Evaluation of information visualization techniques: analysing user experience with reaction cards

Tanja Merčun
University of Ljubljana
Faculty of Arts, Department of Library and Information Science and Book Studies
tanja.mercun@ff.uni-lj.si

ABSTRACT
The paper originates from the idea that in the field of information visualization, positive user experience is extremely important if we wish to see users adopt and engage with the novel information visualization tools. Suggesting the use of product reaction card method to evaluate user experience, the paper gives an example of FrbrVis prototype to demonstrate how the results of this method could be analysed and used for comparing different designs. The authors also propose five dimensions of user experience (UX) that could be gathered from reaction cards and conclude that the results from reaction cards mirror and add to other performance and preference indicators.

Categories and Subject Descriptors
H.5.2. User Interfaces: Evaluation/methodology

General Terms
Performance, Design, Human Factors

Keywords
Information visualization, user experience, reaction cards, evaluation

1. INTRODUCTION
An analysis of empirical studies identified three approaches to the evaluation of information visualization systems: a) evaluating user performance, b) evaluating user experience (UX), and c) evaluating visualization algorithms [10]. In this paper we discuss the second scenario and propose the reaction card method as a possible technique for evaluating UX in information visualization implementations.

Interaction experience or the quality of user's interaction with a service or system presents a major factor in its success [11] and that is especially true for the field of information visualization. As Faisal et al. [6] argue, users' interaction with information visualization applications is a subjective experience going beyond the interface, which typical usability measures cannot fully capture. User's attitude is also important because it is the appeal and positive affection towards an object that affect learnability, preferences, and even performance, consequentially fostering the adoption of novel approaches that otherwise have a disadvantage compared to the already established ones [9]. Besides presenting a major factor in the success of a given visualization [3], UX can also help to evaluate and inform the design of information visualization applications in the early stages of their development. This paper will discuss the possibilities of using reaction cards to evaluate different visualization techniques, proposing a novel way of analysing and interpreting data collected using this method.

BELIV '14, November 10 2014, Paris , France

ACM 978-1-4503-3209-5/14/11…$15.00
http://dx.doi.org/10.1145/2669557.2669565

2. DIMENSIONS OF USER EXPERIENCE
The field of UX examines the quality of information interactions from the perspective of the user [16]. However, as a recently established area with a growing body of conceptual and empirical research in the last fifteen years, there are still numerous questions regarding its theoretical foundations as well as UX evaluation methods and measurements [12].

A number of studies have been dedicated to better understanding and defining the scope of UX. Analysing different dimensions, Mahlke and Thüring [13], for example, identified two types of qualities related to UX: instrumental qualities connected to the usability and usefulness of a system, and non-instrumental qualities that are closely related to the appeal and attractiveness. Law and Schaik [11], on the other hand, defined four dimensions as the main constructs of UX: user's perceived hedonic quality (pleasure-producing product qualities), pragmatic quality (user-perceived usability), beauty (aesthetics), and goodness (overall product quality). In their review of 58 studies between 2005 and 2009, Law et al. [12] found that as many as 42 unique constructs have been measured in the field of UX, 12 of which were used more frequently: flow, aesthetics/beauty, emotion, enjoyment, affect, arousal/valence, hedonic quality, intrinsic motivation, presence, engagement, attractiveness, and satisfaction. Further, their survey among HCI professionals also revealed a number of UX measures, some of them easier and others more difficult to measure, for example: challenge, engagement, surprise, interest, trust, fun, enjoyment, curiosity, ease of use, intuitiveness, usefulness, stimulation, desirability, immersion, efficiency, comfort, benefit, etc.

While these concepts give a good idea of the area covered by UX, there is yet no one definition of UX dimensions and there is still a question of how to measure these concepts and to what extent they are even measurable. Another dilemma that comes to mind is also whether the same set of concepts and dimensions would apply to the field of information visualization or would there be some aspects more and other less fundamental.

3. REACTION CARD METHOD

Developed by Microsoft experts and first reported in 2002 (see [2]), the product reaction card method was driven by the limitations of standard feedback mechanisms such as Likert scales. Described also as a desirability toolkit, reaction cards provide "a way for users to tell the story of their experience, choosing the words that have meaning to them as triggers to express their feelings – negative or positive – about their experience" [1]. This makes the reaction card method an interesting tool for tapping into UX which works best when used along other instruments (to triangulate their findings) or when it is used as a baseline for comparison [1].

