How Can You Do a Card-sorting Study with LOTS of Cards?

PRESENTATION SUMMARY

Confirmation of ability to present:	[x] I will be able to present my poster at the opening reception. I will also choose two 15-minute timeslots on Wednesday and Thursday to discuss my poster
Interest in the Poster Revolution	Yes
Topic Category:	[ ] Usability perspectives [x] Methods and skills [ ] Business and organization [ ] Accessibility and Internationalization [ ] "Outside the box"
Presentation Type	[ ] Business case study [x] How-to Discussion [ ] Overview of a concept, philosophy or methodology [ ] Presentation of design approach or guidelines [ ] Other: _____
Intended Audience:	[x] Everyone [ ] Individuals New to Usability [ ] Experienced Practitioners [ ] Technical/Professional Leadership
Audio Visual Requirements	No special requirements.

ABSTRACT

One of the drawbacks of card-sorting is that it can be very time-consuming for the participants. We present and validate a technique in which each participant only sorts a randomly selected subset of the full set of cards. Through the data analysis, however, you can still get an accurate picture of the relationships among *all* the cards.

GOALS FOR THE POSTER

- Enable anyone who needs to do a card-sorting study with a relatively large number of cards to do the study without exhausting each participant!
- Demonstrate that this sampling technique yields results equivalent to those obtained from the traditional technique where each person sorts all the cards, as long as there are enough participants.

PREVIOUS PUBLICATION OR USE OF THIS MATERIAL

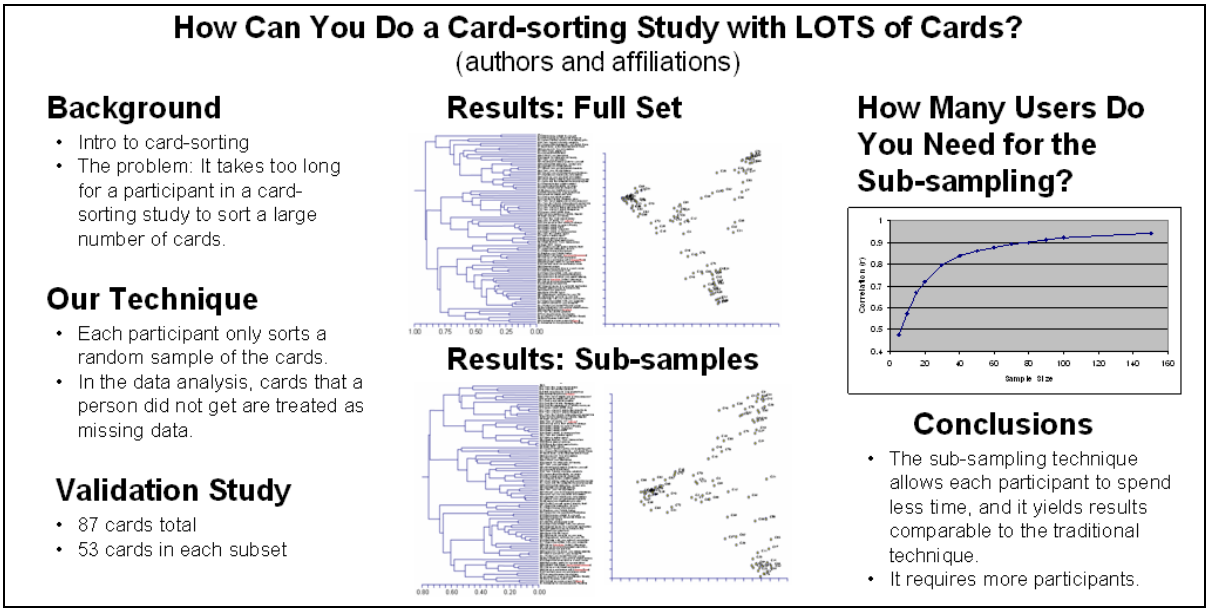
This material has not been presented or published before.

