Development and Evaluation of a Patient Information Website for Childhood Cancer Survivors

Sebastiaan L. KNIJNENBURG a,1, Leontien C. KREMER b,1, A. Birgitta VERSLUYS c, Jaldert R.J. VAN DER BEEK a, Monique W.M. JASPERS a

a Department of Medical Informatics, Academic Medical Center, Amsterdam, The Netherlands
b Department of Pediatric Oncology, Emma’s Children Hospital/Academic Medical Center, Amsterdam, The Netherlands
c Department of Pediatric Hematology and Oncology, Wilhelmina’s Children Hospital, UMC Utrecht, The Netherlands

Abstract. Usability evaluation is an essential step in health care system development, including patient information websites. In this study we compared the think aloud method (TA) and the heuristic evaluation (HE) in a case study on the development of a website for childhood cancer survivors. Both methods managed to uncover all major problems with the website, though additionally HE found cosmetic issues and the TA found problems with website content. These findings contradict earlier studies but may be explained by the inclusion of two double experts in our study, presumably enabling them to take on the role of ‘a patient’ in testing the website. We nevertheless recommend the use of both methods if adequate funding and expertise are available; otherwise when HE usability evaluators are not familiar with the system domain, a work-domain expert could assist them in tackling domain-specific problems.

Keywords. childhood cancer, usability evaluation, heuristic evaluation, think aloud analysis, website

1. Introduction

Due to a major increase in survival rates over the past years, childhood cancer survivors are a new and quickly growing population, with currently approximately 7,000 survivors in Netherlands [1]. As 75% of all five-year survivors suffer from at least one adverse event, adequate life-long follow-up is necessary [2]. In order to improve the quality of care and to successfully conduct clinical research, the DCOG (Dutch Childhood Oncology Group) LATER (Long term effects after childhood cancer) project pursues to follow-up all Dutch five-year survivors and screen them based on evidence-based guidelines [3].

Survivors of childhood cancer are often under-informed about their past disease and potential late effects, which presumably decreases their compliance to screening and
therapy and delaying diagnosis and treatment of late toxicities of their prior cancer
treatment [4, 5]. Patient information websites enable patients to educate themselves
using web-based, personalized information at the time and place of their preference [6].
In healthcare, evaluation of interactive/information systems is critical to their success,
as healthcare systems need to adequately meet requirements and needs of users to
ensure effective use and high quality care [7]. Websites are no exception to this
premise. Several studies suggest a multifaceted, user-centered, iterative approach for
the development and evaluation of patient websites, including focus groups, ‘think
aloud’-analyses and heuristic evaluations [8, 9]. The think aloud analysis is known to
gain the best insight in the cognitive processes of the end users, but at a rather high cost,
whereas heuristic evaluation can give rather good results at much lower costs. It is
however known heuristic evaluation finds a broad range of cosmetic and minor
problems which may disguise more severe problems. In this paper we report a
comparison of these two methods in a case study on a website for childhood cancer
survivors, to find out to which extent both methods differed in their results and to
investigate which method is most effective in relation to associated costs.

2. Background Information

As an earlier study pointed out, only 4/145 survivors are able to find the late effects
information they are looking for on the internet [10]. Because knowledge about
previous disease, treatment and current late effects is low [5], survivors are struggling
online to satisfy their information needs. To tackle this problem, we thought up a
search structure where survivors can review information concerning (potential) late
effects by use of a linked network of diagnoses, cancer treatments and (potential) late
effects information envisioned through three menus with clickable data label entry
fields. All content was generated by members of the Information Service of the DCOG
LATER project.

3. Methods

To evaluate the prototype website regarding usability issues we used two well-known
methods from the field of usability engineering: the think aloud analysis and the
heuristic evaluation [11]. For the think aloud we created three scenarios covering all
the functionality the website offered (using the late effects search structure, finding
information about diagnosis/therapy, looking up outpatient clinic information, etc).
Eight potential end-users were contacted through email to participate in the user tests
[12]. We used a mobile ‘usability lab’ consisting of a laptop with Morae™ software to
capture screen, mouse gestures, mouse clicks and the participant’s face and voice.
Participants were first given a practice task to get accustomed with talking aloud before
starting with the real scenarios. All recorded data were analyzed with the Morae™
software. Firstly a coding scheme was developed bottom-up by one author by
reviewing all occurrences of usability problems. Consequently all problems were
grouped into categories. A second evaluator reviewed all recordings separately and
used the coding scheme to categorize them. Results of both reviewers were compared
and discrepancies were resolved by discussion.
For the heuristic evaluation three evaluators, of which two double experts (on usability and childhood cancer survivors), used the set of Nielsen’s heuristics to systematically inspect the site. Each violation of one of the heuristics was noted together with its severity rating (1–4). To see what kind of differences there were between the heuristic evaluation and the think aloud analysis, we mapped the problems found with both methods to each other and analyzed the results qualitatively.

4. Results

For the heuristic evaluation, the three evaluators listed a total of 40 occurrences of usability problems according to the 10 heuristics described by Nielsen. The usability problems with a severity rating higher than two and those problems found by more than one evaluator are listed in Table 1. The results show that usability problems related to use of external hyperlinks and the search structure were on average rated with a high severity. Other usability issues were related to website content, design consistency over different pages and deviations from standard web design guidelines. The think aloud population consisted of five childhood cancer survivors and three parents of survivors. Average age of the testers was 35 years and duration of follow-up 16 years. All participants had broadband connections available at home and used the internet at least for two hours a week. Testing sessions lasted for around 30 minutes on average. In total the eight testers encountered fifty usability issues which could be categorized into eleven categories (Table 1). The majority of the issues were encountered in the search structure we developed: the system did not work as easily as we expected it to be.

