment 2: Blind Users' Interactions with Sonificated Search Engines

In the experiment 2 a total of 12 problems have been found: 8 problems were identified by both sighted and blind users and 4 problems were identified only by sighted users.

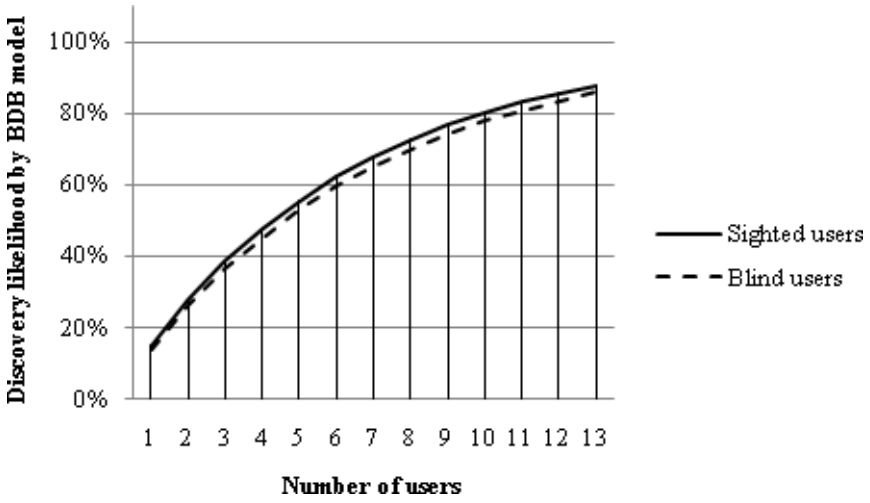
Participants:

- Control group: 4 sighted users (2 male, 2 female, mean age= 25) were involved in the TA analysis of a typical Web search session by using three typologies of graphic layout (Radial, Layered, and Spiral TreeMap) in the visual version of a search engine called WhatsOnWeb.
- Experimental group: 4 blind users (2 male, 3 female, mean age = 28) were involved in the PCTA analysis of a typical Web search session by using three typologies of layout

Table 1. Percentage of problems discovered by sighted and blind users in the experiment 1 calculated by ROI model

Number of users needed for identify more than 80% of problems in the interface	Discovery likelihood of Sighted users	Discovery likelihood of Blind users
1	25%	20%
2	44%	36%
3	58%	49%
4	68%	59%
5	76%	67%
6	**82%**	74%
7	87%	79%
8	90%	**83%**
9	92%	87%
10	94%	89%

Figure 2. Number of users needed for found more of the 80% of problems. For the BDB model 6 sighted users and 8 blind users are needed for create a mixed panel.



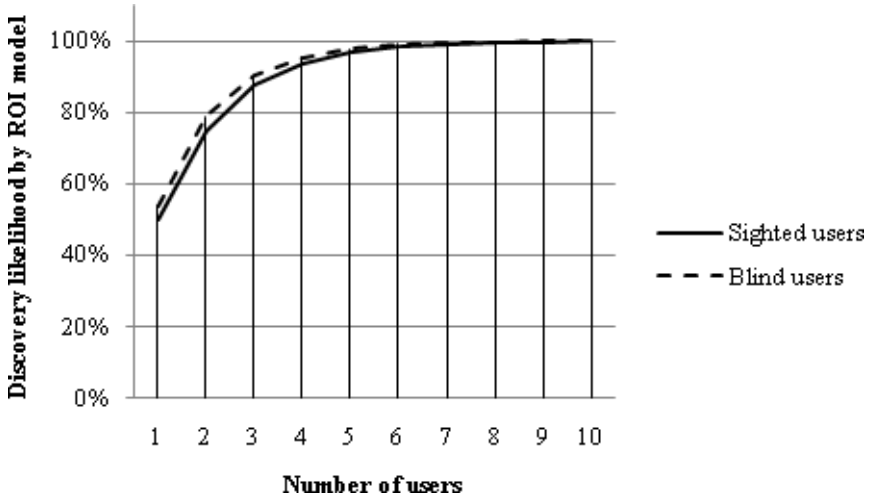
(Radial, Layered and Spiral TreeMap) by means of the PanAndPitchBlinking sonification algorithm in a sonificated version of WhatsOnWeb.

