Using Eye Tracking to Measure User Engagement with a Decision Aid

Chapter · July 2023

DOI: 10.1007/978-3-031-35017-7_5

Doaa Alrefaei
Worcester Polytechnic Institute

Lidan Zhang
Worcester Polytechnic Institute

Gaayathri Sankar
Worcester Polytechnic Institute

Soussan Djamasbi
Worcester Polytechnic Institute

Using Eye Tracking to Measure User Engagement with a Decision Aid

Doaa Alrefaei[1(✉)], Lidan Zhang[1], Gaayathri Sankar[1], Soussan Djamasbi[1], Bengisu Tulu[1], Carlie Flanagan[1], Adam Kalayjian[1], Connie Ge[2], Camarlin Franco[2], Shazeb Meraj[2], and Susanne Muehlschlegel[3]

[1] Worcester Polytechnic Institute, Worcester, MA 01609, USA
{Dalrefaei,lzhang11,gsankar,djamasbi,bengisu,ceflanagan, ankalayjian}@wpi.edu

[2] Department of Neurology, University of Massachusetts Chan Medical School, Worcester, MA 01655, USA
{Geconnie.ge,camarlin.franco,shazeb.meraj}@umassmed.edu

[3] Departments of Neurology, Anesthesiology/Critical Care and Surgery, University of Massachusetts Chan Medical School, Worcester, MA 01655, USA
Susanne.Muehlschlegel@umassmemorial.org

Abstract. Eye tracking has become the gold standard in measuring human attention and information-processing behavior. As such, eye tracking in mixed-methods user experience (UX) research serves as an invaluable tool to learn about user needs and to create actionable insights for improving product and service design during the development cycle. Here, we discuss the iterative process that we used to improve the design of a decision aid (DA) developed to facilitate shared decision making. We explain the use of eye tracking during this process to examine how users processed the information provided by the DA. We also explain how we used eye tracking in a retrospective "think-aloud" protocol to gain insight about users' needs. Our results show that user reactions captured by eye tracking can not only be used to optimize design decisions but also to gather user feedback about their information processing needs.

Keywords: User-centered design · iterative formative studies · retrospective "think-aloud" · area of investigation (AOI) maps · user experience design and evaluation · shared decision-making tool

1 Introduction

Decision aids (DAs) refer to digital or paper-based tools designed to facilitate shared decision-making between patients or their surrogate decision-makers and clinicians (Barry & Edgman-Levitan, 2012). DAs provide information about a specific medical condition, possible treatment paths, as well as pros and cons of different treatment options clinicians (Barry & Edgman-Levitan, 2012). DAs typically support non-recurring context-specific health decisions to help select treatment options that are best aligned with patients' values and preferences. Therefore, the engagement design goal for DAs should focus on helping users to thoroughly process the provided information.

D. D. Schmorrow and C. M. Fidopiastis (Eds.): HCII 2023, LNAI 14019, pp. 57–70, 2023.
https://doi.org/10.1007/978-3-031-35017-7_5

Designing successful user-centric systems requires a series of iterative formative user experience (UX) studies, each with the objective to discover insights for improving the design in the next development cycle. Because the most dominant sense for sighted people is vision, eye tracking offers an effective methodology for measuring how people process visual information provided in a DA. Eye-tracking devices provide a continuous and unobtrusive stream of moment-to-moment objective gaze data about a person's focus of attention on various parts of a visual display without placing an additional burden on users. Consequently, eye tracking has become the gold standard for investigations that rely heavily on measuring visual attention (Djamasbi, 2014; Gaffiero et al., 2019).

In this paper, we present how we evaluate and improve user engagement with a digital DA using a series of iterative formative studies. The main goal of the DA we examined, is to help surrogate decision makers of incapacitated neurocritically ill patients (with severe traumatic brain injury, hemorrhagic, or large ischemic strokes) to prepare for clinician-family meetings. During this meeting, surrogate decision makers are asked to make a goals-of-care decision. That is, surrogates must choose a treatment pathway between two options: survival or comfort. Choosing survival means that the patient will continue receiving invasive medical therapies. Choosing comfort care means that the patient will have life-sustaining measures withdrawn while the patient is provided with medications for comfort; the patient is allowed to pass away with as little suffering as possible (Barry & Edgman-Levitan, 2012; Goostrey et al., 2021; Muehlschlegel et al., 2020).

