Different or alike? Comparing computer-based and paper-based card sorting

Goetz Greve\textsuperscript{1,a}

\textsuperscript{1}Marketing & Sales Department, HSBA Hamburg School of Business Administration, Hamburg, Germany

\textsuperscript{a)Corresponding author: Goetz.Greve@hsba.de}

Abstract: In times of increasing importance of e-commerce solutions, designing for usability has been one focal point of interest of human-computer interaction and marketing research. While internet usability research is progressing tremendously, research on intranet usability is rather scarce. But especially in the firm context, poor usability of intranet portals imposes significant costs on companies as employees need more time obtaining the information they need for their work. The author conducted a laboratory experiment to compare paper-based and computer-based card sorting for usability evaluation and come to the conclusion that there is a significant difference of results regarding the application of the two compared methods optimizing an intranet portal and derive implications for usability testing.

Keywords: Card sort, closed card sorting, information architecture

1. Introduction

Nowadays, employees spend increasing amounts of time and effort interacting with complex web pages on the company-owned intranet. An Intranet is a privately maintained network that can be accessed only by authorized persons, especially employees of the organization that owns it. Intranet portals (or enterprise information portals, corporate portals) can be defined as applications that enable companies to unlock internally and externally stored information, and provide users a single gateway to personalized information needed to make informed business decisions [20]. The intranet portal client is a web browser. Users simply receive a customized homepage that provides a single, integrated view of a wide range of information and presents users with a summary of reports, documents and other information objects [10].

Implementing new software systems or applications in organizations can be a difficult task, and many systems fail to fulfill their goal or have serious limitations, e.g. lack of adequate support to the core tasks of the user or unsuitable designs of user interaction and interfaces [11]. In addition to this, poor usability of intranet
portals imposes significant costs on companies as employees need more time obtaining the information they need for their work. These outcomes can be avoided by applying techniques of usability evaluation during intranet page development or optimization. To a company developing intranet systems, for example, increased employee productivity and efficiency may translate into lower personnel costs as well as reduced training costs. In most cases usability engineering can help reduce development, maintenance, and support cost and shorten development cycles as well [6]. In contrast to this, a study of Forrester Research [23] revealed, that only about half of the surveyed companies were compliant with simple web usability principles such as ‘is the site organized by user goals?’

A way to optimize web site usability is to apply methods of usability testing, e.g. heuristic evaluation, cognitive walkthroughs or card sorting. Usability testing can lead to significant cost due to large sample sizes and offline test conditions. However, the introduction of intranet pages may not be perceived as beneficial enough to justify high introductory costs. As a result of this the development of intranet systems requires low-cost development processes that return high-usability solutions. Thus, companies aim for testing usability within a computer-based environment to easily conduct the test at the workplace of their employees instead of inviting employees to usability testing workshops. Card sorting is one usability testing method that is applicable to a computer-based test environment as well as to a paper-based test environment [19]. Unfortunately, the development and practice of card sorting in information technology has been driven mostly by opinion and anecdotal experience, with little influence from systematic research [28]. Thus, the current research is attempting to compare two different methods for optimizing intranet site usability conducting an experiment and analyze the results to derive managerial and scientific implications.

2. Card Sorting Techniques

Regarding the usability dimension ease-of use information architects and developers of desktop and web-based software applications are faced with the problem of organizing information items, features, and functions to make it easier for users to find them. One way of analyzing human behavior or individual cognitions regarding the categorization of objects is by understanding the categories people use and how they assign concepts to those categories [2]. Cognitive theories assume that individuals employ particular data structures in memory to understand their world [7] [12] on the basis of their experience and particular context [22].

One method to test how people organize and categorize their knowledge is card sorting. Card sorting is a simple technique in usability testing where a group of subject experts or end users are guided to organize content from a web site by grouping items into categories as it makes sense to them [4] [15] [16]. It is a useful approach for optimizing navigability, designing menu structures or workflows. Card sorting can be used to organize an initial website as well as optimizing existing structures.

The paper-based method originally consisted of researchers writing labels representing concepts (either abstract or concrete) on cards, and then asking participants to sort (categorize) the cards into piles that were similar in some way. However, the more concrete and specific the elements are, the more familiar the participant is with the elements and the more likely that s/he will produce much richer sorts. A participant’s direct experience may produce more specific sort. After sorting the cards into piles, the participants are asked to give the piles a name or phrase that would indicate what the concepts in a particular pile had in common.
Basically, two primary methods for performing card sorting exist that can be conducted in a paper-based or computer-based test environment: First, by using an open card sorting approach, participants are given cards showing site content with no pre-established groupings. They are asked to sort cards into groups that they feel are appropriate and then describe each group. Open card sorting is useful as input to information structures in new or existing sites and products. Second, by applying the method of closed card sorting, participants are given cards showing site content with an established initial set of primary groups (e.g. categories). Participants are asked to place cards into these pre-established primary groups. Closed card sorting is useful when adding new content to an existing structure, or for gaining additional feedback after an open card sort.