Discovery rate results: According to the *p* value estimated by the ROI model 3 sighted users (*p* =.5) and 3 blind users (*p* =.54) are needed to find more than the 80% of problems (see Figure 3 and Table 3). While the *p* value, estimated by the BDB model, from a side confirms the ROI model estimation of not-disabled users, from another side shows that practitioners have to add 1 more blind user—4 users for this category (*p*= .34)—to obtain a reliable evaluation (see Figure 4 and Table 4).

Table 2. Percentage of problems discovered by sighted and blind users in the experiment 2 by BDB model

Number of users needed for identify more than 80% of problems in the interface	Discovery likelihood of Sighted users	Discovery likelihood of Blind users
1	15%	14%
2	28%	26%
3	39%	36%
4	48%	45%
5	56%	53%
6	62%	60%
7	68%	65%
8	73%	70%
9	77%	74%
10	**80%**	78%
11	83%	**81%**
13	86%	84%
13	88%	86%

Figure 3. Number of users needed for found more of the 80% of problems. For the ROI model, 3 sighted and blind users are needed for create a mixed panel.



Experiment 3: DS Users' Interactions with a Website

In the experiment 3 a total of 16 problems have been found: 4 problems were identified by both not disabled users and DS users. 9 problems were identified only by not disabled users and 3 problems were identified only by DS users.

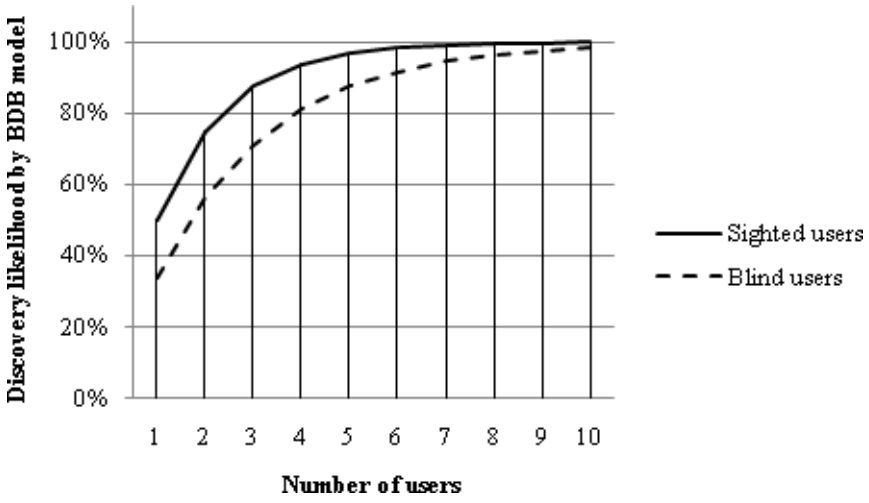
Participants:

- Control group: 6 users (3 male, 3 female, mean age = 26.7) were involved in the TA analysis of the target website.
- Experimental group: 6 users with DS (3 male, 3 female, mean age = 23.4) were in-

Table 3. Percentage of problems discovered by sighted and blind users in the experiment 2 calculated by ROI model

Number of users needed for identify more than 80% of problems in the interface	Discovery likelihood of Sighted users	Discovery likelihood of Blind users
1	50%	54%
2	75%	79%
3	**88%**	**90%**
4	94%	96%
5	97%	98%
6	98%	99%
7	99%	100%
8	100%	100%
9	100%	100%
10	100%	100%

Figure 4. Number of users needed for found more of the 80% of problems in experiment 2. For the BDB model 3 sighted users and 4 blind users are needed for create a mixed panel.