2 Improving the DA with Iterative Formative Studies

The information provided by the pilot digital DA in our study was originally displayed on 18 pages with a left navigation bar (Norouzi Nia et al., 2021). The DA explained the available treatment pathways, visualized an estimated prognosis for the patient via an icon array with data derived from validated disease-specific prediction models, summarized the information important for decision-making in a table to compare treatment options, and provided a worksheet to be completed by the surrogate as a value-clarification exercise. The DA also provided two real patient/family examples.

To improve the design of our DA, grounded in a user-centered approach to product development, we conducted a series of iterative formative UX studies (Djamasbi & Strong, 2019). In two of our four formative studies, we used eye tracking to understand how people process the DA's information. We also used eye tracking to conduct a retrospective "think-aloud" protocol after participants completed reviewing the DA at their own pace. Retrospective "think-aloud" protocol is particularly helpful in the assessment of complex systems because it allows participants to complete the task without interruption (Eger et al., 2007; Schiessl et al., 2003). Similar to generative UX research that uses stimuli such as mockups and/or storyboards to gain a deeper understanding of user needs, retrospective "think-aloud" protocols serve as an ideal tool to engage users in a conversation about their thoughts and feelings.

To facilitate an efficient and effective development process, as customary in formative UX studies after collecting data from a few participants, the research team convened regularly to decide whether the collected data provided enough information to adjust the

DA. This process resulted in four iterative formative studies, through which we collected information to adjust the DA and evaluated the impact of the adjustments we made on users' engagement with the DA.

2.1 Iteration 1: Collecting Information for Revising the Original Pilot DA

The objective of the first iteration was to gather feedback about the pilot DA's content and organization from the clinicians' and surrogates' perspectives. Feedback from clinicians was important to improve the accuracy and flow of information based on clinicians' experience providing this information to their patients and their families. Feedback from the surrogates was important to improve the engagement design with the DA.

We performed remote (Zoom) interviews with seven multidisciplinary clinicians such as neurosurgeons, stroke doctors, and palliative care providers. We also conducted interviews with five surrogate decision makers of prior neurocritical ill patients with the diseases of interest. Participants were provided with a link to the DA via the chat function in Zoom and were asked to share their screens. Using the "think-aloud" protocol each participant reviewed all 18 pages of the DA. Participants were encouraged to provide suggestions for improvements as they reviewed the DA. We also asked them to give us overall feedback at the end of the interview.

The feedback from participants provided strong support and encouragement for the development of the tool. In particular, the availability of accurate and reliable information (e.g., estimated prognosis visualized via the icon array and information about possible treatment pathways and their pros and cons) was considered highly valuable. The feedback also indicated that the DA would benefit from reorganizing some of its content. For example, clinicians suggested changing the order of some of the provided information. They also made suggestions about formatting the textual content in a way that the most important information is moved to the top of the page and emphasized with boldface. Including more images to accompany the text and short video clips that provided the same textual information in a non-textual format was recommended to improve engagement with content by all participants, i.e., both clinicians and surrogates with lived experience. Similarly, it was suggested to provide users with the option to review or skip patient examples.

2.2 Iteration 2: Revising the DA and Evaluating the Changes Made

Based on the feedback gathered in the first iteration, we reorganized some of the DA's content: 1) we reduced the number of DA pages from 18 to 14; 2) we provided users with the option to review or skip patient examples before completing their worksheet; 3) we simplified the language in certain paragraphs; 4) we moved text that was identified by clinicians as important information to top locations on pages and used formatting (e.g., bold text) to draw attention to important information; 5) we added more images with special emphasis on inclusion (i.e., images depicting people of color) to accompany the text; and 6) created six new videos with a physician explaining the content in plain language and embedded these videos in six different pages of the DA.

After implementing these changes, we conducted another remote round of interviews. We recruited one clinician and two surrogates with lived experience who participated in the first iteration of our user study to examine whether they preferred the revised DA over the original one. Again, participants were asked to share their screens while reviewing the DA and providing feedback.

All participants in this second round found that the revised DA, compared to its original version, made it much easier to process the provided information. Participants indicated that they preferred the way information was organized and presented to users in the new version of the DA. They found that the newly added images and videos made the revised DA notably easier to understand than its original version.

