Cogn Process (2011) 12:23–31
DOI 10.1007/s10339-010-0376-6

RESEARCH REPORT

The Bootstrap Discovery Behaviour (BDB): a new outlook on usability evaluation

Simone Borsci · Alessandro Londei · Stefano Federici

Received: 5 July 2010 / Accepted: 12 October 2010 / Published online: 3 November 2010

Abstract The value of λ is one of the main issues debated in international usability studies. The debate is centred on the deficiencies of the mathematical return on investment model (ROI model) of Nielsen and Landauer (1993). The ROI model is discussed in order to identify the base of another model that, respecting Nielsen and Landauer's one, tries to consider a large number of variables for the estimation of the number of evaluators needed for an interface. Using the bootstrap model (Efron 1979), we can take into account: (a) the interface properties, as the properties at zero condition of evaluation and (b) the probability that the population discovery behaviour is represented by all the possible discovery behaviours of a sample. Our alternative model, named Bootstrap Discovery Behaviour (BDB), provides an alternative estimation of the number of experts and users needed for a usability evaluation. Two experimental groups of users and experts are involved in the evaluation of a website (http://www.serviziocivile.it). Applying the BDB model to the problems identified by the two groups, we found that 13 experts and 20 users are needed to identify 80% of usability problems, instead of 6 experts and 7 users required according to the estimation of the discovery likelihood provided by the ROI model. The consequence of the difference between the results of those models is that in following the BDB the costs of usability evaluation increase, although this is justified considering that the results obtained have the best probability of representing the entire population of experts and users.

Keywords Asymptotic test · Bootstrap · Effectiveness · Return of investment · User experience evaluation

Introduction

Nielsen and Landauer (1993) show that, generally, the least number of evaluators (experts or users) required for usability evaluation techniques ranges from three to five. The mathematical model of those authors was run for the problems identified by the experts or users in order to evaluate whether the technique is efficient or cost effective: "observing additional participants reveals fewer and fewer new usability problems" (Turner et al. 2006, p.3084); thus, adding more than four or five users (participants) does not provide an advantage in estimating rates of discovery of new problems in terms of costs, benefits, efficiency and effectiveness. This model, known as return on investment (ROI), is an asymptotic test able to estimate the number of evaluators needed with the following formula:

$$Found_{(i)} = N[1 - (1 - \lambda)^i] \quad (1)$$

Electronic supplementary material The online version of this article (doi:10.1007/s10339-010-0376-6) contains supplementary material, which is available to authorized users.

S. Borsci (✉) · A. Londei · S. Federici
ECoNA—Interuniversity Centre for Research on Cognitive Processing in Natural and Artificial Systems,
Sapienza University of Rome,
Rome, Italy
e-mail: simone.borsci@gmail.com

S. Federici
Department of Human and Education Sciences,
University of Perugia,
Perugia, Italy

In (1), N is the total number of problems in the interface, λ[1] is defined by Nielsen and Landauer (1993, p.208) as "the probability of finding the average usability problem when running a single average subject test" (i.e. individual detection rate), and i is the number of users. As some international studies (Lewis 1994; Nielsen 2000; Nielsen and Mack 1994; Virzi 1990; 1992; Wright and 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 (λ) is at least 0.30. By using this mathematical model, the range of evaluators required for a usability test can be found, and therefore the increase in the number of problems found by adding users to the evaluation can be calculated. For instance, by applying formula (1), practitioners can estimate whether five users are sufficient for an efficient assessment or, otherwise, how many users (n) are needed in order to increase the percentage of usability problems, as follows:

$$Found_{(s)} = [1 - (1 - 0.3)^{s}] = 0.83$$

In this example of a potential application of the formula, provided by Nielsen (2000), the problem detection rate obtained with five users is 0.83 (i.e. 83% of usability problems will be detected). However, we must emphasise that many studies (Lewis 1994; Turner et al. 2006; Virzi 1990; 1992) show λ ranging from 0.16 to 0.42 (see Lewis 2006). Afterwards, the increase in the problem detection rate can be estimated by adding more users to this sample of five, as reported in Fig. 1.The analysis of that hypothetical sample shows that almost 100% of usability problems can be found with 15 users, considering the fact that with just 5 users the likelihood of problem discovery is equal to 83%, but in order to discover less than 20% more usability problems, not yet identified, at least 10 more users need to be added to the evaluation.

