Redesign Based on Card Sorting: How Universally Applicable are Card Sort Results?

Jobke Wentzel[1(✉)], Nienke Beerlage de Jong[2], and
Thea van der Geest[1]

[1] Department of Media, Communication and Organisation,
University of Twente, Enschede, The Netherlands
{m.j.wentzel, t.m.vandergeest}@utwente.nl
[2] Department of Psychology, Health and Technology, University of Twente,
Drienerlolaan 5, 7522 NB Enschede, The Netherlands
n.beerlage-dejong@utwente.nl

Abstract. Card sort studies can facilitate developers to create an information structure for their website or application. In addition, this human-centered design method provides researchers with insights into the target group's mental models regarding the information domain under study. In this method, participants sort cards, with excerpts of the website's or information source's information on them, into piles or groups. Even though the method lends itself for large numbers of participants, it can be difficult to include sufficient participants in a study to ensure generalizability among large user groups. Especially when the potential user group is heterogeneous, basing the information structure on a limited participant group may not always be valid. In this study, we investigate if card-sort results among one user group (nurses) are comparable to the results of a second (potential) user group (physicians/residents).

The results of a formative card sort study that were used to create an antibiotic information application are compared to the results of a second card sort study. This second study was conducted with the aim of redesigning the nurse-aimed information application to meet the (overlapping) needs of physicians. During the first card sort study, 10 nurses participated. In the second card sort study, 8 residents participated. The same set of 43 cards were used in both setups. These cards contain fragments of antibiotic protocols and reference documents that nurses and physicians use to be informed about the use and administration of antibiotics. The participants sorted the cards in individual sessions, into as many categories as they liked. The sorts of both user groups were analyzed separately. Dendrograms and similarity matrices were generated using the Optimal Sort online program. Based on the matrices, clusters were identified by two independent researchers. On these resulting clusters of cards, overlap scores were calculated (between nurse and resident clusters). Differences are compared.

The results show that overall, residents reached higher agreement than the nurses. Some overlap between categories is observed in both card sort data matrices. Based on the nurses' data, more and more specific clusters were created (which in part were observed in the larger residents' clusters).

Based on our findings we conclude that a redesign may not be necessary. Especially when the target group with the lowest prior knowledge levels of the information domain is included in the card sort study, the results can be

translated to other groups as well. However, groups with little knowledge will more likely result in lower agreement in the card sorts. Therefore, a larger sample and/or including participants with low and high knowledge of the information domain is advisable.

Keywords: Card sorting · Re-design · Human-centered design · Information architecture

1   Introduction

Card sorting is a method that is used to establish an information structure for a website or application. In a card sort task, participants sort cards that have excerpts of the information that is to be communicated printed onto them. The resulting sorts (groups of cards) of every participant are compared and analyzed to detect what information topics are related to each other in the eyes of the participants. This method gives insight into the participants' mental models and thus provides a way to create information architectures that meet users' needs [1, 2]. Such an approach may be especially important to take for the development of eHealth applications, which often rely on medical information that is highly expert-driven [3]. Besides patients, health care workers can struggle to process information too. This can be caused by a mismatch between the information and their mental models, causing difficulty in comprehending and applying the information in daily life or clinical practice [4, 5]. Card sorts have been applied with success to create eHealth applications, both for patients and medical professionals [4–6]. In these examples, the human-centered design approach, including card sort studies, enabled a good fit with practice and increased efficiency in information use [7]. One of the possible drawbacks of the card sort approach is that usually a limited number of participants is involved, representing a (often) specific target group. After release, the website or application may prove useful to other user groups as well. This either results in users using an information application that was not designed for them (but possibly works fine), or it forces a re-design. Human-centered design implies certain levels of tailoring of information or features to specific user needs is needed [1]. Following this rationale, developers would need to redesign their application when rolling it out to other groups.

In this paper, we address the question whether a redesign is necessary for different user groups, based on a comparison of card sort results of two user groups. More specifically, we use the case of the information structure of the Antibiotic App, which was driven by a human-centered design approach, including a card sort study among nurses (the main target group) [6]. Physicians and residents (not the app's target group) of the pilot wards who used the app requested to have their own version of the app, quickly after it was introduced onto the wards. To accommodate these new users, a redesign process was started. This process included a second card sort study. In this paper, we describe the analysis of the data of both card sort studies to learn whether card sorting (of identical samples of cards) among different target groups results in substantially different sorts. This study can help researchers and developers decide on

who to include in their formative research target groups, and anticipate on the universal applicability of their user-generated information structure.