Paper-based card sorting experiments need a physical test environment, e.g. a testing facility or rooms prepared for conducting a card sorting experiment. It also allows for observing behavior of participants by a moderator or for interacting with participants, e.g. answering questions or giving insights.

Computer-based card sorting experiments are conducted with the help of a computer, providing an electronic data set of cards and categories. Participants are invited to perform the card sorting on the computer. It allows for comfortably gathering a larger data set of participants, thus providing higher statistical validity to the sample. Compared to paper-based card sorting, it offers an easier and less costly way to reach a broader base of users. To conduct a computer-based card sorting, a variety of web-based card sorting software tools are available, e.g. OptimalSort [17] or CardZort Zone [3] for closed card sorting and Socratic Card Sort [21], and WebSort [27] for both open and closed card sorting.

Nevertheless, if computer-based card sorting offers these advantages, the question arises: Does it lead to the same results than paper-based card sorting or is there a tradeoff between the two approaches?

3. Study Design

For the comparison of computer-based and paper-based card sorting as methods for optimizing intranet usability we conducted a laboratory card sorting experiment as close to real life as possible. A sample of end users is split into an experimental group conducting a computer-based closed card sorting and a control group conducted a paper-based closed card sorting. Each group performed the same card sorting, assigning 51 cards to 13 categories. The conceptual model of our research framework is depicted in Figure 2.

Principally, both approaches, computer-based and paper-based card sorting, are conducted in the same way. Thus, we assume that there is no difference between conducting a paper-based card sorting by hand or computer-based. This outcome would ensure that computer-based usability testing results are equal to paper-based usability testing results of card sorting and consequently can be used as a substitute of the paper-based approach. This finding would allow companies to conduct usability tests computer-based instead of inviting participants to a testing facility. Facility rental fees can be saved, participation fees tend to be lower, thus cost savings can be realized [18]. Thus, we hypothesize the following:

**H0:** There is no significant difference between computer-based and paper-based card sorting methods regarding the mapping of cards.

**H1:** There is a significant difference between computer-based and paper-based card sorting methods regarding the mapping of cards.
For further testing the differences between the two approaches, two measures can be taken into account. First, the consistency score measures the proportion of cards assigned to the same category / different category. As in a closed card sorting predefined categories are available, it can be reckoned that the majority of probands will act in the same way. Thus, this outcome should be equal for both, paper-based as well as computer-based card sorting.

**H2: There is a significant difference between computer-based and paper-based card sorting methods regarding the consistency of assigned cards to categories.**

Second, the complexity score measures the proportion of assigned cards to non-assigned cards to categories. The higher the proportion is, the higher the complexity of the card sorting task. Again, the complexity score of paper-based card sorting should comply with the complexity score of the computer-based card sorting. Thus, we hypothesize that:

**H3: There is a significant difference between computer-based and paper-based card sorting methods regarding the relationship between assignment and non-assignment of cards to categories.**

To test our hypotheses we develop two different performance measures, the complexity score and the consistency score and test the two groups of the experiment for significant difference using the Fisher’s exact test.

### 4. Data Collection

In order to empirically test our conceptual model, we collected data with the help of a global provider of aircraft maintenance, repair and overhaul that wanted to optimize its existing intranet structure. Thus, our hypotheses have been tested under real-life conditions. For the experiment we choose a closed card sorting approach as the company had already implemented an intranet portal.
closed sorting approach items are assigned to pre-defined categories. With this method it can be discovered how users sort content items into each category. This approach is useful to optimize an existing intranet structure by validating that users consistently interpret the category names and groupings as it was intended.

The data collection process has been conducted as follows: First, all participants for the experiment have been recruited by a self-administered form via the landing pages of the company's intranet. 251 self-selected applications have been received. 64 participants have been randomly selected from this sample and have been randomly divided into an experimental group (32 participants), performing the paper-based card sorting and a control group (32 participants), performing the computer-based card sorting. To ensure the quality of the experiment, the number of participants is well above the suggested minimum of 20-30 participants that Tullis and Wood [25] suggested.