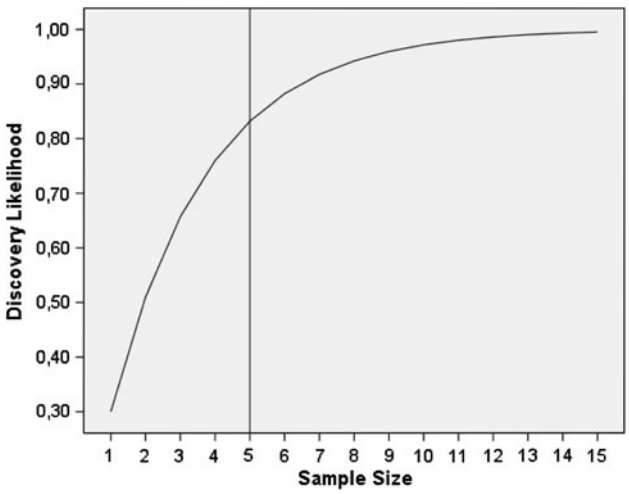

Fig. 1 The asymptotic behaviour of discovery likelihood in relation to our hypothetical sample with $\lambda = 0.30$

The deficiencies of the λ value estimation

As Nielsen et al. (1993) underline when discussing their model, the discoverability rate (λ) 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;
- Type and quality of the methodology used to conduct the test;
- Specific tasks selected;
- Match between the test and the context of real-world usage;
- Representativeness of the test participants;
- Skill of the evaluator.

These factors have an effect on the evaluation of the interaction between system and user that, in our opinion, the ROI model is not able to estimate. Indeed, the ROI model assumes that

All the evaluators have the same probability of finding all problems. As Caulton (2001) states, the ROI model is based on the idea that all types of subjects have the same probability of encountering all potential usability problems, without considering their different evaluation skills. At the same time, the ROI model does not take into consideration the effects of the evaluation methodologies being used, the representativeness of the sample of participants, or, finally, the similarity between the test and its context in the real world. In particular, as Woolrych and Cockton (2001) have claimed, the ROI model fails to integrate all of the individual differences in problem discoverability; in this sense, the probability of the participants encountering all of the problems remains a relevant issue that needs clarification. Recently, Schmettow (2008), while discussing the assumption of a homogeneous detection probability, claimed that it is intuitively unrealistic. This author proposes a beta-binomial distribution, in which the value of "p" is estimated using a process which is able to take heterogeneity into account. However, Schmettow (2008) demonstrates the problems of the assumption of heterogeneity in the ROI model, without proposing a real solution.

[1] Actually, only Nielsen et al. (1993) used λ, instead of p (Lewis 1994; 2006; Virzi 1992; Wright and Monk 1991; Schmettow 2008) in the formula 1, partly because they derived their formula from the "Poisson process" (see Nielsen and Landauer 1993). Many authors (Lewis 1994; 2006; Virzi 1992; Wright and Monk 1991; Schmettow 2008) use the formula (1) written as: $P = 1 - (1 - p)^{n}$, where "P" is the total number of problems in the interface, "p" the probability of finding the average usability problem when running a single average subject test and "n" is the number of participants.

The λ value estimation does not take into account the differences between the systems evaluated. This means that the effect on the evaluation results caused by the properties of the system, the interface lifecycle stage and the methodologies selected for the evaluation are not considered by the model. In fact, 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 (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, is a factor that this model does not consider. 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 (see Lewis 2001). A serious limitation of Nielsen's (and both Landauer and Virzi's) work is that they happened to be working with products for which the value of p across evaluators/users was about 0.3, but as Lewis (1994) showed, it is possible for the composite value of p to be much lower than 0.3. For Lewis (1994), the value was 0.16, and for Spool and Schroeder (2001; see also Lewis 2006) it was 0.029. 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 0.3 (30%).

The international debate on the estimation of the value of λ also shows that the ROI model suffers from an overestimation of λ, or as Woolrych and Cockton (2001) claim, at least from an optimistic estimation of the discovery rate. As Schmettow (2008, p.94) underlines, Lewis (2001) in order to resolve this problem of overestimation "compared several correction terms in application to real data sets. The final suggestion was an equally weighted combination of a simplified Good-Turing (GT) adjustment and a normalization procedure (NORM) proposed by Hertzum and Jacobsen (2003)". This adjustment is able to deflate the overestimated value of λ, estimated using a small sample, but without solving all of the problems, as previously discussed, that Nielsen and Landauer's model generates.