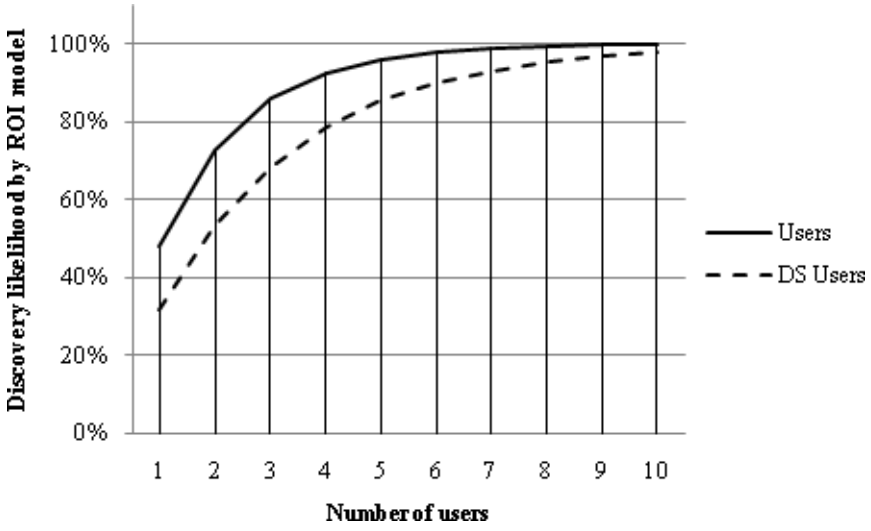
volved in the PCTA analysis of the target website.

Discovery rate results: According to the *p* value estimated by the ROI model 3 not disabled users (p =.48) and 5 DS users (p =.32) are needed to find more than the 80% of problems (see Figure 5 and Table 5).While the *p* value, estimated by the BDB model, from a side confirms the ROI model estimation of not-disabled users, from another side shows that practitioners have to add 3 more DS users—8 users for this category (p= .2)—to obtain a reliable evaluation (see Figure 6 and Table 6).

Table 4. Percentage of problems discovered by sighted and blind users in the experiment 2 calculated by BDB model

Number of users needed for identify more than 80% of problems in the interface	Discovery likelihood of Sighted users	Discovery likelihood of Blind users
1	50%	34%
2	75%	56%
3	**88%**	71%
4	94%	**81%**
5	97%	87%
6	98%	92%
7	99%	95%
8	100%	96%
9	100%	98%
10	100%	98%

Figure 5. Number of users needed for found more of the 80% of problems. For the ROI model 3 users and 5 DS users are needed for create a mixed panel.



