**Evaluating Data Visualization  
An Information Dashboard in the College of Veterinary Medicine**

Clover, K.2, Hines, S2, Maassen, K2  
Kirk, R.1,2, Danielson, J.1,2

{kclover, shines, maassen, rakirk, jadaniel}@iastate.edu

Iowa State University College of Veterinary Medicine1

Human-Computer Interaction Program2

\*These authors contributed equally

**Summary**

We created a prototype that will enable the faculty of the College of Veterinary Medicine to make quick and accurate decisions regarding the curriculum. We tested the prototype using several students from a local summer research opportunity internship. The goal of testing was to determine if the prototype was reported as easy to use as well as to identify changes that could be made to improve the usability and efficiency of the dashboard. We paid particular attention to the topic of which graphs were rated most helpful by the participants and how difficult the participants felt that the tasks were to complete. After completing usability tests, we made changes to the most confusing and unhelpful parts of the dashboard.  
  
**Introduction**

The faculty and staff in the Iowa State University College of Veterinary Medicine are routinely presented with complex decisions relating to the development and implementation of their curriculum. Currently, they must look through several documents from various years that contain information pertaining to different classes and test scores. Information visualization software is being created to assist in the organization of information that is crucial to making these decisions. However, it cannot be assumed that moving this information from paper reports to digital media will reduce the effort involved in decision-making simply by virtue of being a new technology.

The long-term goal of this project is to discover if the information visualization dashboard currently in development will successfully organize information in a way that will simplify the decision-making process for the faculty and administrators in the College of Veterinary Medicine. Specifically, we are interested in whether or not use of the dashboard will result in less time spent on analyzing trends in past and current student performance while taking no extra time to learn the software. Also, we are interested in discovering if the dashboard may lead to more profitable decisions as supported by the displayed information.  
  
**Background**  
*Accuracy*

Barrett, Mondick, Narayan, Vijayakumar, & Vijayakumar (2008) studied the effect of the introduction of a visual dashboard in medical decision making. After their dashboard interface was introduced to the hospital workers, overall drug concentration levels were much more uniform and the appropriate dosage amount was administered a greater percentage of the time. In our study, however, we want to improve curricular decisions rather than medical decisions. Nonetheless, the Barrett et al study supports the theory of our project, which is that information visualizing dashboards may reduce the time and effort needed to make a decision in addition to increasing the accuracy and evidence supporting these choices (2008).

*Menu Design*

A study conducted by Read, Tarrell & Fruhling (2009), compared the expandable index menu in which the participant must click to see the submenu, to the frame-based menu in which all submenus show at the same time. Information from this study has been applied in the design of our prototype, especially in how the menus are configured. The study states that a menu is better at serving its purpose if it has substantial breadth (number of items per menu) and not as much depth (number of levels per hierarchy). In accordance with this research, our prototype is only two levels deep and each level is fairly broad. Read, et al. also found that a frame-based design allowed for easier navigation, item location, and better interpretation of menu labels. In our study, however, we will be using more of an expandable index menu design. Due to the large number of items that needed to be included, we decided to not show all items at the same time while still trying to keep the number of submenus small (Read, 2009).  
  
*Visualization*

In his article *Designing Executive Dashboards*, Thomas Gonzalez (2011) discusses how to make an effective executive dashboard in which the goal is to help participants make more effective decisions in a more timely manner than they would using previous informational methods. Gonzalez explains how “actionable” information gives the participant information that will make him/her able to use specific data to tackle specific tasks or duties in an organization; we have done our best to provide this kind of information in our dashboard. Gonzalez also discusses how to use labeling and various chart types. Because our participants will see the same kinds of graphs many times, the labels don’t have to stand out as much. We used bar and line graphs because this article explains that they are useful for various interpretations. We placed our navigation tree on the left-hand side of the dashboard because this is one of the first places the participants will look (Gonzalez, 2011).  
  
**Methods**

Participants were recruited from all REU (Research Education for Undergraduates) summer programs located at Iowa State University. We recruited twelve students total, all of whom were attending the Human-Computer Interaction REU internship and were involved in projects unrelated to this research project. Two of these participants completed pilot tests. Of the ten participants that completed the study, 5 were male and 5 were female. All participants were 18 to 25 years old. The study was approved by the Iowa State University Institutional Review Board.

A script covering the think-aloud procedure to be used during the test was read to each participant. Participants were then presented a functional prototype of the dashboard system using UserZoom testing software on a laboratory computer. The prototype was created in Dreamweaver. A tree view of the subjects is located on the left of the screen. Any subject can be clicked on to expand it and reveal that subject’s subcategories. Clicking on a subcategory reveals that subcategory’s information in four different graphs. The first graph to the right of the subcategory is the Sparkline. This shows the change in performance over time. Next to the right is the Bar Chart, which shows the current performance. To the right of the Bar Chart is the Annotated Timeline. This is the largest graph, and it shows the performance over time in more detail. The last graph is the Candlestick, which helps visualize the variation and distribution of the data in the subcategory.