Summary of the Result of Iterations 1 and 2

The design objective of the first two iterations of our study was to see whether we can improve the organization of the provided information to enhance user engagement with the DA. To do so we collected feedback from seven multidisciplinary clinicians (e.g., neurosurgeons, stroke doctors, and palliative care providers) who help surrogates with decision-making in the neuro intensive care unit (neuro ICU) to improve the flow of DA's content. We also gathered feedback from five surrogates with lived experience to see how we can improve user engagement with the DA. The feedback from clinicians and surrogates helped us to make major revisions in content organization and presentation to users. The revised DA was ready for a more in-depth evaluation of user engagement with a new set of surrogates which will be discussed in the following sections.

2.3 Iteration 3: Assessing User Engagement with Eye Tracking

The data for this iteration was collected at a university-based hospital in New England. Participants were recruited from the neuro ICU waiting room among individuals whose family member was admitted to the neuro ICU as a patient. We used eye tracking to capture how people read the DA's content. We also used the gaze data from three specific pages deemed to be the most important aspects of the DA by our physician author (i.e., the prognosis page that contained the icon array, the page comparing treatment options, and the worksheet) to conduct a retrospective "think-aloud" session. The retrospective "think-aloud" process allowed both the collection of evaluative user feedback that was cued by their gaze data (Elling et al., 2012) and facilitated engaging users in a conversation that helped us gain a deeper understanding of their thoughts, feelings, and needs.

We invited participants (two men and one woman, mean age = 62.33, SD = 5.13) to a room adjacent to the neuro ICU where we had set up a laptop with a Tobii X-60 eye tracker attached to the laptop screen. The laptop used Tobii Pro Lab software (version 1.18) to calibrate the eye tracker for each participant. This setup allowed us to capture participants' gaze data as they reviewed the web-based DA. Participants were randomly assigned to review one of the three available DAs (i.e., traumatic brain injury, large acute ischemic stroke, and hemorrhagic stroke). Once participants finished reviewing the DA, we showed them their gaze replay for the three aforementioned pages. We ask them to recall their thought processes as they reviewed their gaze videos. Finally, we asked participants to tell us about their overall impression and experience with the DA.

All three participants reported an overall positive experience with the DA. The feedback gathered through the retrospective "think-aloud" for three specific pages, indicated that we could further improve the user experience of DA by adjusting the content of the table that compared treatment options. Participants indicated that they were partially confused by the information presented in the table. For example, user feedback during the retrospective "think-aloud" and exit interviews included statements such as: "I am not sure what to do with the table" and "The table is confusing". This table was displayed twice in the DA, first on page 9 and then in the worksheet which was provided as the last page of the DA.

The collected eye-tracking data was analyzed by focusing on fixation, saccade, and visit metrics. Fixations refer to slow eye movements that we use to process visual information. Saccades refer to fast eye movements that we use to change our focus of attention. Visits refer to a sequence of fixation and saccades that we use to view an area of investigation (AOI), i.e., a specific area on a visual display (Djamasbi, 2014).

Figure 1.a shows the fixation duration heatmap for page 9 where the comparison table for goals of care is shown for the first time. This color heatmap displays participants' cognitive effort by visualizing their fixation duration intensity from highest (red) to medium (yellow) to lowest (green). The heatmap in Fig. 1.a reveals a scattered viewing pattern. This heatmap also shows that the bottom cells in the table, which explain reasons for avoiding survival and comfort, received more intensive fixations as indicated by more and larger red color clusters in these areas. Similarly, the heatmap indicates more intense fixation on the column heading "survival" compared to the column heading "comfort."

Figure 1.b displays the order by which various sections of the table were reviewed by the three participants. The map in Fig. 1b. was developed by calculating the average time to first fixation for each AOI (Djamasbi, 2014). The AOIs in Fig. 1.b are displayed as dark gray boxes. The AOI gaze order map shows that participants reviewed the cells related to survival before looking at the cells that provided information for comfort. They also show that users did not read the page title ("Comparing and summarizing different goals of care") and table titles ("Survival" and "Comfort") before reading the content of the table. This viewing behavior is important because designing the page in a way to encourage users to read the titles before looking at the content makes it easier for them to understand the provided information (Djamasbi et al., 2012).

The fixation order as well as the scattered viewing pattern in the heatmap in Fig. 1 indicates that participants were looking around for information. This interpretation is supported by the results of the retrospective "think-aloud" which indicated that users were not quite sure how to process the content of the table.