<table>
<thead>
<tr>
<th>Think aloud usability problem category descriptions</th>
<th>N</th>
<th>Associated findings of the heuristic evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘Information’ header wrongly interpreted</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Use of unknown abbreviations</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Users cannot find/do not notice submenu</td>
<td>4</td>
<td>Submenu is not clearly visible</td>
</tr>
<tr>
<td>User cannot return to previous page</td>
<td>5</td>
<td>No ‘Back’ at search structure explanation</td>
</tr>
<tr>
<td>User cannot proceed to full information</td>
<td>5</td>
<td>Automatic change of results not clear when</td>
</tr>
<tr>
<td>User wants to display results by clicking</td>
<td>12</td>
<td>using search structure</td>
</tr>
<tr>
<td>Selection is gone after browser ‘back’</td>
<td>10</td>
<td>Search structure ‘Back’ dysfunctions</td>
</tr>
<tr>
<td>Users see a selection displayed, which is not the</td>
<td>8</td>
<td>Search structure selection does not</td>
</tr>
<tr>
<td>actual active selection</td>
<td></td>
<td>correspond with results</td>
</tr>
<tr>
<td>Users cannot choose between hyperlinks</td>
<td>3</td>
<td>2 hyperlink lines linking to the same site</td>
</tr>
<tr>
<td>Users do not perceive external links as such</td>
<td>4</td>
<td>External links are not depicted as such</td>
</tr>
<tr>
<td>(and get lost)</td>
<td></td>
<td>Deeplinks do not link to correct information</td>
</tr>
<tr>
<td>Search bar does not give expected results</td>
<td>9</td>
<td>Homepage ‘blocks’ not clearly clickable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inconsistent use of hyperlink colours</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inconsistent placement of search box</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Logo’s on homepage not clickable</td>
</tr>
</tbody>
</table>

N: amount of occurrences. S: severity score (1–4).

Users did not notice automatic changes in selections, did not fully comprehend the use of hyperlinks to internal and external web pages and had troubles using the overall menu structure.
In Table 1 the findings of the TA and the HE are lined up against each other, where similar problems are placed on the same rows. All the major usability issues discovered in the heuristic evaluation are covered by the think aloud and vice versa. The four additional issues revealed by heuristic evaluation concerned cosmetic design flaws, as where the two additional problems revealed by think aloud were related to website content. However, the two methods differ in nature in the description of revealed problems. The results of the think aloud method mainly reveal practical problem descriptions of issues that real end users encounter and their underlying causes (i.e., ‘user cannot click on ‘back’ button), whereas the results of the heuristic evaluation are more solution-based, focusing on things the patient website should do, but does not.

The time spent on both analyses differed a lot. The heuristic evaluation took each evaluator approximately one hour, where after it took another hour to aggregate the results. The think aloud analysis costs around 30 minutes to perform, but took a lot more time to analyze: two evaluators both went through all the participants audiovisual recordings to transcribe the usability problems the participants uncovered. This took at least another 30 minutes per participant per evaluator, even though we used a sophisticated software suite specialized at this kind of tasks. We estimate the total time spent on the think aloud analysis at least 30 hours (excluding travel time to/from the participants).

5. Discussion

Usability evaluation of healthcare systems is an essential step in human-centered design and may improve implementation, system adoption amongst end-users and ultimately user satisfaction. In this study we compared two well-known usability evaluation methods, the think aloud method and the heuristic evaluation, on their usefulness and effectiveness in the user-centered development process of a patient information website for childhood cancer survivors. Both the heuristic evaluation and think aloud analysis found all major problems. The workload of the think aloud analysis was much higher than that of the heuristic evaluation.

When comparing usability evaluation methodologies, the interpretation of the differences needs to be done carefully. For example, one can ask whether the problems that were found by HE and not by TA were a ‘false alarm’ of the usability experts or an omission in the system use of the TA-participants? In our current case study we think both the answers can explain the difference and that the main difference between the two methods lies in the eye of the beholder: the HE resulted in uncovering the major problems and more smaller, cosmetic issues, resulting from the usability expertise from the evaluators, as where the think aloud analysis resulted in more major problems with the practical system use. Beuscart-Zéphir et al. found similar results in a study combining heuristic evaluation and a think aloud analysis, where the think aloud sessions provided hints about the causes and severity of the problems detected by the heuristic evaluation [13]. However, in studies using a think aloud analysis after a heuristic evaluation, results showed the input from end-users was irreplaceable to uncover remaining problems in the design of the websites [8, 14, 15]. These studies contradict our results as they showed that the expert heuristic evaluators focused more on problems related to information design, whereas the patients had more concerns related to system navigation, access and computer operating functions, clarifying problems with user interaction and revealing mismatches between designers’ ideas and
users’ actual strategies in approaching the computer-supported tasks: a result not found in our study. The difference in the outcome of these studies and our study might be explained by the inclusion of two double experts in our study, presumably enabling them to take on the role of ‘a patient’ in testing our prototype website.

Practically, we would recommend the use of the combination of both methods for healthcare usability evaluations if adequate funding and expertise are available. As the heuristic evaluation is less costly and is also able to find all major usability issues, it might be a good alternative for testing interactive healthcare websites in circumstances where time and resources are limited, since skilled experts can yield high quality results in a limited amount of time. In situations that usability evaluators are not familiar with the system domain, a work-domain expert could assist the evaluator in tackling domain-specific problems. The work-domain experts in these so-called ‘participatory heuristic evaluations’ may help usability evaluators in considering how the system contributes in reaching certain goals in that particular area of skills. Generally speaking, health care applications should be designed to match end users’ characteristical way of thinking, reasoning and speaking and most of the recurring problems in their design can be prevented by a human-centered design approach.

References