Taking into account the deficiencies of the ROI model, in a usability evaluation practitioners must consider that the estimation of λ could have a variable range of values and as a consequence, that this model cannot guarantee the reliability of the evaluation results obtained by the first five participants.

This analysis allows us to provide an alternative model to the ROI one based on the probabilistic behaviour in the evaluation. As its first feature, our alternative model should be able to take into account the probabilistic individual differences in problem identification. The second feature of our model is that it must consider 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 model is that in order to calculate the number of evaluators needed for the evaluation, it must consider the representativeness of the sample (as regards the population of all the possible evaluation behaviours of the participants). Our model is based on the statistical inference methods known as bootstrapping.

A new look: the Bootstrap Discovery Behaviour

Bootstrapping is a general approach to statistical inference based on building a sampling distribution for a statistic by resampling from the data at hand. 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 (for a general introduction to bootstrapping methods see 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

$$p(x_1 V x_2) \tag{2}$$

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:

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

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:

$$p[\neg(\neg x_1 \Lambda \neg x_2)] = 1 - [1 - p(x_1)] * [1 - p(x_2)] \quad (4)$$

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:

$$p(x_1 \mathrm{V} x_2) = 1 - [1 - p]^2 \quad (5)$$

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

$$p(x_1 \mathrm{V} x_2 \mathrm{V} \ldots x_L) = 1 - [1 - p]^L \quad (6)$$

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. Our idea is that the bootstrap model should be able to grant a more reliable estimation of the probability of identifying a problem.

Experiment

In order to test the Bootstrap Discovery Behaviour:

- Two experimental groups are asked to evaluate a target interface: 25 experts by means of a cognitive walkthrough (CW) technique and 20 users by using the thinking aloud (TA) technique.
- Using the "Fit" function in Matlab software (http://www.mathworks.com), we applied a bootstrap with 5000 samplings. The results of each subsample were obtained by a random order of evaluators (experts and users) with repetition.
- In order to identify the best fit of the data within a 95% confidence interval, the result of each bootstrap sampling allowed us to estimate three parameters: (i) the probable number of problems found (p): this value was obtained as the normalized mean number of problems found by each subgroup of subjects; (ii) the maximum number of problems that all possible samples could identify (a), known as the maximum limit; and (iii) the value of the known term q.

Participants

Experts group

This group was comprised of 25 experts (10 males, 15 females, mean age = 26.6) with different levels of expertise: 10 experts had more than 3 years of experience and 15 had less than 1 year of experience. All the experts evaluated the target website with a CW technique.

Users group

Twenty students from Sapienza University of Rome (5 males, 15 females, mean age = 21.3) were involved in the TA analysis of the target website.

Evaluation techniques

Cognitive walkthrough

This starts with a task analysis that allows (a) the sequence of steps a user should take in order to accomplish a task to be specified and (b) the system responses to the actions to be observed. Once the task analysis is over, the expert simulates the actions of the potential user and identifies the problems the user is supposed to find. As Rieman, Franzke, and Redmiles (1995) claim, this technique is based on three elements: "(1) a general description of who the users will be and what relevant knowledge they possess, (2) a specific description of one or more representative tasks to be performed with the system, and (3) a list of the correct actions required to complete each of these tasks with the interface being evaluated" (p. 387).

The experts perform the walkthrough by asking themselves a set of questions for each subtask (Lewis and Rieman 1993; Polson et al. 1992; Wharton et al. 1994):

- The user sets a goal to be accomplished with the system (for example, checking the spelling of this document).
- The user searches the interface for currently available actions (menu items, buttons, command-line inputs, etc.).
- The user selects the action that seems likely to lead to progress towards the goal.
- The user performs the selected action and evaluates the system's feedback for evidence that progress is being made towards the current goal.

Thinking aloud

Known as verbal protocol analysis, this had a large application in the study of consumer and judgement-making processes (Bellman, Park 1980; Bettman 1979; Biehal and Chakravarti 1982a; b; 1986; 1989; Green 1995; Kuusela et al. 1998). In describing this user-based evaluation process, Kuusela and Pallab (2000) state: "The premise of this procedure is that the way subjects search for information, evaluate alternatives, and choose the best option can be registered through their verbalization and later be analysed to discover their decision processes and patterns. Protocol data can provide useful information about cue stimuli, product associations, and the terminology used by consumers" (p. 388). The TA can be performed according to two main different experimental procedures: the first procedure, and the most popular, is the concurrent verbal protocol, with which data are collected during the decision task; the second procedure is the retrospective verbal protocol, with which data are collected when the decision task is over. Our experimental work has used the concurrent TA because it is one of the most frequently applied techniques of verbal reporting used in HCI studies. Indeed, in the concurrent TA, users express their problems, strategies, stress and impressions without the influence of a "rethinking" perception, as happens in retrospective analysis (Borsci and Federici 2009; Federici and Borsci 2010; Federici et al. 2010a; b).