Employed at the end of a testing session, participants are typically asked to select from a set of cards (adjectives) those that best reflect their experience with the system. After making a first selection, they are sometimes requested to comment on their top five choices, thus eliciting commentary and providing a better insight into user's experience. In its original design, there was a list of 118 cards to be used in the method; however, several studies (for example [17]) have reported using only a limited selection of those adjectives.

Results of reaction cards can be analysed and used in a variety of ways, but they are still most commonly presented in form of a word cloud [4, 7, 8]. The most comprehensive but basic description of possible analysis has so far been provided by Barnum and Palmer [10], who reported on using the following methods in their experiments with reaction cards: a) the number of positive cards b) the percentage of positive vs. negative chosen words c) mapping the frequency of a few chosen cards on a radar graph d) presenting the frequency of selected cards in a graph or a word cloud, and e) creating meaningful clusters.

4. COMPARING VISUALIZATIONS WITH REACTION CARDS: A CASE STUDY

Apart from our own research, we have been able to find only one similar study [7] that has employed the reaction card method to assess and compare different information visualization techniques. Using all 118 adjectives, they asked the participants to choose 5 cards that best described a specific visualization. They learned that while users found both techniques easy to use, they saw the indented tree technique as familiar, organized, straightforward, and simplistic but also as boring, dull, and busy, while the graph was described as intuitive but also as annoying and complex.

Even though our study was primarily focused on evaluating usability and applied the reaction card method using only a small set of adjectives intended for soliciting commentary, the results suggested that the method holds interesting potential also for information visualization. Using results of our study, the following part investigates possible ways in which reaction cards can tell a story of UX.

4.1 Study design

Evaluating 4 hierarchical layout implementations in FrbrVis prototype (Figure 1) [14], our usability study also included reaction cards in order to examine user's perceptions and experience with individual visualizations. Next to the four visual designs, a Baseline prototype was also included in the study to compare our approach to a more traditional bibliographic information system approach currently used in libraries. Each of the 120 study participants worked with 3 randomly counterbalanced prototype designs, which meant that each prototype design was tested 72-times.

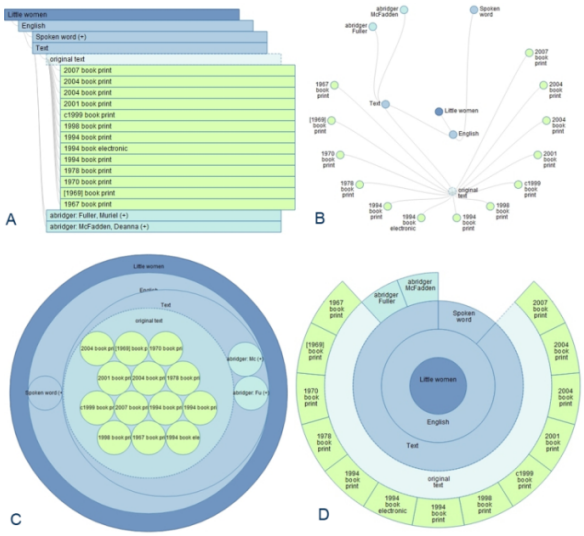


Figure 1. 4 hierarchical prototype designs: A. Indented list - B. Radial tree - C. Circlepack – D. Sunburst

The reaction card method was modified from its original, using only 29 adjectives (14 positive, 15 negative – see Table 1) which were considered appropriate for describing interactions with visual designs. After completing 10 tasks in a given prototype design, participants were asked to choose any number of reaction cards that were listed on a sheet of paper in a random order. These were then used also at the end of each session when participants were requested to elaborate more on the three designs they tested with the help of selected adjectives.

Table 1. A set of reaction cards used in the study.

time consuming	unfriendly	useless
frustrating	clumsy	hard to use
quick to understand	inefficient	difficult to understand
informative	unappealing	efficient
innovative	organised	useful
complex	deficient	appealing
easy to use	transparent	fun
advanced	convenient	interesting
illogical	attractive	intimidating
opaque	logical	

4.2 Analysing reaction cards

4.2.1 The selection of individual cards

The analysis of individual adjectives and how often they have been selected is by far the most common application of reaction cards. Looking at examples in literature we can see that the results of such examinations have typically been reported in form of

word clouds and presented the only quantitative outcome. While word clouds provide a quick overview of the most frequently selected adjectives at first glance, they do not, however, communicate results in a very clear manner that would allow a more detailed comparison. Figure 2, for example, displays word clouds for our five prototype designs. The colours reveal that Baseline and Radial prototype received more negative responses than other designs, and the large typed words tell us that Radial design was seen as interesting, innovative, and logical, Circlepack and Sunburst were appraised as quick to understand, while Baseline and Indented tree were assessed as more informative and quick to understand.