1.a. Fixation duration heatmap 1.b. AOI map

Fig. 1. Iteration 3: Fixation duration heatmap and AOI map for page 9

The heatmaps in Fig. 2 display the distribution of attention on the table on page 9 and the same table on the last page of the DA (page 14, the worksheet page). Figure 2 also displays three quantitative eye-tracking metrics that capture cognitive effort for reviewing the tables on page 9 and the worksheet page: average visit duration, saccade-to-fixation frequency, and fixation-visit-duration. Average visit duration captures the average amount of time for every single visit (every single time that the table area was visited). Higher average visit durations indicate more effort expended to view the table. Saccade-to-fixation frequency (i.e., total number of saccades/total number of fixations) reveals changes in focus. Higher values of saccade-to-fixation frequency indicate more changes in focus when viewing the table (e.g., more intense search behavior) (Wu et al., 2015). Fixation-to-visit duration (total fixation duration/total visit duration) reveals the effort expended to read the content of the table. The larger the value of fixation-to-visit duration, the more effort is expended to read the content (Pool & Ball 2005).

Because the table is presented to users twice, we expected to observe differences in viewing patterns and behavior when the table was viewed for the first time on page 9, compared to the second time that it was viewed on the last page of the DA (the worksheet page). Due to familiarity with the table, it is reasonable to expect that participants would exhibit less cognitive effort (i.e., less intense viewing behavior) when they view the table on the worksheet page. The data displayed in Fig. 2, however, does not support this expectation. As shown in Fig. 2, participants' fixations on both tables had a similar scattered pattern. Additionally, quantitative eye-tracking metrics did not show a decreasing trend in cognitive effort. For example, the results showed that on average people exhibited more cognitive effort when viewing the table on the worksheet page (56,576 ms) compared to when they viewed it for the first time on page 9 (46,213 ms). While saccade-to-fixation frequency values indicated slightly fewer changes of focus on the worksheet table compared to the table on page 9 (51% vs. 57%), fixation-to-visit duration ratios indicated that participants expended slightly more effort reading the table in the worksheet (67%) compared to the effort they expended to read the table on page 9 (64%). The information summarized in Fig. 2 indicates that participants did not exhibit an overall decrease in cognitive effort when they reviewed the table for the second time on the worksheet page. These eye-tracking results suggest that participants may have not fully understood what the table intended to communicate. Participants' feedback, which indicated confusion about the table, supported the interpretation of the eye-tracking results.

Visual Engagement with the Icon Array

We also investigated how participants visually processed the icon array. An Icon array is a graphical depiction of probabilities and proportions, which uses a matrix of icons. (e.g., probability of dying from an injury and/or surviving it with serious disabilities, etc.) (Scalia et al., 2021).This information is essential for surrogates to understand the probability of a predicted outcome to help them select a treatment pathway that most likely matches the values and preferences of the patient. Hence, the icon array in our DA is presented to users before the table that compares treatment options on page 9.

Figure 3 displays the heatmap for the icon array used in the DA to visualize the probability of death, survival with severe disability, and survival with minimal or no disability, all derived from disease-specific, validated prediction models. As shown in

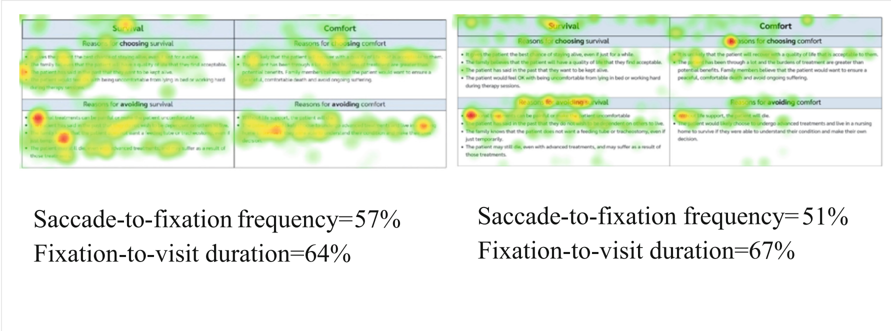


Fig. 2. Iteration 3: Attention, search, and reading behavior

Fig. 3, the aggregated gaze patterns cover mostly the right side of the icon array indicating that participants mostly focused on the icon array legend. This viewing behavior can be explained by the fact that the only information conveyed by the graph on the icon array is the visual representation of the proportion of three possible outcomes (death, survival with severe disability, and survival with no or minimal disability). The explanations for these outcomes are provided by text in the legend of the icon array.