Each test was performed at the laboratory of cognitive psychology of the Sapienza University of Rome with a specific setting represented in Fig. 2.

Apparatus

Each participant uses an Intel-Pentium 4 computer with a RAM of 4 GB, a Gforce 8800 (video), and a Creative Sound Blaster X-Fi (audio). The monitor is a SyncMaster900p, 19", and the speakers are two Creative GigaWorks T20 Series II. Each test is video recorded with a Sony 3 megapixels and each user screen movement is recorded by the CamStudio 20 screen recorder. A 28″ Sony monitor is used as a control of user action by the expert. Each user used Internet Explorer 8 as a browser.

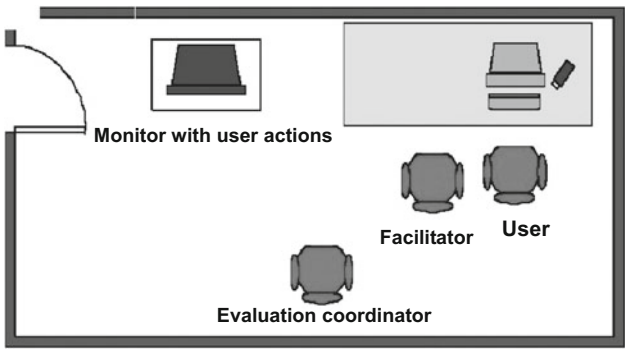

Fig. 2 The users' experimental settings for the TA analysis

Target websites

http://www.serviziocivile.it was chosen as the target website. It was selected from those websites of the Italian Public Administration considered accessible by the CNIPA evaluation (http://www.pubbliaccesso.gov.it/logo/elenco.php). We chose serviziocivile.it for two main reasons: (1) from a structural point of view, it offers a high quantity of information collected in a large number of pages; (2) from the point of view of the analysis we had to carry out, the fact that the website's target users are people between 18 and 28 years old eased our enrolment possibilities in order to form the usability evaluation samples involved in user-based evaluations.

The expert-based and user-based analyses were carried out on four scenarios. These scenarios were created and approved by three external evaluators with more than 3 years of experience in the field. These evaluators did not participate in the experimental sessions.

Procedure

Experts group

In a meeting with all experts, the evaluation coordinator (the second author of this paper) presented the procedure, goals and scenarios provided by three external experts with more than 5 years of experience in accessibility and usability evaluation. Then, all experts were invited to evaluate the system through a CW and to provide independent evaluations.

Users group

After 20 min of free navigation as training, users started the TA evaluation following four scenarios (see Appendix). The evaluation coordinator reported all problems identified in the TA session, and checked and integrated the report using the video of verbalization and mouse action recorded by CamStudio.

Measurements and tools

We compared the number of evaluators needed in order to achieve identification of 80% of problems, applying both the ROI and our bootstrap model. The analysis was carried out by SPSS 16 and Matlab software.

Alternative model of estimation

In order to identify the number of experts and users needed to detect more than 80% of problems, we must obtain the best fit with our results. Our model must also provide an

estimation of those parameters able to represent the properties of the interface and the representativeness of the sample. The bootstrap analysis was used in order to obtain the following parameters:

- *All the possible discovery behaviours of participants.* Considering our 5,000 possible bootstrap samples (with repetition), at each bootstrap step a subsample composed of collected data (i.e. the identified problems) presented in a random order was selected. The maximum value of collected problems represents our maximum limit value (indicated below in (8) as *a*). This value indicates the representativeness of our sample.
- *A rule in order to select the representative data.* As representative data for the subsamples, we used the normalized mean of the number of problems found by each subsample (indicated below in (8) as *p*). As already mentioned, *p* is the estimated probability of the detection of a generic problem by an evaluator in the chosen population.