Figure 2. Word clouds for individual prototype designs.

Classical graphs, on the other hand, enable us to also observe more precise distances between selected adjectives within an individual visualization or in comparison to another design. Figure 3 shows the selection of reaction cards for visual prototype designs (average for the four visualizations) in comparison to the Baseline design in our study. The figures tell us that participants in the study experienced Baseline prototype as similarly informative, useful and efficient as visual designs, but significantly less quick to understand, logical, easy to use, interesting, innovative, fun and attractive. It is also possible to see that more than 20% of participants found the Baseline prototype as time consuming, clumsy and complex, which was far more than in case of visual designs.

A similar comparison also reveals more details about the differences among the four visualizations. To name just a few:

- Indented list was least often selected as attractive, interesting, fun, innovative, or advanced, but most often as useful, efficient, organised, transparent, and informative.
- Radial tree design received no highest values for positive cards and was, compared to the other three visualizations, described as least appealing, useful, efficient, convenient, transparent, organised, easy to use, and quick to understand as well as most time consuming, complex, opaque, clumsy, difficult to understand, and hard to use.
- While Circlepack was noted as most fun of the four as well as most illogical and unappealing, its values for most of the cards were generally average, not standing out in one or the other direction.
- Sunburst was seen as most quick to understand, innovative, advanced, convenient, logical, interesting, and attractive. Interestingly, it was the only visualization that received no lowest scores in positive cards as well as no highest score in negative cards.

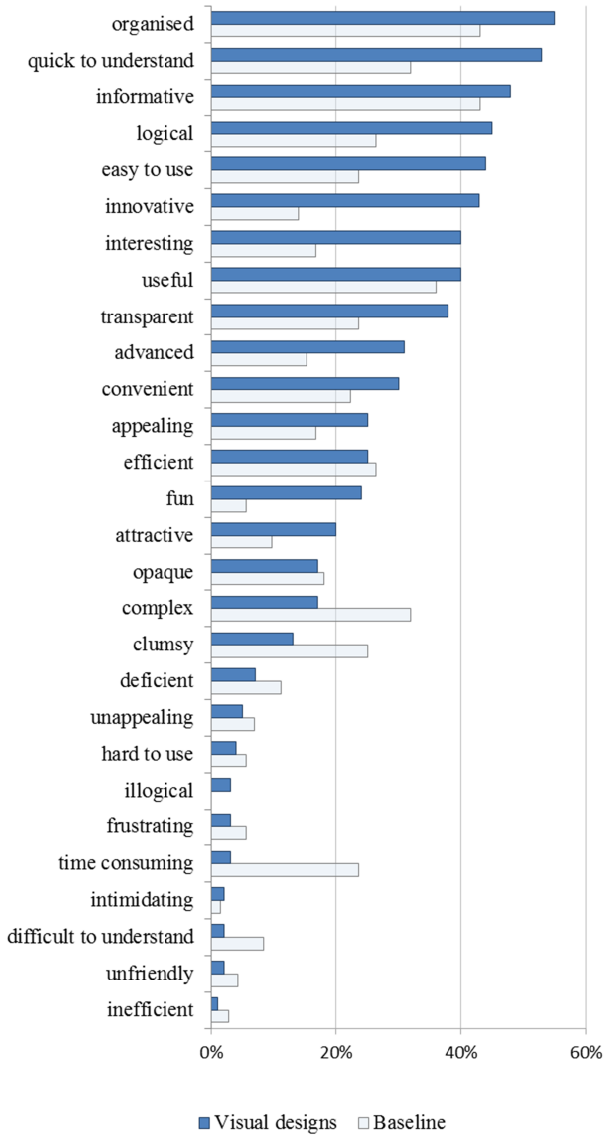


Figure 3. Reaction cards selected for visual designs in comparison to Baseline.

Although examining the selection of individual cards is the most common approach, an exploration of results from our study showed that other types of analysis can also be very informative.