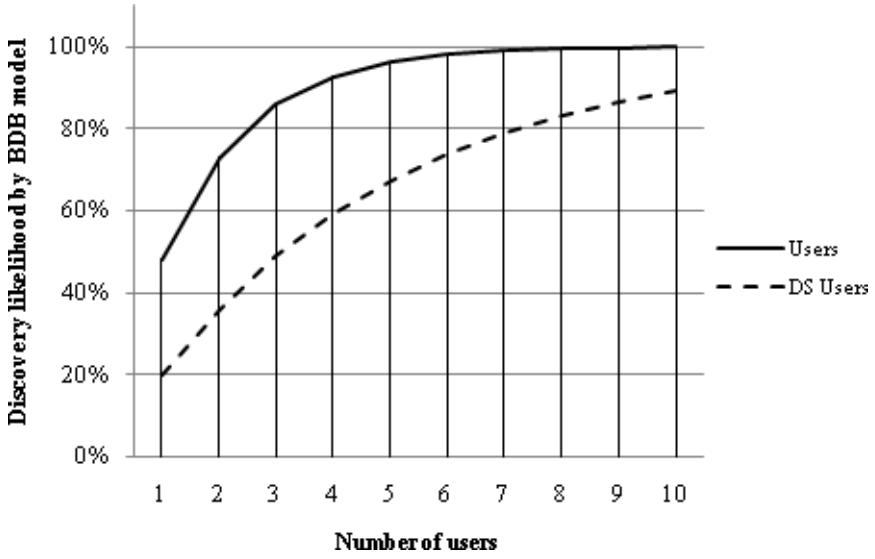
HOW TO APPLY THE BDB MODEL FOR COMPOSE A MIXED PANEL OF USERS

Our findings show that the BDB model is more accurate than the ROI in considering the abilities of the different categories of users in finding interaction problems. As showed in the experiment 1, which has been conducted on a website with many interaction problems, a great difference between the number of subjects recommended for the evaluation through the ROI model and the number of subjects estimated through the BDB model have been found—the site has been subsequently redesigned and some problems previously retrieved in other tests previously carried out (Borsci, et al., 2011) have been fixed. On the other hand, in the experiment 2, in which the tested

Table 5. Percentage of problems discovered by DS and not disabled users in the experiment 3 by ROI model

Number of users needed for identify more than 80% of problems in the interface	Discovery likelihood of users	Discovery likelihood of DS users
1	48%	32%
2	73%	54%
3	**86%**	69%
4	93%	79%
5	96%	**85%**
6	98%	90%
7	99%	93%
8	99%	95%
9	100%	97%
10	100%	98%

Figure 6. Number of users needed for found more of the 80% of problems. For the BDB model 3 users and 8 DS users are needed for create a mixed panel.



interface has been designed by following a user-centred process, the number of problems found by the two groups of users is quite low: in this case the BDB model confirmed the data obtained through the ROI model for both groups of users.

Moreover, the experiment 3, in which the users identified a great set of different problems, shows a high variance between the results obtained through the BDB model and those obtained through the ROI model, especially for the evaluations made with the users with DS. In this case the difference between the two models seems to be due to both different approaches and strategies used by the subjects with DS towards the navigation tasks, leading them to identify a very different set of problems compared to the control group. Therefore, our results do not confirm the Nielsen's (2000) indication about testing usability with

Table 6. Percentage of problems discovered by DS and not disabled users in the experiment 3 by BDB model

Number of users needed for identify more than 80% of problems in the interface	Discovery likelihood of users	Discovery likelihood of DS users
1	48%	20%
2	73%	36%
3	**86%**	49%
4	93%	59%
5	96%	67%
6	98%	74%
7	99%	79%
8	99%	**83%**
9	100%	87%
10	100%	89%

two groups of users, i.e., that a range from 3 to 4 participants for each group can be considered as a sufficient number for a reliable evaluation. Indeed by applying the BDB model the following ranges of users for each category has to be considered by practitioners for a reliable evaluation: from 3 to 10 of not disabled participants, a range from 4 to 11 blind users, and a range from 5 to 8 DS users. Taking into account these results practitioners may start their interaction evaluations with a mixed sample composed of 5 users for each kind of group, and when users find a high number of interaction problems in the tested interfaces we recommend to add at least 5 more not disabled users, 6 blind, and 3 DS users.

CONCLUSION

In this work we have shown the application of a new tool called the BDB model, which is able to support the UCD by both endorsing the PFO approach and promoting the Design for All philosophy. The BDB model has been developed to support an interaction evaluation specialist to create a good mixed sample of disabled and not disabled users to obtain a reliable assessment of the UX during the design process. The use of UCD process driven by a matching between users' need and the prototype's features, guarantees to the designer the possibility to obtain a final product with a great degree of UX satisfaction. In this sense, the results obtained with the BDB model clearly show that the practitioners aiming to obtain a complete evaluation of an interface have to both consider all the possible divergent navigation strategies and recruit a mixed panel of users for identifying more than the 80% of the interaction problems. Comparing our results with the indications given by the Nielsen (2000) model, at least 5 subjects for each category are needed to conduct a complete evaluation, even though practitioners have to expect a necessary increase in number for the category of blind and DS users. Practitioners are suggested to start the evaluation with a mixed sample of five users for each category. Moreover we suggest to the evaluation specialists the application of the BDB model in order to obtain an exact estimation on how many users for each category have to be added to the sample to obtain at least 80% of the interaction problems. In this way, the use of the BDB model should help practitioners to optimize the evaluation process by involving disabled users.