The model expressed below in 8 represents the best fit of the data obtained by the bootstrapped subsamples of expert and user groups:

$$F(L) = N_t\left[a - (1-p)^{L+q}\right] \quad (8)$$

In (8), *Nt* represents the total number of problems in the interface and the q variable expresses the hypothetical condition $L = 0$ (an analysis without evaluators). In other words, since F does not vanish when $L = 0$, $F(0)$ represents the amount of evident problems that can be effortlessly detected by any subject, and q the possibility of detecting a certain number of problems that have already been identified (or are evident to identify) and were not fixed by the designer:

$$F(0) = N_t[a - (1-p)^{q}] \quad (9)$$

The value q represents the properties of the interface from the evaluation perspective. This is at least the "zero condition" of the interface properties.

Results obtained by applying the ROI model

Experts identified 46 problems with a value of λ equal to 0.26. The number of experts needed to find 80% of problems equalled 6 (Fig. 3).

Users identified 39 problems with a value of λ equal to 0.22. The number of users needed to find 80% of problems equalled 7 (Fig. 4).

Results obtained by applying the BDB

Experts group

Applying our model to the data obtained by the experts with the classic CW, we obtained the probable discovery likelihood expressed in Fig. 5.

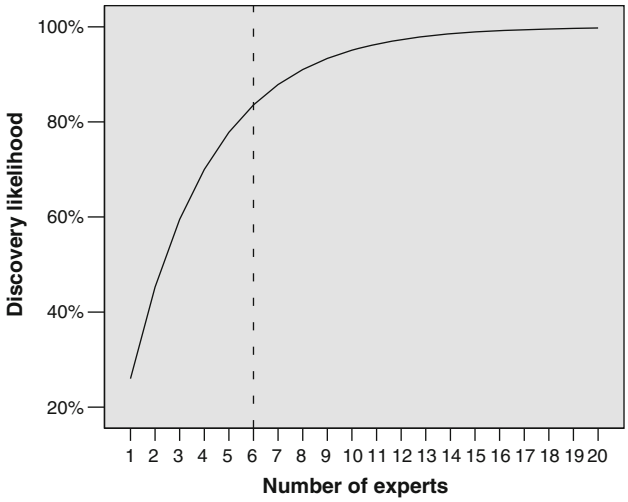

Fig. 3 The discovery likelihood of the experts group

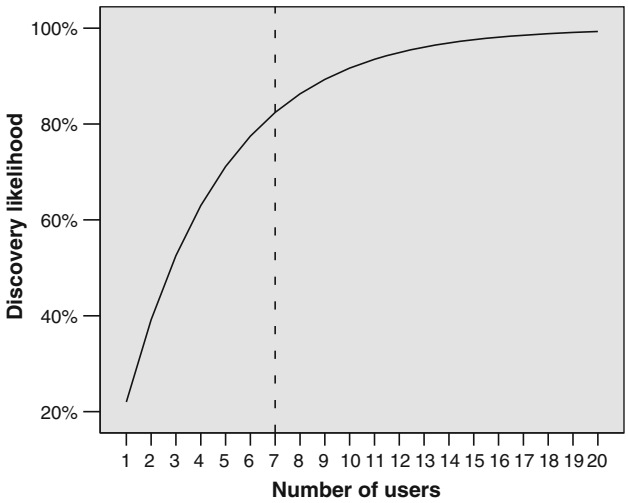

Fig. 4 The discovery likelihood of the users group

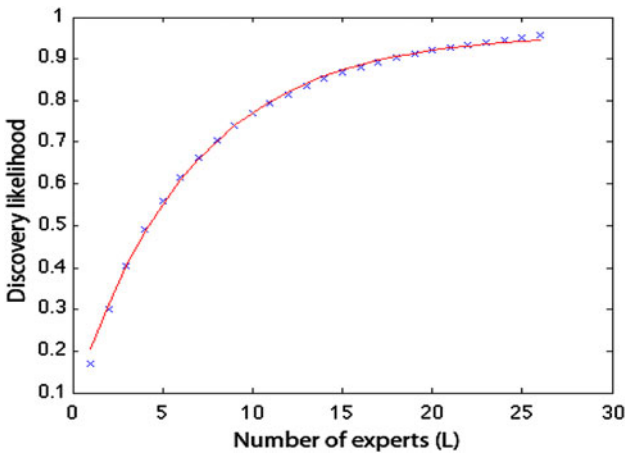

Fig. 5 Discovery likelihood of experts with CW1 estimated by a 5,000 bootstrap sampling

Table 1 The values of the parameters a, p and q calculated by the bootstrap of experts data

Parameters	Values	Confidence interval
a	0.9623	(0.9583–0.9664)
p	0.1414	(0.1387–0.1440)
q	0.8356	(0.7706–0.9006)

The values of the parameters needed for calculating the model (a, p, q) are reported in Table 1 below:

Our results show that 13 experts are needed for the evaluation in order to identify more than 80% of problems (the number of known problems q is 3.77).