People's fixation intensity tends to decrease from top to bottom when they read textual information that is presented in a list format (such as the textual information in the icon array). As shown in Fig. 3, the explanation for the legend (the paragraph above the legend) received relatively less intense attention compared to the two first two items of the three-item legend. The color clusters on the legend indicate that participants read the first two items of the legend with relatively similar intensity. They also indicate that the first two items of the legend were viewed more intensely than the last item of the legend.

The eye-tracking results and feedback from participants indicated that we did not need to make any changes to the icon array.

Fig. 3. Iteration 3: Fixation duration heatmap for Icon array

2.4 Iteration 4: Assessing User Engagement with Eye Tracking After 3rd DA Revision

The results of the user study in iteration 3 provided a number of actionable insights for revising the table that compared the survival and comfort treatment options. To make the table easier to process, we simplified its content in several ways. For example, to make it clearer that the treatment goal must be considered from the patient's point of view we changed the subtitles in the table from "Reasons for choosing <treatment goal>" to "The patient may choose <treatment goal> because:"). We also simplified the language in bulleted points (e.g., changed "The patient has said in the past that they don't wish to be dependent on others to live" to "The patient does not wish to be dependent on others to live"), and balanced the number of bulleted points for each section in the table (i.e., three bulleted points in each cell). Additionally, we simplified the title of page 9, where this table is shown for the first time, to create an easier-to-read summary of the entire page. The title of page 9 was changed from "Comparing and summarizing different goals of care" to "Comparing different goals of care". The table in the worksheet was the exact copy of the table on page 9 with one exception; The table was simplified on the worksheet by removing the column titles Survival and Comfort. Because patients were already familiar with the table, we did not expect any problem removing the table titles on the worksheet. Removing the table titles helped simplify the content-heavy worksheet page.

After making these adjustments, we resumed eye-tracking data collection in the same hospital, using the same room and setup. We again recruited individuals whose family member was admitted as a patient to the neuro ICU. Four new participants (three men and one woman, mean age = 44.25, SD = 20.04) were recruited. Again, participants reported a positive overall experience with the DA. The feedback from the retrospective "think-aloud" indicated that the adjustments made to the table improved engagement with its content because participants no longer reported any confusion about the table.

Figure 4.a displays the heatmap and Fig. 4.b displays the AOI map for the revised page. The heatmap in Fig. 4.a shows that participants had more intense fixations (red and yellow color clusters) on the survival column than on the comfort column. Figure 4.b shows the AOI map representing the order by which table content was viewed. As shown in Fig. 4.b, participants first looked at the page title, then at the titles of the table before looking at the table content. The observed viewing order suggests that the changes made to the page had a positive impact on how people reviewed it. As mentioned before, because titles and subtitles provide the summary of the content, they are important in effective communication of information (Djamasbi et al., 2012).

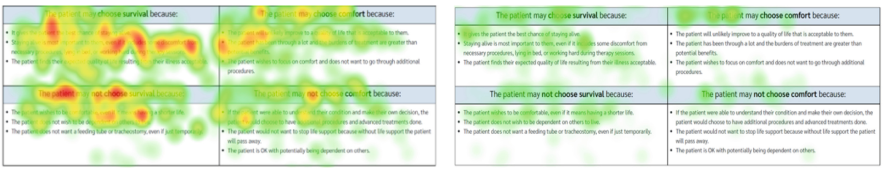
a. Heatmap b. AOI map

Fig. 4. Iteration 4: Fixation duration heatmap and AOI maps for page 9

Figure 5 displays heatmaps for the table when it was viewed by participants for the first time (Fig. 5.a) and for a second time on the worksheet (Fig. 5.b). Because the table titles were not included in the worksheet, qualitative metrics were measured for the area of the table that was common on both pages (i.e., the table AOI in Fig. 5 does not include table titles). The heatmaps show that in this iteration participants read the column that explained the survival goal of care more carefully on both tables as evidenced by the spread and intensity of the color clusters in the tables.

As mentioned in iteration 3, because of familiarity with the table (i.e., participants already reviewed the table on page 9), we expected them to read the content of the table on the worksheet with less intense viewing patterns or cognitive effort. The eye-tracking results summarized in Fig. 5 support our expectation. For example, the heatmaps show less intense color clusters covering fewer areas on the worksheet table compared to the table on page 9.