Data has been collected in one single week, conducting both, the paper-based and the computer-based card sorting. For the experiment group conducting the paper-based card sorting all items have been printed on separate cards as well as all categories have been printed on separate cards. In total 32 sets of cards and categories have been produced, one set for each participant of the experiment group. A complete list of all items and categories can be found in table 2. All participants are end-users of the company’s intranet system with direct experience from working with the system. These participants will clearly have an edge to other potential evaluators, such as usability experts, when it comes to knowledge how the intranet portal should be organized.

<table>
<thead>
<tr>
<th>Cards</th>
<th>Customer Events</th>
<th>IQ Move @ TS</th>
<th>Pictures &amp; Graphics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advertising</td>
<td>Customer Questionnaires</td>
<td>LHT Connection</td>
<td>Presentations</td>
</tr>
<tr>
<td>Ameco Beijing</td>
<td>Customer Satisfaction Survey</td>
<td>LHT Group Events</td>
<td>Product Catalogue</td>
</tr>
<tr>
<td>Application Styleguide</td>
<td>Design</td>
<td>LHT Logistics</td>
<td>Ready to Sign</td>
</tr>
<tr>
<td>Archive</td>
<td>Design Manual</td>
<td>LHT Philippines</td>
<td>Sounds</td>
</tr>
<tr>
<td>Branding &amp; Naming</td>
<td>Event Feedback</td>
<td>LHT Products</td>
<td>Trademarks</td>
</tr>
<tr>
<td>Brochures</td>
<td>Event Management</td>
<td>Market Outlooks</td>
<td>TS/B - Sales Service</td>
</tr>
<tr>
<td>Calendars</td>
<td>Event Registration</td>
<td>Market Studies</td>
<td>TS/M - Marketing</td>
</tr>
<tr>
<td>Christmas Card</td>
<td>Fairs</td>
<td>Miscellaneous</td>
<td>TS/X - Key Account VIP</td>
</tr>
<tr>
<td>Christmas E-Card</td>
<td>Giveaways</td>
<td>Online Research</td>
<td>Video clips</td>
</tr>
<tr>
<td>Company Logos</td>
<td>Internet</td>
<td>Past events</td>
<td>Weekly News</td>
</tr>
<tr>
<td>Competitive Competence (Co-Co)</td>
<td>Internet Styleguide</td>
<td>Peoplesoft</td>
<td>Weekly Sound</td>
</tr>
<tr>
<td>Conferences</td>
<td>Intranet</td>
<td>Picture Database</td>
<td></td>
</tr>
</tbody>
</table>

At the start of the test, participants are instructed to sort the cards by matching them to one of the categories. A pile of randomly sorted cards has been distributed to each participant. Additionally, the 13 categories have been randomly arranged in front of each participant. The basic data collected is simply which cards each user put into which categories. 32 participants took part in the online computer-based card sorting experiment. For the experimental group we chose to use the software OptimalSort [17].
5. Results

Based on the collected data two scores have been calculated: First, we analyzed how many participants have assigned the cards to the same category in relation to the participants assigning cards to a different category. The higher the percentage per card is, the lower the consistency of assignments. Thus, the consistency score shows the goodness of fit of the usability design for the sampled participants based on the applied card sorting method. With 49 of 51 cards assigned differently, computer-based card sorting seems to work in a less optimal manner than paper-based card sorting with 44 cards differently assigned. The results of the computer-based experiment are in line with Tien et al. [24], who stated that the absence of paper-based cards in a computerized test environment can influence the cognitive performance of the participants. Additionally, we assumed that the procedural variations between the two forms of administration (paper-based / computer-based) do substantially affect the outcome of the test [9]. Second, we calculated the overall assignment of cards to categories in relation to the non-assignment. The higher the percentage per card is the higher the complexity of the assignment. Thus, the complexity score shows a general lack of usability regarding certain cards and categories. The overall results do not show high failure rates, thus it can be assumed that the general overall complexity of the cards is rather low (see Appendix A). Nevertheless, it can be noted that the computer-based card sorting results in higher rates of non-assigned cards to categories.

To compare for significant differences between computer-based and paper-based card sorting, we calculated the mean of the assigned cards to categories and run a t-test, resulting in a low significant difference (p = .057) between computer-based and paper-based card sorting. Thus, we have to reject H0 for the overall experiment.