Participants were then asked to complete a list of 11 tasks to the best of their abilities while speaking openly about their thought processes and decisions. Task 1 required participants to examine scores for the subject Microbiology. Task 2 asked how many subcategories microbiology had. Task 3 asked seven questions comparing current performance score, variation, and change over time among the three subcategories of Microbiology. In Task 4, participants examined the scores for three different subjects. Task 5 asked for the most recent score of each subject. Task 6 asked how many subcategories each subject had. In Task 7, participants answered seven questions, similar to those in Task 3, comparing a subcategory from each of the subjects. Task 8 asked participants to examine the scores of six subjects. Task 9 asked for the most recent score of each subject. Task 10 asked for the number of subcategories of three of the six subjects. Task 11 asked seven questions similar to those in Task 3, comparing one subcategory from each of the six subjects.

Tasks included determining trends using graphs, counting subcategories, and comparing data from different subcategories. When they completed a task, they wrote down the answer and clicked the success button to move on to the next task. After the tasks had all been addressed, a survey was distributed with questions regarding the participants’ opinions of the dashboard design, miscellaneous background data, and the self-reported ease of use of the design. We referred to *Research Methods* when writing the survey questions to ensure that the survey was well written and would allow us to accurately measure responses (Lazar, J., Feng, J. H., & Hochheiser, H. 2010).

Using UserZoom we tracked and recorded mouse click data during the test, and we also recorded the completion time of each task. We compared the accuracy of the participants’ decisions to the optimal result. We determined which tasks were more difficult to complete by observing the overall accuracy on each specific task. We looked for trends in the surveys and written observations to determine how helpful specific aspects of the prototype, such as graphs, were to the participants. We referred to *Handbook of Usability Testing* when setting up our procedure for testing (Rubin, J. & Chisnell, D. 2008).  
  
**Results**

After looking through our data, we focused on several specific topics. We created graphs to determine the trends in topics such as the usefulness of our various graphs, the preference for graphs, the number of clicks per task, the time for task completion, and the accuracy of task performance.

In analyzing the usefulness of the graphs, we discovered that the annotated timeline received the highest rankings for usefulness in every one of the five categories measured, such as finding data and making decisions. The Sparkline graph had the lowest ranking in four out of five of these categories.  
  
Figure 1: This graph shows the usefulness of each graph in five different areas. Participants rated each graph on each of the categories, 7 being good and 1 being bad. The Annotated Timeline was the best overall and the Sparkline was the worst overall.

We looked through the survey responses to determine how many participants expressed a like or dislike for specific graphs. From this we discovered that the Annotated Timeline was the graph that participants mentioned liking the most. The Sparkline and Bar Chart were both disliked the most, however, two participants mentioned liking the Bar Chart, whereas no participant mentioned liking the Sparkline. The candlestick received more evenly spread opinions, with three participants expressing a like towards the graph and three expressing dislike.

Figure 2: This graph shows the preference for graphs amongst participants. We analyzed the survey results and noted if a participant mentioned liking or disliking a graph.

We made all of the following comparisons using a Paired Samples T-Test, and considered a p value of < .05 to be significant.

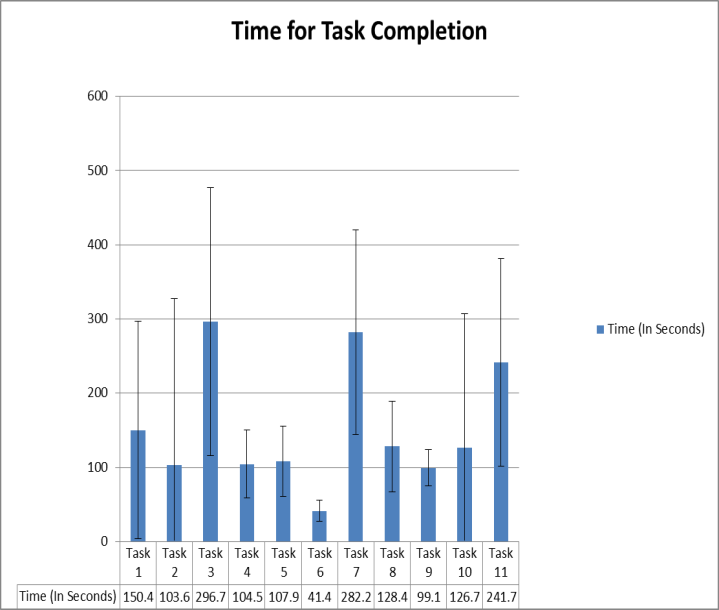
Using Userzoom, we were able to measure the number of clicks participants made on each task. We found that Task 2 and Task 6 both had a significantly lower number of clicks than the other tasks. Task 2 was significantly different from five of the other tasks and Task 6 was significantly different from two other tasks. In looking at the time it took participants to complete each task we found that Task 6 took less time to complete and was significantly different from two other tasks. In looking at the data, we discovered that there were three significantly different levels of difficulty: easy, medium, and hard. Tasks 2 and 6 were in the easy category, Tasks 5 and 7 were in the medium category, and Tasks 8 and 9 were in the hard category.