The quantitative measures of average visit duration, saccade-to-fixation frequency, and fixation-to-visit duration also supported our expectation that the table in the worksheet, compared to the table on page 9, would be reviewed with less cognitive effort. As shown in Fig. 5, the average visit duration was notably shorter for the table on the worksheet (16,403 ms) than the average visit duration for the table on page 9 (50,956 ms). Similarly, saccade-to-fixation frequency and fixation-to-visit duration ratios indicated



Average visit duration = 50,956 ms
Saccade-to-fixation frequency= 48%
Fixation-to-visit duration=54 %
5.a. Page 9

Average visit duration = 16,403 ms
Saccade-to-fixation frequency= 41%
Fixation-to-visit duration=48%
5.b. Last page (worksheet)

Fig. 5. Iteration 4: heatmaps, attention, search, and reading behavior

less intense search and reading behavior for the table in the worksheet compared to the table on page 9 (41% vs 48% and 48% vs 54%). These results suggest that the adjustments made were effective in improving how the table communicated information to users. Participants' feedback indicating that they knew how to use the table supported the interpretation of these eye-tracking results.

Visual Engagement with the Icon Array

Because no changes were made to the icon array in iteration 4 (i.e., the results of iteration 3 indicated no adjustment was needed), the icon array heatmap in iteration 4 was created with the eye movement data for all seven participants in iterations 3 and 4 (Fig. 6).

The heatmap for the icon array in iteration 4 shows that the aggregated viewing pattern of all participants is similar to the viewing patterns of the first 3 participants in iteration 3 (Fig. 3). The gaze patterns in Fig. 6 cover mostly the textual information with participants' most intense fixation covering the first two items of the three-item legend. There are intense fixations (red clusters) on the textual information that reveals percentages (39 and 28 out of hundred) on the survival items (first two items of the legend) but no red cluster on the percentage for those who die from their injury (last item of the legend). Within the first two items of the legend, the larger number of red and yellow clusters on the first item of the legend indicates that the explanation for survival with no or mild disabilities was viewed more attentively than the explanation for survival with severe disabilities. The observed viewing pattern for the legend items could be due to their presentation order (i.e., attention decreases from top to bottom on lists). It could also be due to their content indicating that participants attended the information on survival outcomes more intensely than information on death due to injuries. The eye-tracking data in iteration 4 shows that attention to information about survival (compared to attention to information about comfort) was more intense on the tables on pages 9 and 14. While these results suggest that surrogates may exhibit attentional bias toward information about survival, future research is needed to examine such a possibility more directly.

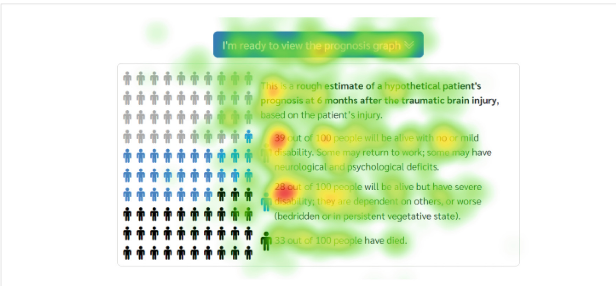


Fig. 6. Fixation duration heatmap for Icon array

Summary of the Result of Iterations 3 and 4

The objective of iterations 3 and 4 was to examine user engagement with the DA at a deeper cognitive level. The average time to review the DA for all seven participants in iterations 3 and 4 was 27.71 min. The analysis of eye movements showed that participants reviewed the DA without skipping any textual or image-based communication.

Because the comparison table appears twice (on pages 9 and 14) in the DA, we expected participants to exhibit less intense cognitive effort the second time they review the table. In iteration 3, participants' viewing patterns and behavior did not support our expectation; participants' fixation patterns and viewing behavior did not show an overall decreased trend in cognitive effort. In iteration 4, however, after we revised the DA, the eye-tracking results suggested that participants reviewed the table on page 14 with less cognitive effort. The differences in cognitive effort between reviewing the table on page 9 and page 14 suggest that the changes made to pages 9 and 14 were effective in helping users process the provided information more easily. User feedback supported the above interpretation derived from these eye-tracking results.