YOUR BACKGROUND IN THIS MATERIAL

The first presenter manages the usability department of a large company. He has a PhD in Engineering Psychology and MA in Experimental Psychology. He has over 25 years of experience in the usability field, having managed usability groups at several other companies prior to his current position. He has published over 30 technical articles in the usability or human factors field. He is an active member of UPA, ACM SIGCHI, and HFES. He has been conducting card-sorting studies since the early 1980's.

The second presenter is a principle in a small company dedicated to the development of tools for Web site design and development. One such tool supports on-line card-sorting. He is also a professor emeritus of Psychology at a major university, where he taught HCI and UCD courses for 20 years. In addition, he has published articles on UCD and edited a book on user interface design. He served on the UPA Board of Directors from 2000-2003.

POSTER LAYOUT



DETAILED DESCRIPTION OF POSTER CONTENT

Introduction

Card-sorting is a common technique for learning the logical relationships that users perceive among a set of objects. For example, it is commonly used for input to the organization of a new website. In that case, the cards would provide brief descriptions of each candidate page in the site. The participants in the card-sorting study would then sort the cards into the logical groups that they perceive, and, typically, name those groups. Various analysis techniques, such as hierarchical cluster analysis and multi-dimensional scaling, can then be used to get a sense of the relationships among the cards.

Card-sorting studies originally were literally that: manual exercises in which participants sorted physical cards (typically index cards) on a table. Today, a variety of online tools are available for conducting card-sorting studies (e.g., EZSort, WebSort, WebCAT, Socratic CardSort, Classified, CardZort). These typically provide some type of online interface for dragging and dropping representations of the cards into various categories. Both manual and online techniques are still used today, although the advantages of the online method seem to outweigh its disadvantages.

One drawback of the card-sorting technique, whether manual or online, is that it can take a relatively long time for any one participant to sort a large number of cards. Our experience with many card-sorting studies has been that the amount of time it takes any one participant to sort the cards tends to increase in

a linear fashion with the number of cards. Although the time per card clearly varies depending upon the complexity of the material being sorted, we have found that about one minute per card is a reasonable "rule of thumb". So the number of cards that can be included in a study is usually limited by how much time you can convince each participant in the study to devote to the exercise. Obviously one thing that influences how much time the participants are willing to devote is how motivated they are. One common technique to motivate participants is to provide some kind of incentive, such as money, the chance to win a prize, etc. We have seen cases where participants are only willing to spend 20-30 minutes, and other cases where they would devote up to 1 ½ hours. But the former is much more common than the latter. In fact, with online card-sorting studies in particular, we have found that it is far easier to get many participants to do a relatively short exercise (e.g., 15-20 minutes) than it is to get even a small number of participants to do a long exercise (e.g., 1 hour).

As an example, assume you decide that, given the motivation of the participants in your card-sorting study, you expect them to be willing to spend up to 30 minutes on the exercise. Using our rule of thumb, that means you can only ask them to sort about 30 cards. But what if you have 50 pages for the website that you are designing, and consequently 50 cards? Previously, the only solutions we have seen involve "winnowing" the full set of cards down to a manageable size—30 in this case. You then have to make assumptions about how the deleted cards would have been sorted, had they been included.

Our Technique

The technique we have developed for dealing with these cases of larger numbers of cards is to have each participant in the study sort a randomly selected subset of the total set of cards. Our initial test, to be described in this poster, used a 60% sample for each participant. So, for example, if you had 100 cards total, each participant would sort 60 cards randomly selected from the 100. Clearly, larger samples would also work (e.g., 80%); we have yet to determine whether smaller samples would work (e.g., 40%).

After a participant does their card-sorting exercise, a "distance matrix" for that participant is built. The entries in this matrix reflect the "perceived distance", for that participant, among all pairs of the full set of cards. If two cards were placed in the same group by that participant, that pair gets a distance of 0. If two cards were placed in different groups, that pair gets a distance of 1. In addition, if *either* member of the pair was not included in the cards that participant sorted, that pair is assigned a "missing" value. After these matrices are built for each participant, they are then averaged across all the participants. As with any calculation of an average, cells with "missing" values are not included in the calculation of the average—they neither add to it nor detract from it. The hope is that you have enough participants in the study that all cells (i.e., all pairs of cards) get a meaningful average (i.e., at least some participants did get that particular pair as a part of their random subset). Obviously this could be a problem with a small number of participants and a large number of cards. Once this average matrix has been created, the data can be analyzed using the same techniques as with a traditional card-sorting exercise.