At the same time this model should allow both implementation and evaluation of technologies by a user-driven process.

Summarizing, the key points that make BDB better than ROI are the following:

- The BDB model computes probabilistic individual differences in problem identification.
- The BDB considers evaluated interfaces as an object *per se*, not like in an ordinal sequence, differently from the ROI, which considers the rate of the evaluators starting from the number of problems found by the first evaluator (evaluation condition).
- The BDB computes all evaluated interfaces aligning them to a start zero point independently from the first evaluator.
- The BDB model extends the representativeness of the sample resampling the population assessing all possible participants' evaluation behaviours in order to provide the number of evaluators needed for an interaction evaluation.
- The BDB provides a more accurate number of users needed for a specific interaction evaluation with and without disability.

REFERENCES

Bohman, P. R., & Anderson, S. (2005). *A conceptual framework for accessibility tools to benefit users with cognitive disabilities.* Paper presented at the International Cross-Disciplinary Workshop on Web Accessibility. Chiba, Japan.

Borsci, S., Kurosu, M., Federici, S., & Mele, M. L. (2013). *Computer systems experiences of users with and without disabilities: An evaluation guide for professionals*. London, UK: CRC Press.

Borsci, S., Londei, A., & Federici, S. (2011). The bootstrap discovery behaviour (BDB): A new outlook on usability evaluation. *Cognitive Processing*, *12*(1), 23–31. doi:10.1007/s10339-010-0376-6

Caulton, D. (2001). Relaxing the homogeneity assumption in usability testing. *Behaviour & Information Technology*, *20*(1), 1–7. doi:10.1080/01449290010020648

Charlton, J. I. (1998). *Nothing about us without us: Disability oppression and empowerment*. Berkeley, CA: University of California Press.

Di Giacomo, E., Didimo, W., Grilli, L., & Liotta, G. (2007). Graph visualization techniques for web clustering engines. *IEEE Transactions on Visualization and Computer Graphics*, *13*(2), 294–304. doi:10.1109/TVCG.2007.40

Di Giacomo, E., Didimo, W., Grilli, L., Liotta, G., & Palladino, P. (2008). *WhatsOnWeb+: An enhanced visual search clustering engine.* Paper presented at the IEEE Pacific Visualization Symposium. Kyoto, Japan.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, *7*(1), 1–26. doi:10.1214/aos/1176344552

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, *87*(3), 215–251. doi:10.1037/0033-295X.87.3.215

Fairweather, P., & Trewin, S. (2010). Cognitive impairments and Web 2.0. *Universal Access in the Information Society*, *9*(2), 137–146. doi:10.1007/s10209-009-0163-2

Federici, S., & Borsci, S. (2010). Usability evaluation: models, methods, and applications. In J. Stone & M. Blouin (Eds.), *International Encyclopedia of Rehabilitation*. Buffalo, NY: Center for International Rehabilitation Research Information and Exchange (CIRRIE). Retrieved from http://cirrie.buffalo.edu/encyclopedia/article.php?id=277&language=en.

Federici, S., Borsci, S., & Mele, M. L. (2010). Usability evaluation with screen reader users: A video presentation of the PCTA's experimental setting and rules. *Cognitive Processing*, *11*(3), 285–288. doi:10.1007/s10339-010-0365-9

Federici, S., Borsci, S., Mele, M. L., & Stamerra, G. (2010). Web popularity: An illusory perception of a qualitative order in information. *Universal Access in the Information Society*, *9*(4), 375–386. doi:10.1007/s10209-009-0179-7

Federici, S., Borsci, S., & Stamerra, G. (2010). Web usability evaluation with screen reader users: Implementation of the partial concurrent thinking aloud technique. *Cognitive Processing*, *11*(3), 263–272. doi:10.1007/s10339-010-0358-8

Federici, S., Micangeli, A., Ruspantini, I., Borgianni, S., Corradi, F., & Pasqualotto, E. (2005). Checking an integrated model of web accessibility and usability evaluation for disabled people. *Disability and Rehabilitation*, *27*(13), 781–790. doi:10.1080/09638280400014766

Feng, J., Lazar, J., Kumin, L., & Ozok, A. (2008). *Computer usage by young individuals with down syndrome: An exploratory study.* Paper presented at the 10th International ACM SIGACCESS Conference on Computers and Accessibility. Halifax, Canada.