The heatmaps for the icon array in iterations 3 and 4 showed similar viewing patterns indicating that the information provided in this graph was thoroughly reviewed. The heat maps (Figs. 3 and 6) also showed more intense attention to the legend that described the possibility of survival with mild or no disability. Attention to survival was also observed in heatmaps in Fig. 5. After revising the table to make it easier to process, the column in the table that described survival was covered with more fixations on page 9 (Fig. 5.a) and more intense fixations on page 14 (Fig. 5.b).

The similarity of viewing patterns in Fig. 3 (the icon array heatmap generated for the first three participants) and Fig. 5 (the icon array heatmap generated for all seven participants) shows that we were able to capture the overall viewing behavior for the icon array with only three participants. Similarly, we were able to use the eye movement data of a small number of participants (i.e., the first 3 participants) to test the DA and generate actionable insight for revising it. The ability to generate actionable insight for design and the ability to evaluate the impact of revisions on cognitive effort with only a small number of participants (in our case 3 to 4 participants), highlights the value of eye tracking in iterative formative studies which by mere nature have small sample sizes.

3 Discussion

Here we showed four iterative formative user studies to gather actionable insight for improving the engagement design with the DA and evaluating the changes made. The DA in our study was designed to help surrogate decision makers of patients with severe brain injuries to make a goals-of-care decision. We used the "think-aloud" protocols in the first two rounds and eye tracking in the last two. During the eye-tracking recordings, we did not use the "think-aloud" protocol to avoid interruption of participants' interaction with the DA. On average, participants required 27.71 min to review the DA without interruption. Eye-tracking results showed that no textual or image-based communication was skipped by users. This is notable because research shows that textual information is rarely reviewed thoroughly (Djamasbi et al., 2016).

Our mixed methods analysis showed that our iterative revisions effectively improved participants' engagement with the DA. The results showed that users looked at the survival table column more intensely than the comfort treatment column. A similar attention pattern was observed for the icon array: the survival outcome received more attention than death. The observed viewing patterns may indicate surrogate decision makers' attentional bias toward survival-related information. They may also reflect the tendency to view information from top-to-bottom (e.g., legends in the icon array) or left-to-right (e.g., table content). Future research is needed to examine these possibilities.

Our study results highlight the value of eye tracking in evaluating engagement design. For example, similar gaze patterns captured by heatmaps in Figs. 3 and 5, suggest that stable viewing patterns can be captured with a small number of participants (e.g., n = 3 in our case). Similarly, our results show that capturing visual information processing behavior (e.g., via heatmaps, eye tracking metrics, and retrospective "think-aloud") even from a small number of participants can generate valuable insight for revising the DA and/or testing the effectiveness of its revisions.

A recent industry report points out the need for more user experience research that can generate actionable insight for improving products during the development cycle (User Zoom, 2022). The iterative "design, test, and revise" process in our study shows how to take advantage of various techniques (e.g., "think-aloud" protocol, eye tracking, retrospective "think-aloud") to gain insights for improving the engagement design of a DA. The "think-aloud" protocol in the first two iterations allowed us to gather feedback and suggestions for improvement from participants as they were viewing each page of the DA. This process allowed us to learn about reorganization and simplification of the DA's content. The eye-tracking protocol allowed us to investigate engagement with the revised DA without interrupting participants. This process allowed us to capture how the provided information was processed in the moment. Finally, the retrospective "think-aloud" helped us collect participants' feedback on specific pages by cueing them with their own gaze data.

4 Strengths and Limitations

Our study has important strengths. The mixed methods approach used in our study facilitated the evaluation of the engagement with the DA at a cognitive level not possible with the more traditional UX methods. It also facilitated a deeper and more nuanced understanding of user information processing needs. Our small sample size in each iteration could be viewed as a limitation; as such, the result should be considered with caution. However, the objective of iterative studies in our project was not to find statistically significant differences in the results, rather we intended to provide actionable insights for the development team in a timely manner. Conducting multiple studies with small sample sizes to provide timely and cost-effective insight for the development team is grounded in user-centered approach to product development (Albert & Tullis 2013; Djamasbi & Strong 2019). Additional limitations include that we did not collect user feedback for every single page of the DA, but limited our retrospective "think-aloud" investigations to the three most important pages of the DA.