2 Method

The results of a previous card-sort study with nurses [6] are compared to the results of a later but similar card sort study with a different participant group (medical residents). In the second study, eight residents of the same hospital the nurses in the first study worked at participated. The participants sorted a set of 43 cards with excerpts of medical protocols, reference books, guidelines, and other antibiotic-related information sources printed onto them. In individual settings, every participant created piles or groups of cards that were logical to him/her. The participants were free to create as many groups of cards that they thought was necessary. Prior to participating, participants were instructed about the study's goal and provided their informed consent to participating in the study.

The groups of cards that every participant created were entered in and analyzed with the online Optimal Sort program [8]. In this analysis, a focus was put on the cluster analysis; clusters of cards emerged based on how often two cards were placed together in the same group by the participants. Based on the similarity matrices and dendrograms (tree-like visual representation of clusters, produced by the program), two independent researchers defined the clusters for both participant groups (nurses, residents). The clusters were created by selecting the cards with high similarity (placed together by > 50 % of the participants). Outliers (no participant majority exists for any combination) were placed in the cluster where their similarity scores with the cards are highest. Next, the independent researchers compared the identified clusters to ensure the procedure renders consistent results. The final clusters of both studies (nurses, residents) were analyzed for similarities by calculating an overlap score, based on a similarity calculation applied in card sort analysis; the Jaccard score [9]. In this calculation, the amount of items that occur in both clusters is divided by the total amount of items in both clusters, then multiplied by 100 [9]. An overlap score of 75–100 was considered sufficient to label the two clusters as identical. Scores between 50–75 represent substantial overlap, scores of 25–50 moderate overlap and scores between 0–25 indicate that there is no noticeable overlap. Differences between the clusters are described and interpreted based on the contents of cards that the two participant groups did not agree upon. In the aforementioned analysis, agreement between participants to place two cards in one group was considered, not the similarities between the names participants gave to every pile of cards they created.

3 Results

The similarity matrices and dendrograms of both studies as produced by Optimal Sort can be viewed online via [10, 11].

3.1   Initial Analysis - Dendrograms

Based on the dendrograms, first results indicate that overall, similar groups are created by both participant groups. The dendrograms show that, with a cutoff percentage of 50 %, the nurses' and residents' data both result in 3 overall clusters. Roughly, nurses made a cluster of instructions during/for use (preparation, administration), description or general information, and special cautions or checks. The residents identify similar clusters, but with a separate cluster on dosage instead of cautions or checks, at this very basic agreement level.

3.2   In-depth Analysis – Proximity Matrices

To get more in-depth information on which cards are placed together and thus constitute a cluster, the proximity matrices were analyzed. These analyses resulted in five clusters in the residents' similarity matrix, and seven in the nurses' matrix (see Table 1).

Table 1. Nurses'and residents'cluster overlap scores

Nurse cluster [a] (amount of cards) ↓	Resident clusters [a] (amount of cards) →	Preparation - administration (15)	Particularities in administration (6)	Dose and checks (5)	Particularities – indication (8)	General description (9)
Administration (9)		50	7.1	0	0	0
Preparation - dose (11)		36.8	13.3	14.3	0	0
Particularities (9)		0	15.4	7.7	6.3	38.5
Cautions (6)		0	0	0	40	15.4
Type (2)		0	0	0	0	22.2
Side effects (3)		0	0	0	37.5	0
Checks (3)		0	12.5	33.3	0	0

[a]Clusters were defined based on similarity matrix scores > 50 %, or highest as possible for outliers.

The researcher provided names for each cluster, based on the cards placed in it. Roughly, the groups of cards that are recognized in both card sort outcomes include a cluster of cards containing information on preparation and administering the medication, cautions, particularities, and general descriptions. The nurses' clusters tend to be more specific (and smaller). Also, the average agreement percentages in the matrices (see outcomes via online display [10, 11]) show that the residents' data results in clearer and stronger clusters. The resident *Preparation and administration* cluster overlaps with the nurses' *Administration* (overlap score 50) and *Preparation- dose* (overlap score 36.8) cluster. However, based on the residents' sorts, dose is associated with some of the checks that need to be done (to adjust proper dosing), and is a separate cluster (*Dose and checks*). The nurse cluster *Particularities* holds information cards that can apply to other categories; the greatest overlap with the residents' clusters is with *General description* (overlap score 38.5), but there is also overlap with *Particularities – indication* (overlap score 6.3), *Dose and checks* (overlap score 7.7), and

Particularities in administration (overlap score 15.4). The nurses' more specific clusters of *Cautions* and *Type* are together with *Particularities* related to the somewhat generic or overarching resident cluster *General description* (overlap scores 15.4, 22.2, and 38.5, respectively.

During the card sort sessions and while analyzing the results it became clear that nurses had a tendency to group cards with information they were unfamiliar with into one group labeled 'for the physician'. The residents did the same with cards that hold information that is (in their eyes) very specific to the nurse tasks.