User group

The data for the user group were processed as for the experts' one. The probable discovery likelihood of the user sample is reported in Fig. 6.

The parameters obtained are reported in Table 2:

Applying the results to Equation (8), the result shows that 20 users are needed for the evaluation in order to identify more than 80% of problems (the number of known problems q is 6).

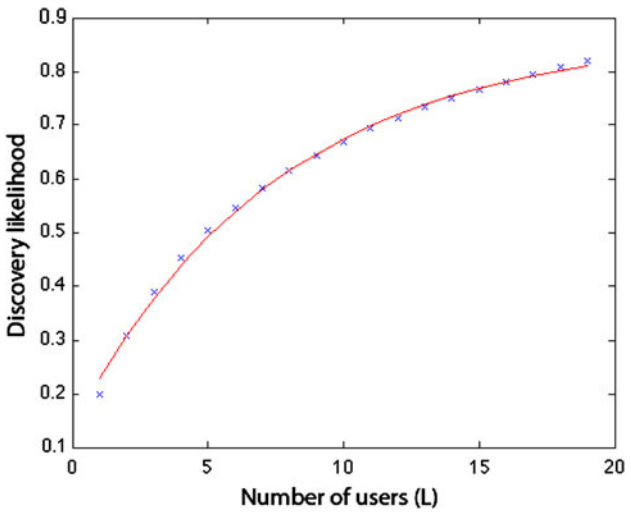

Fig. 6 Discovery likelihood of users with TA estimated by the bootstrap analysis

Table 2 The values of the parameters a, p and q calculated by the bootstrap of users data

Parameters	Values	Confidence interval
a	0.8691	(0.8440–0.8942)
p	0.1235	(0.1116–0.1355)
q	2.3910	(2.0980–2.6850)

The convergent validity of the BDB model

In order to verify the significance of the results obtained through the BDB model,[2] a convergent validity test was carried out by comparing the results with those obtained using the MC method (Lewis 2001; 2006).

The results (Tables 3 and 4) show that there is barely an overlap between the λ values obtained using the two techniques. These results confirm the validity of the BDB model with respect to the MC method.

The aim of this work is not to provide a new method for estimating the value of λ, since the BDB model does not discuss the discovery rate obtained using practitioners' data. The BDB model should be applied when the λ value has already been estimated using a test with three or five users, and for calculating how many users are required in order to detect more than 80% of the errors in a target interface. By doing so, the BDB model enlarges the perspective of analysis by adding two new parameters which are not considered in the classic estimation model: this model considers all of the possible discovery behaviours of participants (a) and encompasses a rule for the selection of representative data (q). These parameters take into account the variability of the different interfaces (q) and the different behaviours of the samples used in a usability study (a). However, our model does not supersede the λ value estimation obtained using the classic ROI model or by GT adjustment. In fact, by using the BDB model, practitioners might receive confirmation that the number of users/experts involved in their sample test is already sufficient for a reliable evaluation.

Discussion

The results given by the BDB are very different from those obtained by the ROI model. For the ROI model, the estimation of the costs of the usability evaluation required a sample composed of 6 experts and one of 7 users, while applying the BDB the estimation shows that a sample of 13 experts and one of 20 users are needed in order to identify more than 80% of problems. As a consequence, in following the BDB model, there is an increase in the usability evaluation costs with respect to the data provided by the

[2] In the review phase of this work, a reviewer claimed that "The authors should do a Monte Carlo resampling exercise to assess the extent to which randomly selected sets of 6 experts (for the CW data) and 7 users (for the TA data) find or fail to find at least 80% of the problems discovered by the full samples," since, according to the reviewer's opinion, "The authors simply state the different sample size estimates and appear to assume that the BDB are correct without further, evaluation or any tests of significance". In accordance with the reviewer's suggestions, we have added this section.