4.2.2 The overall number of selected cards

The sheer number of selected cards can be an indicator of positive attitude toward a certain design. Comparing the results in Table 2 to other performance and preference results from the study, it is clear they correspond to a large degree: Baseline prototype was ranked lowest and was closely followed by the Radial design, while the differences between Circlepack, Indented list, and Sunburst were small. Sunburst, the design chosen as the preferred by most participants, also in this measure received the highest number of positive cards.

Table 2. The total number of selected cards per design.

BASELINE	*INDENTED*	*RADIAL*	*CIRCLEPACK*	*SUNBURST*
359	456	423	474	491

4.2.3 Ratio of positive and negative cards

The ratio of positive and negative cards (Table 3) even better mirrors the overall results of usability testing as Indented list and Sunburst always ranked highest with only minimal differences between them, while Circlepack always ranked 3rd and Radial design and Baseline proved to be significantly inferior. The fact that Sunburst and Indented list received over 90% of positive cards therefore tells us that both designs evoked also positive UX.

Table 3. Percentage of selected positive and negative cards.

	BASELINE	*INDENTED*	*RADIAL*	*CIRCLEPACK*	*SUNBURST*
+	71%	94%	76%	88%	92%
-	29%	6%	24%	12%	8%

4.2.4 Dimensions

Looking at the individual cards results, we felt that different designs elicited different kinds of UX, meaning that while some received a high number of positive reactions due to their perceived usefulness and organisation, others drew more positive responses related to interest and appeal.

Analysing the original set of 118 reaction cards, we saw the potential in grouping them by 5 different dimensions (questions regarding the choice of dimensions are addressed in the Discussion section):

- perceived ease of use (clear, effortless, friendly, intuitive,…),
- perceived usefulness (helpful, relevant, valuable, meaningful…)
- perceived efficiency (effective, responsive, time-saving, fast,…)
- appeal (appealing, attractive, desirable, impressive, novel,…)
- engagement (engaging, exciting, entertaining, inviting, motivating, inspiring,…)

These dimensions could also be applied to our limited set of reaction cards, the results of which we then illustrated using a radar chart. For graphical presentation, positive and negative reaction cards were separated and while both aspects could be presented in a graph, we feel that creating dimensions based on the ranges of positive cards may be sufficient as the charts with negative cards basically showed only their mirror image.

Figure 4, for example, maps the values collected in our study to the five constructed dimensions. The presentation enables us to compare all prototype designs at once and it can be observed that, for example, Indented list scored relatively low on engagement and appeal, but high on usefulness, efficiency, and also ease of use. In terms of values achieved on individual dimensions, the graph also gives us a good idea of which visualizations have, according to reaction cards, provided an overall lower UX (Baseline and Radial design).

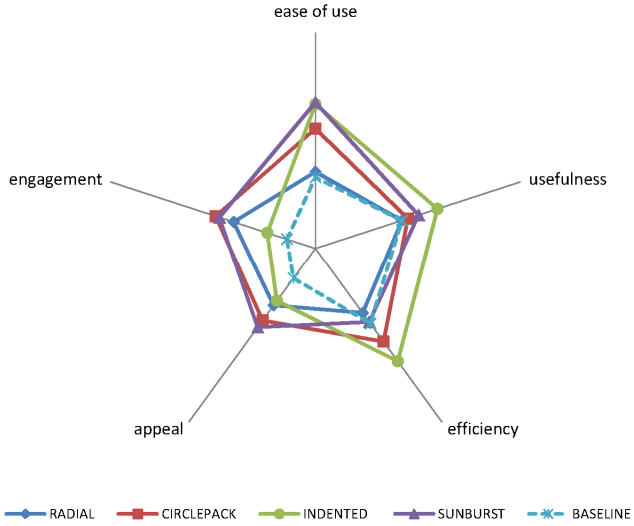
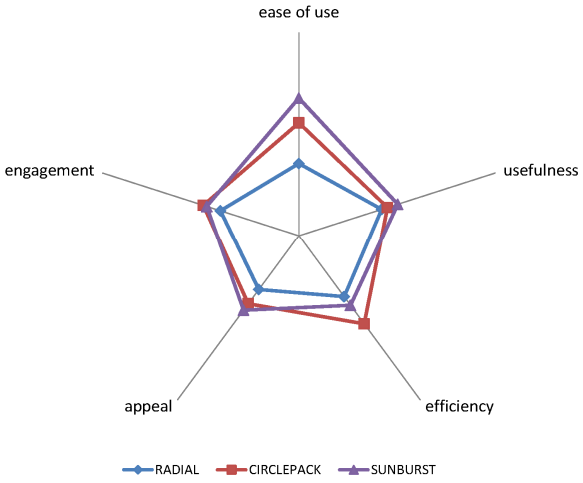


Figure 4. Radar graph comparing prototype designs on 5 dimensions.