5 Conclusion

Because our DA provides crucial information about the continuation or withdrawal of life-sustaining measures for a neurocritically ill patient, it is critical to effectively present the information needed for decision-making. The DA's content must be presented in a way that can be easily processed by surrogate decision makers. The results of our study show that including eye tracking in iterative formative studies can serve as a valuable and feasible methodology for assessing how provided content is reviewed by users. The results also show that the retrospective "think-aloud" protocol provides an important tool for gaining insights about users' needs and preferences.

Acknowledgement. . Research reported in this publication was supported by the National Institute of Nursing Research of the National Institutes of Health under Award Number R21NR020231. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

Albert, B., Tullis, T.: Measuring the user experience: collecting, analyzing, and presenting usability metrics. Newnes (2013)

Barry, M.J., Edgman-Levitan, S.: Shared decision making — the pinnacle of patient-centered care. N. Engl. J. Med. **366**(9), 780–781 (2012). https://doi.org/10.1056/NEJMp1109283

Djamasbi, S.: Eye tracking and web experience AIS. Trans. Hum.-Comput. Interact. **6**(2), 37–54 (2014)

Djamasbi, S., Chen, P., Shojaeizadeh, M., Rochford, J.: Text simplification and generation Y: an eye tracking study. In: SIGHCI 2016 Proceedings, p. 12 (2016)

Djamasbi, S., Siegel, M., Tullis, T.S.: Faces and viewing behavior: an exploratory investigation. AIS Trans. Hum. Comput. Interact. **4**(3), 190–211 (2012)

Djamasbi, S., Strong, D.: User experience-driven innovation in smart and connected worlds. AIS Trans. Hum. Comput. Interact. **11**(4), 215–231 (2019)

Eger, N., Ball, L.J., Stevens, R., Dodd, J.: Cueing retrospective verbal reports in usability testing through eye-movement replay. In: Proceedings of HCI 2007 The 21st British HCI Group Annual Conference University of Lancaster, UK 21, pp. 1–9, September 2007

Elling, S., Lentz, L., de Jong, M.: Combining concurrent think-aloud protocols and eye-tracking observations: an analysis of verbalizations and silences. IEEE Trans. Prof. Commun. **55**(3), 206–220 (2012). https://doi.org/10.1109/TPC.2012.2206190

Gaffiero, D., Elander, J., Maratos, F.: Do individuals with chronic pain show attentional bias to pain-related information? An early stage systematic review of the eye-tracking evidence. Cogn. Psychol. Bull **4**, 37–45 (2019)

Goostrey, K.J., et al.: Adapting a traumatic brain injury goals-of-care decision aid for critically ill patients to intracerebral hemorrhage and hemispheric acute ischemic stroke. Crit. Care Explor. **3**(3), e0357 (2021). https://doi.org/10.1097/cce.0000000000000357

Muehlschlegel, S., et al.: Goals-of-care decision aid for critically ill patients with TBI. Neurology **95**(2), e179 (2020). https://doi.org/10.1212/WNL.0000000000009770

Norouzi Nia, J., Varzgani, F., Djamasbi, S., Tulu, B., Lee, C., Muehlschlegel, S.: Visual hierarchy and communication effectiveness in medical decision tools for surrogate-decision-makers of critically Ill traumatic brain injury patients. In: Schmorrow, D.D., Fidopiastis, C.M. (eds.) HCII 2021. LNCS (LNAI), vol. 12776, pp. 210–220. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-78114-9_15

Poole, A., Ball, L.J.: Eye tracking in human-computer interaction and usability research: current status and future prospects. Encyclopedia Hum.-Comput. Interact. 211–219 (2005)

Scalia, P., et al.: Comparing the impact of an icon array versus a bar graph on preference and understanding of risk information: results from an online, randomized study. PLoS ONE **16**(July) (2021). https://doi.org/10.1371/journal.pone.0253644

Schiessl, M., Duda, S., Thölke, A., Fischer, R.: Eye tracking and its application in usability and media research. MMI-interaktiv J. **6**(2003), 41–50 (2003)

UserZoom. (2022). State of UX report: Bridging the gap between UX insights and business impact (2022). https://info.userzoom.com/rs/293-RDJ-600/images/UserZoom_State_of_UX_22.pdf

Wu, X., Xue, C., Zhou, F.: An experimental study on visual search factors of information features in a task monitoring interface. In: Kurosu, M. (ed.) HCI 2015. LNCS, vol. 9171, pp. 525–536. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-21006-3_50