The Validation Study

Although this technique seems sound, in principle, we wanted to validate it with a real card-sorting study. We chose a card-sorting study being done in support of a redesign of the Intranet for the first presenter's corporation. A total of 87 cards were identified for the exercise. Some examples are as follows:

- Submit an Invoice (Accounts Payable)
- Download anti-virus software & definitions
- Replace your ID badge
- Change your healthcare provider
- Get a business credit card
- Find and reserve a conference room
- Request tuition reimbursement
- Report a problem with your computer

The study was conducted online using a specially modified version of WebSort that allows for each participant to receive a random subset of the full set of cards. For this test, we decided to use a 60%

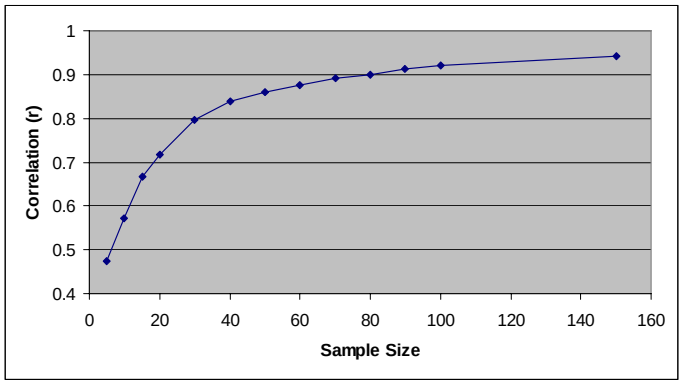
subset, or 53 cards randomly selected from the full set of 87. In order to validate the sub-sampling technique, we actually did two card-sorting studies: a traditional card-sorting study where each participant sorted the full set of 87 cards, and the sub-sampling study where each participant received a randomly selected set of 53 cards to sort. The participants in both studies were employees of the first presenter's company. They were recruited via an email message that is sent every morning to all employees which included a link to the study. For each study, the incentive for participation was entry in a drawing for a $100 gift check. Once the average distance matrix had been created for each study, the data were analyzed using hierarchical cluster analysis and multi-dimensional scaling.

Results

A total of 367 people completed the card-sorting exercise using the full set of 87 cards. A total of 343 people completed the card-sorting exercise using the sub-sampling approach where each person sorted 53 cards. (The study using the full set of 87 cards was posted online for about five business days, while the sub-sampling study was posted for only one. Our goal was to get approximately equal numbers of participants in both studies.)

In the actual poster, we will present the results of the hierarchical cluster analysis and multi-dimensional scaling for both studies so that they can be compared to each other. The two sets of results are virtually identical to each other. The correlation between the two average distance matrices is r=.97, which, of course, is highly significant (p<.001).

We also analyzed the data from the sub-sampling study to determine how large a sample is needed in order to get reliable data (i.e., comparable to the full set). From the 343 participants in the sub-sampling study, we took random samples ranging from n=5 to n=150, calculated the average distance matrices for at least five such samples, and correlated those with the average distance matrix from the full study. Those correlations are shown in the graph below. As expected, the sub-sampling approach requires more participants to get reliable data than the traditional approach. In fact, the correlations do not reach r=.90 until a sample size of 80. However, one can also argue that the return, in terms of increased correlation, begins to diminish somewhere around a sample size of 30-40. This can be compared to the findings of Tullis and Wood (2004), who showed that using the traditional card-sorting technique, the correlations reach r=.90 at a sample size of 15.



References

<reference suppressed for anonymity>.

EZSort: http://www-3.ibm.com/ibm/easy/eou_ext.nsf/Publish/410
WebCAT: http://zing.ncsl.nist.gov/WebTools/WebCAT/overview.html
WebSort: http://www.websort.net/
Socratic CardSort: http://www.sotech.com/main/eval.asp?pID=123
Classified: http://www.infodesign.com.au/usabilityresources/classified/
CardZort: http://condor.depaul.edu/~jtoro/cardzort/index.htm