Figure 5 reveals another interesting aspect related to this type of results visualization: the shape of the graph. The two graphs show that while Radial tree, Circlepack, and Sunburst created a balanced (although differently large) shape, the two more traditional designs (Indented list and Baseline) formed a skewed shape which demonstrates that the 5 dimensions of UX are unevenly supported. This makes us conclude that while Indented list and Sunburst are quite equal in all other measures, Sunburst seems to provide a more all-around UX.

Another possible use of dimensions would also be to examine how different groups of users experience the same visualization. Figures 6 and 7 illustrate this by comparing the scores of science & engineering students to those by humanities & social science students for two visualizations. Interestingly, the Indented list was seen more positively by humanities & social science students, but the overall shape of the graph indicates comparable perception by both user groups. In case of Radial tree, on the other hand, it is possible to observe some differences between the groups both in shape of the graph as well as in some of the values. Similar analysis could also be carried out by gender, by the level of experience or by some other criteria.



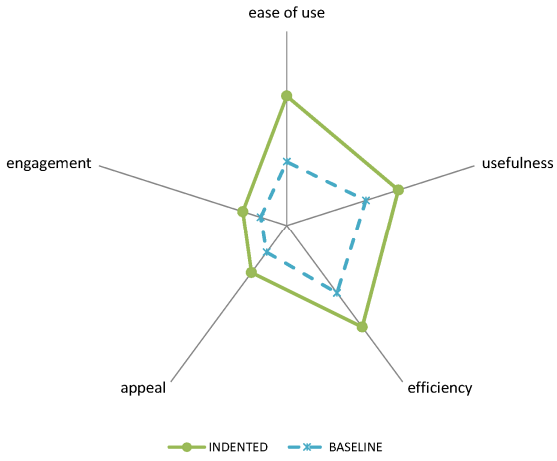


Figure 5. Two different shapes of radar graphs.

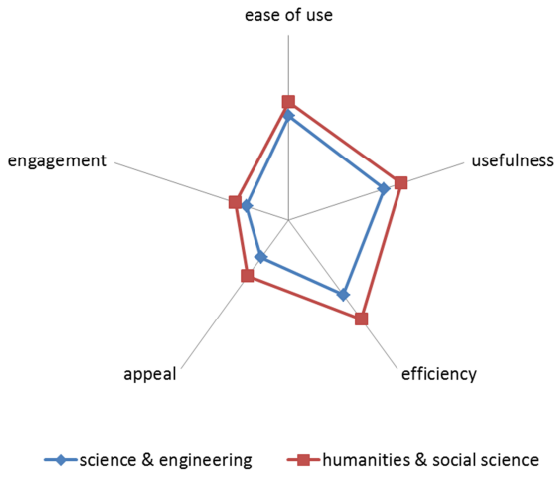


Figure 6. Perceptions of Indented list design by two different user groups.

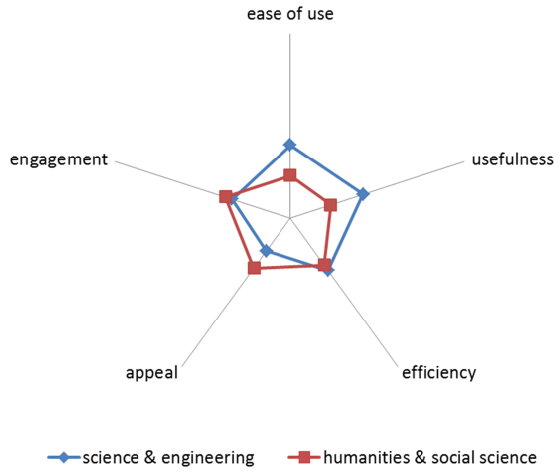


Figure 7. Perceptions of Radial tree by two different user groups.

4.2.5 Surface area

Presenting dimensions in a radar graph also enables a calculation of its surface area, which gives a mathematical expression to the overall UX. Typically used in benchmarking analysis, the method is known as SMOP – "surface measure of overall performance" and is calculated using the following formula [15]:

((P1*P2)+(P2*P3)+(P3*P4) +....+(Pn*P1)) * sin(360/n)/2

where P is the data point on the axis of the radar chart and n is the number of axes.