Table 3 A comparison of the value of λ as obtained using BDB model with the ones obtained using MC resampling with 3, 6, 10 and 20 users

BDB λ value with 20 users	MC λ value with 20 users	MC λ value with 10 users	MC λ value with 6 users	MC λ value with 3 users
0.123	0.119	0.243	0.266	0.364

Table 4 Comparing the value of λ as obtained by BDB model with the ones obtained using MC resampling with 3, 6 and 13 experts

BDB λ value with 13 experts	MC λ value with 13 experts	MC λ value with 6 experts	MC λ value with 3 experts
0.141	0.165	0.235	0.313

ROI model. However, the increase in costs enlarged the evaluator perspective by providing a more reliable set of results. Actually, the BDB approach allows the behaviour of the whole population (parameter a), the representativeness of the sample data (i.e. the problems found expressed by the parameter p) and the different properties of the interface (parameter q) to be taken into account.

Conclusion

The BDB, while respecting the assumption of the ROI model, opens a new perspective on the discovery likelihood and on the costs of usability evaluation. Indeed, the possibility of considering both the properties of the interface and the representativeness of data grants the practitioner a representative evaluation of the interface. A practitioner can run a test by applying the BDB model after the first five experts and users i.e. the ROI model in order to estimate the parameters a, p, and q and the number of evaluators needed for an evaluation that considers the specific properties of the interface and the representativeness of the sample. In this sense, in the evaluation a practitioner can take into account both the BDB model and the ROI one. Our perspective offers a new model for the usability evaluation, guaranteeing the representativeness of the data and overcoming the deficiencies of the ROI model. In this sense, the increase in costs is justified by the possibility of obtaining representativeness of the entire potential population with a small sample.

Appendix: User scenarios

1. A friend of yours is enrolled on a 1-year activity in social service. You are interested in finding more information about social service activities and acquiring information in order to apply for a one-year job. Go to the website http://www.serviziocivile.it/, find that information and download the documents for the job application.
2. A friend of yours, who lives in Rome, has some internet connection problems, so he or she telephones you for assistance. In fact, he or she is interested in social service work, but he or she does not know where the office is and when it is open in order to present his curriculum vitae. Go to the website http://www.serviziocivile.it/ in order to find that information for him or her.
3. You are interested in social service activities, so you go to the website http://www.serviziocivile.it/ in order to see whether this website offers a newsletter service, even though you are not enrolled on the social service activities. If the newsletter service requires you log in, sign up to the newsletter.
4. A friend of yours is working on a 1-year social service project in the Republic of the Philippines. You are interested in applying for a job on this project. Go to the website http://www.serviziocivile.it/ in order to find information about the project and whether is possible to obtain a job.

References

Bellman JR, Park CW (1980) Effects of prior knowledge and experience and phase of the choice process on consumer decision processes: a protocol analysis. J Consum Res 7(3):234–248

Bettman JR (1979) An information processing theory of consumer choice. Addison-Wesley, Cambridge

Biehal G, Chakravarti D (1982a) Experiences with the Bettman-park verbal-protocol coding scheme. J Consum Res 8(4):442–448

Biehal G, Chakravarti D (1982b) Information-presentation format and learning goals as determinants of consumers' memory retrieval and choice processes. J Consum Res 8(4):431–441

Biehal G, Chakravarti D (1986) Consumers' use of memory and external information in choice: Macro and micro perspectives. J Consum Res 12(4):382–405

Biehal G, Chakravarti D (1989) The effects of concurrent verbalization on choice processing. J Mark Res 26(1):84–96

Borsci S, Federici S (2009) The partial concurrent thinking aloud: a new usability evaluation technique for blind users. In: Emiliani PL, Burzagli L, Como A, Gabbanini F, Salminen A-L (eds) Assistive technology from adapted equipment to inclusive environments—AAATE 2009, vol 25. Assistive technology research series. IOS Press, Florence, pp 421–425. doi:10.3233/978-1-60750-042-1-421

Caulton D (2001) Relaxing the homogeneity assumption in usability testing. Behav Inf Technol 20(1):1–7. doi:10.1080/01449290010020648

Efron B (1979) Bootstrap methods: another look at the jackknife. Ann Statist 7(1):1–26. doi:10.1214/aos/1176344552

Federici S, Borsci S (2010) Usability evaluation: models, methods, and applications. International encyclopedia of rehabilitation. Center for International rehabilitation research information and exchange (CIRRIE), Buffalo. http://cirrie.buffalo.edu/encyclopedia/article.php?id=277&language=en. Accessed 20 Sept 2010