Feng, J., Lazar, J., Kumin, L., & Ozok, A. (2010). Computer usage by children with down syndrome: Challenges and future research. *Transactions on Accessible Computing*, *2*(3), 1–44. doi:10.1145/1714458.1714460

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189–198. doi:10.1016/0022-3956(75)90026-6

Fox, J. (2002). *An R and S-plus companion to applied regression*. Thousand Oaks, CA: SAGE.

Garrett, J. J. (2003). *The elements of user experience: User-centered design for the web*. New York, NY: New Riders Press.

Goodstein, R. L. (1963). *Boolean algebra*. Oxford, UK: Pergamon Press.

Hartson, H. R., Andre, T. S., & Williges, R. C. (2001). Criteria for evaluating usability evaluation methods. *International Journal of Human-Computer Interaction*, *13*(4), 373–410. doi:10.1207/S15327590IJHC1304_03

Hertzum, M., & Jacobsen, N. E. (2003). The evaluator effect: A chilling fact about usability evaluation methods. *International Journal of Human-Computer Interaction*, *15*(4), 183–204. doi:10.1207/S15327590IJHC1501_14

ISO. (1998). *ISO 9241-11: Ergonomic requirements for office work with visual display terminals*. Geneve, Switzerland: ISO.

ISO. (2010). *Ergonomics of human-system interaction – Part 210: Human-centred design for interactive systems*. Geneve, Switzerland: ISO.

Lewis, C. (2005). HCI for people with cognitive disabilities. *ACM SIGACCESS Accessibility and Computing*, *83*, 12–17. doi:10.1145/1102187.1102190

Lewis, J. R. (1994). Sample sizes for usability studies: Additional considerations. *Human Factors*, *36*(2), 368–378.

Lewis, J. R. (2001). Evaluation of procedures for adjusting problem-discovery rates estimated from small samples. *International Journal of Human-Computer Interaction*, *13*(4), 445–479. doi:10.1207/S15327590IJHC1304_06

Lewis, J. R. (2006). Sample sizes for usability tests: Mostly math, not magic. *Interaction*, *13*(6), 29–33. doi:10.1145/1167948.1167973

Luckasson, R., Borthwick-Duffy, S., Buntinx, W. H. E., Coulter, D. L., Craig, E. M., & Reeve, A. (2002). *Mental retardation: Definition, classification, and system of supports* (10th ed.). Washington, DC: AAMR.

Mele, M. L., Federici, S., Borsci, S., & Liotta, G. (2010). Beyond a visuocentric way of a visual web search clustering engine: The sonification of WhatsOnWeb. In Miesenberger, K., Klaus, J., Zagler, W., & Karshmer, A. (Eds.), *Computers Helping People with Special Needs* (*Vol. 1*, pp. 351–357). Berlin, Germany: Springer. doi:10.1007/978-3-642-14097-6_56

Nielsen, J. (1994). Heuristic evaluation. In Nielsen, J., & Mack, R. L. (Eds.), *Usability inspection methods*. New York, NY: John Wiley & Sons.

Nielsen, J. (2000). *Why you only need to test with 5 users*. Retrieved May, 20th, 2011, from http://www.useit.com/alertbox/20000319.html.

Nielsen, J., & Landauer, T. K. (1993). *A mathematical model of the finding of usability problems.* Paper presented at the Conference on Human factors in computing systems: INTERACT and CHI 1993. Amsterdam, The Netherlands.

Nielsen, J., & Mack, R. L. (Eds.). (1994). *Usability inspection methods*. New York, NY: John Wiley & Sons.

Norman, D. A. (1988). *The psychology of everyday things*. New York, NY: Basic Books.