A comparison of different information visualization designs by surface area in our case (Table 4) clearly showed that Sunburst, by this measure, provided the best UX.

Table 4. Surface area of radar graphs.

BASELINE	*INDENTED*	*RADIAL*	*CIRCLEPACK*	*SUNBURST*
0,142	0,369	0,214	0,362	0,390

5. DISCUSSION

While all the examined analysis approaches give some feedback on user's experience, we believe that mapping by dimensions gives a new and added value to reaction cards that has so far not been considered. The creation of new variables even enables a statistical comparison of visualization designs by individual dimensions.

However, much more extensive research will be needed in order to verify and establish not only the correct dimensions but also the sets of adjectives that constitute each dimension as our exploratory set of dimensions was done without much prior in-depth analysis. We have recently done some exploration into using factor analysis for recognizing the dimensions and saw promising results on the small set of adjectives used in our study. Nevertheless, our previous studies were formed with a different purpose in mind, which means that the results cannot be generalized. A new study will be needed, preferably with a larger set of adjectives, in order to clearly identify the dimensions of UX

as expressed through reaction cards. Such a study could also serve to identify principal components of each dimension, which would help reduce the number of reaction cards to a more manageable set. We believe that a smaller number of reaction cards should be used, not only because of the time and effort required from the users but also because our observations show that users get confused by the choice of very similar adjectives.

Another area of research should also focus on individual reaction cards, examining how well they cover the concepts that have been identified in literature as essential for UX. This would be especially interesting for the field of information visualization where we would also need to establish whether there are some aspects that are specific or more important compared to visualizations. Since UX has not really been the focus of research in the field of information visualization [5], there is still much work to be done in defining the UX aspects within information visualization.

Our current investigations suggest that not all adjectives included in the original set of 118 reaction cards would be particularly useful for the field of information visualizations, but many of the adjectives encompassed in the full set are easily applied also to the experiences with visualizations. Regarding the choice of adjectives and the set of dimensions, the results from our tests indicate that engagement and appeal play an important role in the acceptance of visualizations. As working with information visualization is still a relatively novel experience, we observed that these two aspects presented an important factor in the way users reacted to the particular visualization. Another concept that should be more closely examined in context of UX for information visualization is also notion of fluidity [5], which is not really well represented in the original set of reaction cards. The idea encompassed in fluidity is the seamless and natural interaction where the user is able to stay in the flow of visualization, immersed in the process and with a sense of control. What would be the proper adjectives to express these ideas will still need to be tested in future studies, as well as the weight each of the dimensions has on the overall UX.

6. CONCLUSION

Showcasing the possibilities and potential of product reaction cards for analysing UX, this paper aimed to introduce the method as an interesting option also for the field of information visualization. While reaction cards can also be used just as a tool for eliciting user's comments that help gain a deeper understanding of how they feel about the visualization and why, more quantitative analysis is also possible that might, on one hand, give some more tangible results concerning UX and, on the other hand, lead us towards better comprehension of UX in the context of information visualization.

Going beyond the commonly used word cloud, we feel that our proposed mapping of reaction cards to dimensions could present a useful tool for studying the properties of individual visualization implementations as well as for comparing different designs, giving more solid ground to the field of UX. Furthermore, the overall experience is the final indicator of user's interaction with the visualization and this method also shows potential for providing us with such an overview.

These methods of analysis could also be used in other studies that work with product reaction cards, however, being only our first experiment, the validity of identified dimensions and their possible use in a variety of contexts (for example analysing UX within a specific design by age, experience, etc) still needs to be tested in the future. Interestingly, we can report that the results gathered from reaction cards reflect well all other performance and preference data collected in our usability study, which gives additional support to future use and development of the reaction card method.

7. ACKNOWLEDGMENTS

The FrbrVis prototype was developed in cooperation with dr. Trond Aalberg from NTNU, Norway. The research was partially funded by grant J5-4155 of Slovenian Research Agency. A word of appreciation also to dr. Maja Žumer for all the discussions that resulted in this paper.

8. REFERENCES

[1] Barnum, C.M., and Palmer, L.A. 2010. More than a feeling: understanding the desirability factor in user experience. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems*, 4703-4716. ACM. DOI= http://dx.doi.org/10.1145/1753846.1754217

[2] Benedek, J., and Miner, T. 2002. Measuring desirability: New methods for evaluating desirability in a usability lab setting. In *Proceedings of UPA 2002 Conference*.