3.3   Disagreement and Cluster Overlap (Per Participant Group)

Some of the cards showed high disagreement among participants; participants did not agree on which cards the card should be grouped with. To get more insight into this, we checked in the similarity matrices [10, 11] which cards had high (> 50) similarity scores with cards outside their cluster. The residents' data shows that the *Preparation and administration* and *Particularities in administration* clusters (residents) are related, as similarity scores of cards of these two separate clusters range between 12–62 (median 37, mean 37.7). Card 38 (additional checks, belongs to dose and checks cluster) was placed with cards from the *Preparation and administration* cluster often (similarity score median 25, mean 29.1). In addition, card 57 (acute responses, belongs to *Particularities – indication* cluster) is placed with cards from the *Particularities in administration* cluster often (similarity score median 37, mean 33).

The nurses' data shows overall more disagreement; the similarity matrices [10, 11] show that many cards are placed with cards outside their own cluster. In fact, six cards in this matrix were grouped together with every other card by at least one participant. These were the cards 54. Compatibility, 71. Pregnancy, 70. Contra-indications, 73. Interactions, 38. Extra checks, and 39. Blood gas tests. In the residents' data matrix, this (a card being grouped with every other card by at least one participant) did not occur. An overlap between the *Administration* and *Preparation – dose* clusters is observed through the 8 cards of the former category which are placed often (similarity score > 50) with 9 cards of the latter cluster. Many cards of the *Particularities* cluster relate to the cards in *Cautions* (54. Compatibility, 55. Incompatibility, 71. Pregnancy, 68. Kinetic information, 36. Descriptions, 67. Characteristics) or *Type* (68 Kinetic information, 66. CHF advice, 40. Renal insufficiency). The *Checks* cards 38 (Extra checks) and 39 (Blood gas test) are related to the *Caution* cluster. These observations include cards which are placed with another card by at least 50 % of the participants. Smaller overlap (10–50 % agreement) outside the cluster was also observed.

4   Discussion

The results show that overall, similar groups are created by both user groups. However, some subgroups are placed together by nurses but not by residents (e.g., cards on preparation and dose) and vice versa (dose and checks). In addition, the nurse sorts showed less agreement and similarities than the residents' sorts. Regarding the research objective of determining whether re-design is needed, based on card sort results, these data provide useful insights. Especially in medical areas where the target groups have a

fairly high level of (formal) training, redesign may not be necessary. Overall, overlap does exist and even when main clusters are not fully equal, subcategories are formed similarly by both groups. If different user groups have mutual understanding and agreement on each other's' (other target groups') tasks and responsibilities, this likely also contributes to universally applicable information structures. A need for redesign may become more urgent when user groups that have less knowledge of the subject wish to start using the application. In other studies, participants who are unfamiliar with the system and its contents created more diverse groups and labels in card sort studies than more knowledgeable and experienced participant groups [12]. Some of the cards were less relevant to the nurses than to the residents (e.g., information on renal insufficiency, kinetic characteristics of the drug), and vice versa (e.g., information on the preparation of an intravenous drug which is done by the nurse). These cards could be seen as a bit 'difficult', as they are not a prominent or relevant part of daily tasks. Interestingly, physicians still grouped all the cards similarly, whereas many difficult or less relevant cards were grouped in very diverse ways by the nurses. An explanation for this could be that residents or physicians are forced to internalize the protocols and reference books to a greater extent than nurses, in order to make save decisions on patient treatment. Nurses, however, mostly perform their regular tasks and only look up the information when something out of the ordinary happens. To accommodate the mental models of the nurses, who, in this case, seem a bit fuzzier than the residents', the information structure should enable cross-referencing to various themes or even include some description of the information structure to facilitate searching.

Possibly, including representatives of the target group with the lowest level of prior knowledge of the information as a participant group in formative card sort studies results in the best chances of universally applicable and comprehensible information structures. However, because participants who have little prior experience with the content need more time and effort to comprehend and process the cards, fewer cards can be used in a study. In addition, it is likely that more participants are needed to be able to identify clusters made by a majority. The reason for this is that there are probably more differences between inexperienced participants' sorts than between the sorts of a sample of participants who are knowledgeable on the subject of the cards that are sorted. In our study, the sorts of both professions were comparable, so designs are based ideally on user studies with both experienced and inexperienced participants. Their precise profession or background may matter less than their experience with the information that is communicated.

In the presentation, we will address the question whether a redesign for different user groups is necessary, based on a comparison of card sort results of our two user groups. Besides the conclusions presented in this paper, we will elaborate in the presentation on the differences between both user groups that we found. In addition, we will discuss how such results have implications for a good comprehension of the tasks to be supported by information applications.