Schalock, R. L., & Luckasson, R. (2004). American association on mental retardation's definition, classification, and system of supports and its relation to international trends and issues in the field of intellectual disabilities. *Journal of Policy and Practice in Intellectual Disabilities*, *1*(3-4), 136–146. doi:10.1111/j.1741-1130.2004.04028.x

Schmettow, M. (2008). *Heterogeneity in the usability evaluation process.* Paper presented at the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity. Liverpool, UK.

Spool, J., & Schroeder, W. (2001). *Testing web sites: Five users is nowhere near enough.* Paper presented at the Human Factors in Computing Systems: CHI 2001. Seattle, WA.

Stephanidis, C. (2001). User interfaces for all: New perspectives into human-computer interaction. In Stephanidis, C. (Ed.), *User Interfaces for All: Concepts, Methods, and Tools* (pp. 3–17). Mahwah, NJ: Lawrence Erlbaum Associates.

Virzi, R. A. (1990). Streamlining the design process: Running fewer subjects. *Human Factors and Ergonomics Society Annual Meeting Proceedings, 34*, 291-294.

Virzi, R. A. (1992). Refining the test phase of usability evaluation: How many subjects is enough? *Human Factors*, *34*(4), 457–468.

Wright, P. C., & Monk, A. F. (1991). A cost-effective evaluation method for use by designers. *International Journal of Man-Machine Studies*, *35*(6), 891–912. doi:10.1016/S0020-7373(05)80167-1

ADDITIONAL READING

Albert, D. M. (1999). Psychotechnology and insanity at the wheel. *Journal of the History of the Behavioral Sciences*, *35*(3), 291–305. doi:10.1002/(SICI)1520-6696(199922)35:3<291::AID-JHBS6>3.0.CO;2-1

Andronico, P., Buzzi, M., & Leporini, B. (2004). *Can I find what I'm looking for?* Paper presented at the 13th International World Wide Web Conference on Alternate Track. New York, NY.

Annett, J. (2002). Subjective rating scales in ergonomics: A reply. *Ergonomics*, *45*(14), 1042–1046. doi:10.1080/00140130210166762

Annett, J. (2002). Subjective rating scales: Science or art? *Ergonomics*, *45*(14), 966–987. doi:10.1080/00140130210166951

Ascott, R. (1995). The architecture of cyberception. In Toy, M. (Ed.), *Architectural Design* (pp. 38–40). London, UK: Academy Editions.

Baber, C. (2002). Subjective evaluation of usability. *Ergonomics*, *45*(14), 1021–1025. doi:10.1080/00140130210166807

Brooke, J. (1996). SUS: A "quick and dirty" usability scale. In Jordan, P. W., Thomas, B., Weerdmeester, B. A., & McClelland, I. L. (Eds.), *Usability Evaluation in Industry* (pp. 189–194). London, UK: Taylor & Francis.

Bruner, J. S., & Postman, L. (1949). On the perception of incongruity: A paradigm. *Journal of Personality*, *18*(2), 206–223. doi:10.1111/j.1467-6494.1949.tb01241.x

Clark, R., Williams, J., Clark, J., & Clark, C. (2003). Assessing web site usability: Construction zone. *Journal of Healthcare Information Management*, *17*(2), 51–55.

De Kerckhove, D. (1995). *The skin of culture: Investigating the new electronic reality*. Toronto, CA: Somerville.

De Kerckhove, D. (2001). *The architecture of intelligence*. Berlin, Germany: Birkhäuser.

Drury, C. G. (2002). Measurement and the practising ergonomist. *Ergonomics*, *45*(14), 988–990. doi:10.1080/00140130210166915

Hutchins, E. L., Hollan, J. D., & Norman, D. A. (1985). Direct manipulation interfaces. *Human-Computer Interaction*, *1*(4), 311–338. doi:10.1207/s15327051hci0104_2

Jordan, P. W. (1998). *An introduction to usability*. London, UK: Taylor and Francis.