[3] Cawthon, N., and Moere, A.V. 2006. A conceptual model for evaluating aesthetic effect within the user experience of information visualization. In *IV '06: Proc.of the conference on Information Visualization*, 374-382. IEEE. DOI= http://dx.doi.org/10.1109/IV.2006.4

[4] De Guzman, E.S., and Schiller, J. 2011. How does this look? Desirability methods for evaluating visual design. In *HCI International 2011 – Posters' Extended Abstracts*, 123-127. Comm. in Computer and Information Science, 173. Springer. DOI= http://dx.doi.org/10.1007/978-3-642-22098-2_25

[5] Elmkvist, N., Moere, A. V., Jetter, H.-C., Cernea, D., Reiterer, H. and Jankun-Kelly, T. 2011. Fluid interaction for information visualization. Information Visualization, 10, 4 327-340. DOI=http://dx.doi.org/10.1177/1473871611413180

[6] Faisal, S., Craft. B., Cairns, P., and Blandford, A. 2008. Internalization, qualitative methods, and evaluation. In *BELIV '08 Proc. of the 2008 Workshop on BEyond time and errors: novel evaLuation methods for Information Visualization*. ACM. DOI= http://dx.doi.org/10.1145/1377966.1377973

[7] Fu, B., Noy, N.F., and Storey, M. 2013. Indented tree of graph? A usability study of ontology visualization techniques in the context of class mapping evaluation. In *The Semantic Web – ISWC 2013*, 117-134. DOI= http://dx.doi.org/10.1007/978-3-642-41335-3_8

[8] Govaerts, S., Verbert, K., Klerkx, J., and Duval, E. 2010. Visualizing activities for self-reflection and awareness. In *Advances in Web-Based Learning, ICWL 2010*, 91-100. DOI= http://dx.doi.org/10.1007/978-3-642-17407-0_10

[9] Julien, C., Leide, J., and Bouthillier, F. 2008. Controlled user evaluations of information visualization interfaces for text retrieval: literature review and meta-analysis. *J. Assoc. Inf. Sci. Technol.* 59, 6, 1012-1024. DOI= http://dx.doi.org/10.1002/asi.20786

[10] Lam, H., Bertini, E., Isenberg, P., Plaisant, C., and Carpendale, S. 2012. Empirical studies in information visualization: seven scenarios. *IEEE Trans. Visual Comput. Graphics* 18, 9, 1520-1536. DOI= http://dx.doi.org/10.1109/TVCG.2011.279

[11] Law, E.L., and van Schaik, P. 2010. Modelling user experience – An agenda for research and practice. *Interact. Comput.* 22, 5, 313–322. DOI= http://dx.doi.org/10.1016/j.intcom.2010.04.006

[12] Law, E.L., van Shaik, P., and Roto, V. 2013. Attitudes towards user experience (UX) measurement. *Int. J. Human-Computer Studies*. DOI= http://dx.doi.org/10.1016/j.ijhcs.2013.09.006

[13] Mahlke, S., and Thüring, M., 2007. Studying antecedents of emotional experiences in interactive contexts. *In Proceedings of the CHI Conference*, 915–918. ACM. DOI= http://dx.doi.org/10.1145/1240624.1240762

[14] Merčun, T., Žumer, M., and Aalberg, T. 2012. FrbrVis: an information visualization approach to presenting FRBR work families. In *TPDL'12 Proceedings of the Second international conference on Theory and Practice of Digital Libraries*, 504-507. LNCS, 7489. DOI= http://dx.doi.org/10.1007/978-3-642-33290-6_60

[15] Mosley, H., and Mayer, A. 1999. Benchmarking national labour market performance: a radar chart approach. Report prepared for European Commission, Discussion Paper FS I 99 –

[16] O'Brien, H.L., and Toms, E.G. 2013. Examining the generalizability of User Engagement Scale (UES) in exploratory search. *Inf. Process. Manage.* 49, 1092-1107. DOI= http://dx.doi.org/10.1016/j.ipm.2012.08.005

[17] Schønau-Fog, H., and Bjørner, T. 2012. "Sure, I would like to continue": a method for mapping the experience of engagement in video games. *Bulletin of Science, Technology & Society* 32, 5, 405–412. DOI= http://dx.doi.org/10.1177/0270467612469068