To determine the accuracy achieved on each benchmark task, determined how many participants correctly completed each task. We found that the first two tasks had a lower accuracy than the remaining tasks. From the survey and observations, we learned that many participants had a difficult time determining trends and understanding the data due to a lack of obvious labeling.

**Discussion**

From this study, we learned that the Annotated Timeline was the most helpful to and favored by the participants. The Sparkline, however, was not as helpful to or favored by participants. The first two tasks had less accuracy than the rest of the tasks, which can be seen in Figure 4. This might represent a learning curve. There was also some confusion among some users as to how to complete Task 2, which may have contributed to the low accuracy on that task. Overall, the participants’ performance became more accurate after the first two tasks, and remained accurate for the remaining tasks.  
  
Figure 3: This graph shows the accuracy of performance on each task. After the first two tasks, which had a lower accuracy, the tasks remained highly accurate. The 1st, 4th, and 8th tasks are not shown in this figure because they only required the participant to examine various subcategories. No answer was required for these tasks, so success could not be measured.

Tasks 3, 7, and 11 seem to be the tasks that took the most time. Tasks 3 and 11 also had a fairly high number of clicks. Tasks 3, 7, and 11 required participants to answer seven questions, which could be the reason for the higher numbers. Tasks 1, 4, and 8 all required participants to examine the scores for various subjects. These tasks seem to have a higher number of clicks than the other tasks, and the time taken to complete these tasks is similar to most of the other tasks. From this we see that these tasks seemed to require more clicks, but did not require more time.

 (a.)

(b.)

Figure 4: (a) This graph shows the average completion time (measured in seconds) for each task. (b) This graph shows the average number of clicks for each task.

Both of the graphs have error bars showing the standard deviation above and below each average.

We also learned that there are three levels of difficulty (easy, medium, and hard). Tasks 2 and 6 are in the easy category. Task 6 is significantly low in time and clicks, and Task 2 is significantly low in the amount of clicks. Tasks 5 and 7 are in the medium category. Task 5 seems to be fairly low in both number of clicks and time on task. Task 7 seems low in number of clicks, but has a higher time on task. Tasks 8 and 9 are in the hard category. They have an average time on task, but they both have a high number of clicks.

From the survey and observations, we learned that participants did not like the lack of labeling. From this we have learned that clear labels are very important. Through this study we learned many things that will help us come closer to creating an easy to use dashboard that helps in making faster and accurate decisions.  
  
**Implications and Future Work**

We will use what we have learned from testing this prototype in improving future versions of the dashboard. We have already implemented these results into the newest dashboard. The candlesticks were moved to be easily matched with the associated subcategory. Because the Annotated Timeline was so helpful, it has been made larger and has been placed where it will be easily seen first. Because the Annotated Timeline is now bigger, the labels should be easier to understand. The Bar Charts were replaced with bullet charts. These charts will hopefully be more helpful to participants, especially in determining performance. The Sparklines were kept, but they are more in line with the candlesticks and bullet charts. The layout makes it easy to associate the correct subcategory with all of its graphs. This change will hopefully create less dependence on the Sparkline, making it a more supportive graph, rather than a main graph.  
  
**Acknowledgements**  
We thank our advisor Dr. Jared Danielson and our graduate mentor Ryan Kirk for supporting us in this project along with the Summer Program for Interdisciplinary Research and Education. We also acknowledge the National Science Foundation for financially supporting this project through NSF Grant IIS-0851976.

**References**  
  
 [1] Barret, J. S., Mondick, J. T., Narayan, M., Vijayakumar, K., Vijayakumar, S. (2008). Integration of modeling and simulation into hospital-based decision support systems guiding pediatric pharmacotherapy. *BMC Medical Informatics and Decision Making, 8,1-12.* Doi: 10.1186/1472-6947-8-6.

[2] Gonzalez, Thomas. "Marketing036304." *OoCities - Geocities Archive / Geocities Mirror*. Web. 06 July 2011. <http://www.oocities.org/shaw4545/aa/marketing036304.htm>.

[3]Lazar, J., Feng, J. H., & Hochheiser, H. (2010). *Research Methods: In Human-Computer Interaction.* United Kingdom: Wiley.

[4] Read, Aaron, Tarrell, Alvin, Fruhling, Ann. "Exploring User Preference for the Dashboard Menu Design," hicss, pp.1-10, 42nd Hawaii International Conference on System Sciences, 2009.

[5]Rubin, J., & Chisnell, D. (2008). *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests.* Indianapolis, Indiana: Wiley Publishing