Federici S, Borsci S, Mele ML (2010a) Usability evaluation with screen reader users: a video presentation of the pcta's experimental setting and rules. Cogn Process 11(3):285–288. doi:10.1007/s10339-010-0365-9

Federici S, Borsci S, Stamerra G (2010b) Web usability evaluation with screen reader users: implementation of the partial concurrent thinking aloud technique. Cogn Process 11(3):263–272. doi:10.1007/s10339-009-0347-y

Fox J (2002) An r and s-plus companion to applied regression. SAGE, California

Goodstein RL (1963) Boolean algebra. Pergamon Press, Oxford

Green A (1995) Verbal protocol analysis. Psychologist 8(3):126–129

Hertzum M, Jacobsen NE (2003) The evaluator effect: a chilling fact about usability evaluation methods. Int J Hum Comput Interact 15(4):183–204. doi:10.1207/S15327590IJHC1501_14

Kuusela H, Pallab P (2000) A comparison of concurrent and retrospective verbal protocol analysis. Am J Psychol 113(3):387–404

Kuusela H, Spence MT, Kanto AJ (1998) Expertise effects on prechoice decision processes and final outcomes: A protocol analysis. Eur J Mark 32(5/6):559

Lewis JR (1994) Sample sizes for usability studies: additional considerations. Hum Factors 36(2):368–378

Lewis JR (2001) Evaluation of procedures for adjusting problem-discovery rates estimated from small samples. Int J Hum Comput Interact 13(4):445–479

Lewis JR (2006) Sample sizes for usability tests: mostly math, not magic. Interactions 13(6):29–33. doi:10.1145/1167948.1167973

Lewis C, Rieman J (1993) Task-centered user interface design: a practical introduction. http://users.cs.dal.ca/~jamie/TCUID/tcuid.pdf. Accessed 20 Jun 2010

Nielsen J (2000) Why you only need to test with 5 users. www.useit.com/alertbox/20000319.html. Accessed 20 Jun 2010

Nielsen J, Landauer TK A mathematical model of the finding of usability problems. In: Proceedings of the INTERACT '93 and CHI '93 Conference on human factors in computing systems, Amsterdam, 24–29 Apr 1993. ACM, New York, NY, USA, pp 206–213

Nielsen J, Mack RL (eds) (1994) Usability inspection methods.Wiley, New York

Polson PG, Lewis C, Rieman J, Wharton C (1992) Cognitive walkthroughs: a method for theory-based evaluation of user interfaces. Int J Man Mach Stud 36(5):741–773. doi:10.1016/0020-7373(92)90039-N

Rieman J, Franzke M, Redmiles D Usability evaluation with the cognitive walkthrough. In: Conference companion on human factors in computing systems, Denver, Colorado, United States, 1995. ACM, 223735, pp 387–388. doi:10.1145/223355.223735

Schmettow M Heterogeneity in the usability evaluation process. In: Proceedings of the 22nd British HCI group annual conference on people and computers: culture, creativity, interaction—Volume 1, Liverpool, United Kingdom, 2008. British Computer Society, 1531527, pp 89–98

Spool J, Schroeder W Testing web sites: Five users is nowhere near enough. In: CHI '01 extended abstracts on human factors in computing systems, Seattle, Washington, 2001. ACM, 634236, pp 285–286. doi:10.1145/634067.634236

Turner CW, Lewis JR, Nielsen J (2006) Determining usability test sample size, vol 2. International encyclopedia of ergonomics and human factors, Second edn. CRC Press, Boca Raton

Virzi RA (1990) Streamlining the design process: running fewer subjects. Human factors and ergonomics society annual meeting proceedings 34:291–294

Virzi RA (1992) Refining the test phase of usability evaluation: how many subjects is enough? Hum Factors 34(4):457–468

Wharton C, Rieman J, Lewis C, Polson PG (1994) The cognitive walkthrough method: a practitioner's guide. In: Nielsen J, Mack RL (eds) Usability inspection methods. Wiley, New York, pp 105–140

Woolrych A, Cockton G Why and when five test users aren't enough. In: Vanderdonckt J, Blandford A, Derycke A (eds) Proceedings of IHM-HCI 2001 conference, Toulouse, FR, 10–14 Sept 2001. Cépadčus Éditions, pp 105–108

Wright PC, Monk AF (1991) A cost-effective evaluation method for use by designers. Int J Man Mach Stud 35(6):891–912. doi:10.1016/s0020-7373(05)80167-1