4.1   Limitations

This study suffers from some limitations that need to be taken into account. First, the number of cards used in this study (43) may have been insufficient to enable participants to form various, substantive clusters. In a previous study, samples of 100 cards proved to be about the maximum of cards (if not too much) to sort [12]. Therefore, in this study, fewer cards were used. Likewise, the amount of participants is not high, but not unusual for these types of studies [13]. The ideal amount of cards and participants to include in a study depends on the content (difficulty, variation of cards), and the representativeness of the participant sample. The limited amount of cards allowed us to create the clusters manually (based on the similarity matrices). Bigger samples (of cards and participants) are too complex to analyze manually but they do enable statistical analysis like hierarchical cluster analysis or factor analysis [9].

4.2   Future Work

Possibly, card-sort studies are a low-threshold method to learn whether redesigns are needed. In future studies, the need for additional user studies for a re-design that matches the needs of a new target group can be further investigated. This study shows that in fields with highly protocolled information, inter-professional differences are low. Rather, (job) experience and good understanding of the other target group's tasks and responsibilities might be at play. It is worthwhile to explore to what extent designs can be attuned to information domain novices or experienced users. Our future work will (among other things) further focus on determining the best analytic approach for analyzing card sort data, including in non-web-development settings, as a means of qualitative data structuring [14].

5   Conclusion

Based on our findings, we conclude that card-sorting is a quick method to determine whether a redesign for the information structure is needed for different target groups. However, even when the target groups' tasks differ, results will be more similar with increased participant experience and mutual understanding of the tasks an responsibilities of other user groups. In such cases, redesign is unnecessary. Therefore, if user groups are heterogeneous, and some users may have low levels of information and task experience, it is best to include these users in design studies, regardless of the user target group they represent to ensure the broadest relevance or universal information structure match of the application.

References

1. Maguire, M.: Methods to support human-centred design. Int. J. Hum Comput Stud. **55**, 587–634 (2001)
2. Morville, P., O'reilly, T.: Information Architecture for the World Wide Web. O'Reilly, Sebastopol (2007)
3. Van Velsen, L., Wentzel, J., Van Gemert-Pijnen, J.E.W.C.: Designing eHealth that matters via a multidisciplinary requirements development approach. JMIR Res. Protoc. **2**(1), e21 (2013)
4. Verhoeven, F., Karreman, J., Bosma. A., Hendrix, H.R.M., van Gemert-Pijnen, L.E.W.C.: Toward improved education of the public about methicillin-resistant Staphylococcus aureus: a mental models approach. Inter. J. Infect. Control **6**(1) (2010)
5. Verhoeven, F., Hendrix, R.M., Daniels-Haardt, I., Friedrich, A.W., Steehouder, M.F., van Gemert-Pijnen, J.E.: The development of a web-based information tool for cross-border prevention and control of Methicillin Resistant Staphylococcus aureus. Inter. J. Infect. Control **4**(1), 11 (2008)
6. Wentzel, J., van Velsen, L., van Limburg, M., de Jong, N., Karreman, J., Hendrix, R., van Gemert, J.E.W.C.: Participatory eHealth development to support nurses in antimicrobial stewardship. BMC Med. Inf. Decis. Making **14**(1), 45 (2014)
7. Wentzel, J., van Drie-Pierik, R., Nijdam, L., Geesing, J., Sanderman, R., van Gemert-Pijnen, J.: Antibiotic information application offers nurses quick support. Am. J. Inf. Control (2016). (in press)
8. Optimal Workshop - Optimal sort. https://www.optimalworkshop.com/
9. Capra, M.G.: Factor analysis of card sort data: an alternative to hierarchical cluster analysis. Proc. Hum. Factors Ergon. Soc. Ann. Meet. **49**(5), 691–695 (2005). SAGE Publications
10. Nurse card sort study results: similarity matrix and dendrogram. https://www.optimalworkshop.com/optimalsort/mrsa-net/68158f/shared-results/ba49d9fe334fe8f3ba780902dc37e6a8
11. Resident card sort study results: similarity matrix and dendrogram. https://www.optimalworkshop.com/optimalsort/mrsa-net/b75e78/shared-results/805474e2179ad3089f370cbb6c73eb58
12. Wentzel, J., Müller, F., Beerlage-de Jong, N., van Gemert-Pijnen, J.: Card sorting to evaluate the robustness of the information architecture of a protocol website. Inter. J. Med. Inf. **86**, 71–81 (2016)
13. Nielsen, J.: Card Sorting: How Many Users to Test (2004). http://www.useit.com/alertbox/20040719html. Accessed 2 March 2015
14. Coxon, A.P.M.: Sorting data: Collection and analysis, vol. 127. Sage Publications, Thousand Oaks (1999)