To compare both card sorting methods analyzing the observed frequencies of this multi-cell classification with those expected from our hypotheses H2 (consistency) and H3 (complexity), normally, Pearson’s chi-square (\( \chi^2 \)) test would be first choice. It tests a null hypothesis that the frequency distribution of certain events observed in a sample is consistent with a particular theoretical distribution. However, many authors have shown that for small sample sizes the Yates-correction for continuity [29] or the Fisher’s exact test should be preferred [26, 5]. Fisher’s exact test [8] is a statistical significance test used in the analysis of contingency tables where sample sizes are small. It is called exact test, because the significance of the deviation from a null hypothesis can be calculated exactly, rather than relying on an approximation that becomes exact in the limit as the sample size grows to infinity, as with many statistical tests. A rule of thumb is that the expected count in each cell of the table should be greater than 5 before Pearson’s chi-squared test is used. Therefore, we choose the Fisher’s exact test to analyze the experimental and control group.

We tested the hypothesis based on the significance of the p-values of the Fisher’s exact test in Appendix A. Regarding the consistent assignment of items (consistency score) the level of significance is < .05 for 6 items, which accounts for approx. 12% off all items. Thus, we have to reject H2. Hence, we have to admit a significant relationship between the card sorting method and differently assigned items. Comparing computer-based and paper-based card sorting it can be seen that the number of differently assigned cards is higher within online card sorting than offline card sorting. It could be concluded, that computer-based card sorting leads to a higher number of different possibilities how to assign cards to categories than offline card sorting. This finding may be based on the different possibilities of paper-based
and computer-based card sorting to visualize the sorting task. The level of significance regarding the assignment / non-assignment of items is < .05 for one item, which is approx. 2 % off all items (complexity score). Thus, we also have to reject H3. Looking at the percentage of non-assigned cards, computer-based card sorting delivers a higher number of non-assigned cards.

6. Managerial Implications

Several interesting insights and managerial implications can be derived from our findings and can help address management’s common misconceptions. Even if computer-based card sorting offers at first sight advantages of cost and time saving our findings indicate that computer-based and paper-based card sorting does not necessarily lead to equal results. This finding is, on the one hand, in line with Molich et al. [13] and Molich and Dumas [14] for a common comparison of usability testing methods. But on the other hand, is in contrast with the findings of Bussolon et al. [1], who compared computer-based and paper-based card sorting techniques and did not reveal any differences in results. We suggest that especially complex sorting tasks with demanding card associations could be better done with the help of paper-based card sorting as participants are able to arrange cards and categories in a way they think. Also, paper-based card sorting together with the presence of an administrator offers the possibility for probands to ask for help. The complexity of interpersonal interaction between subject and administrator cannot be replaced by a computerized test [24]. However, one should keep in mind that helping participants could also lead to biased results that are not exclusively optimized by the user. It should be noticed, that the differences between paper-based and computer-based card-sorting are significant, but e.g. regarding the consistency of the sorting task, approx. 6% of items differently assigned, the deviation is rather low.

Nevertheless, a computer-based card sorting may be useful for carrying out studies of large samples. It also allows a higher degree of standardization across administration by different examiners and at different locations. Thus, utility and practicability of employing computerized card sorting seems to be obvious. However, the results make it clear, that a computerized card sorting test should not be administered to participants alone in a room or even at their desktop PC in the office. Rather, participants need to be monitored during the test as well as e.g. given verbal instructions or the possibility to ask questions.

7. Conclusion

This paper contributes to the existing literature on card sorting techniques by providing insights into the comparability of paper-based and computer-based card sorting. In this article, we have, first, developed a conceptual framework for comparing two different methods of card sorting for usability optimization, second, empirically tested that framework; and third, found out that we can assume differences between the application of the two tested card sorting methods. Thus, we show what one may realistically expect from applying paper-based and computer-based card sorting techniques in regard to comparability.

However, this study opens a variety of avenues for future research: First, data were collected from a sample of end users of one company in one industry. This may restrict the applicability of the results of the experiment to other populations. However, since the end users are intranet users of one firm, and our research scenario was designed to induce the participants into the experience of conducting sorting tasks, the use of an end user sample should not present a serious threat to
the validity of this study. Second, it can be assumed that procedural variations of administration between paper-based card sorting with possible administrative help and computer-based card sorting without can lead to differences in the outcome of the test. Third, it can be assumed, that cultural diversity in multinational companies can lead to different outcomes regarding the usability of intranet pages from country to country. Fourth, future research could use our complexity and consistency score as a starting point for developing metrics of web site complexity. More extensive studies are needed to test the effects of different combinations of textual categories, graphics, icons, links and additional design factors on perceptions of intranet usability.

Acknowledgments

The author would like to thank Mrs. Daniela Wanner and an anonymous company for supporting the data collection process.

Appendix

Table 2. Results of the closed card sorting experiment
References