Karwowski, W. (Ed.). (2006). *International encyclopedia of ergonomics and human factors* (2nd ed.). Boca Raton, FL: CRC Press. doi:10.1201/9780849375477

Kirakowski, J. (2002). Is ergonomics empirical? *Ergonomics*, *45*(14-15), 995–997. doi:10.1080/00140130210166889

Krug, S. (2000). *Don't make me think! A common sense approach to web usability*. Indianapolis, IN: New Riders.

Monk, A., Wright, P., Haber, J., & Davenport, L. (Eds.). (1993). *Improving your human computer interface: A practical technique*. New York, NY: Prentice Hall.

Nielsen, J., & Loranger, H. (2006). *Prioritizing web usability* (2nd ed.). Berkeley, CA: New Riders Press.

Nielsen, J., & Pernice, K. (2009). *Eyetracking web usability*. Berkeley, CA: New Riders.

Norman, J. (2002). Two visual systems and two theories of perception: An attempt to reconcile the constructivist and ecological approaches. *The Behavioral and Brain Sciences*, *25*(1), 96–144.

Shneiderman, B. (1983). Direct manipulation: A step beyond programming languages. *Computer*, *16*(8), 57–69. doi:10.1109/MC.1983.1654471

Shneiderman, B. (1987). Direct manipulation: A step beyond programming languages. In Shneiderman, B. (Ed.), *Human-Computer Interaction: A Multidisciplinary Approach* (pp. 461–467). Burlington, MA: Morgan Kaufmann. doi:10.1109/MC.1983.1654471

Stanton, N. A., & Young, M. S. (1999). *A guide to methodology in ergonomics: Design for human use*. London, UK: Taylor & Francis.

APPENDIX 1

Box 1. Bootstrap Discovery Behaviour model code for Matlab

```
function [Nsubj085,baseerr,c,gof]=BDB(errors, NBS)
  % BDB    Calculates the number of subjects for the detection of 85% of
  % problems in BDB approach by bootstrap iterations.
  %
  % [Nsubj085,baseerr,c,gof]=BDB(errors, NBS)
  %
  % errors: matrix with total amount of subjects (Nsub) rows and number of
  %         errors (Nerrors) columns.
  % NBS: number of bootstrap iterations.
  % Nsubj085: estimated number of subject to reveal the 85% of errors
  % baseerr: bias of errors considered as certainly found with no subjects.
  % c: fit object that encapsulates the result of fitting (from function
  % FIT).
  % gof: structure with fitting statistical information (from function
  % FIT).

    Nsubj=size(errors,1);
    Nerrors=size(errors,2);

    bootstrap=zeros(Nsubj,NBS);

    % Bootstrap loop
    for b=1:NBS,
        exptrial=zeros(1,Nerrors);
        ind=ceil(rand(Nsubj,1)*Nsubj);
        for k=1:Nsubj,
            exptrial=exptrial|errors(ind(k),:);
            bootstrap(k,b)=sum(exptrial');
        end
    end
    results=mean(bootstrap,2);
    stdresults=std(bootstrap,0,2);

    resultsnorm=results/Nerrors;

    % Fit of the averaged errors
    s = fitoptions('Method','NonlinearLeastSquares','Robust','LAR','Lower',[0
0 -Inf],'MaxFunEvals',2000,'MaxIter',1000,'StartPoint',[0.5 0.5 0.5]);
    f = fittype('a-(1-p)^(x+q)','options',s);
```

```
    [c,gof]=fit((1:Nsubj)',resultsnorm,f);

    plot((1:Nsubj)',resultsnorm,'x',(1:Nsubj)',c((1:Nsubj)'),'r-');

    % Find the desired parameters
    if(c.a-0.85>0)
        Nsubj085=log(c.a-0.85)/log(1-c.p) - c.q;
    else
        Nsubj085=NaN;
    end

    baseerr=Nerrors*(c.a-(1-c.p)^c.q);

    plot((1:Nsubj)',resultsnorm,'x',(1:Nsubj)',c((1:Nsubj)'),'r-');
end
```