

Evaluation of a Data Dashboard to Support Resident Learning and Competency Assessment

Scott Vennemeyer¹, Milan Parikh¹, Shuai Mu, MDes¹, Benjamin Kinnear, MD, MEd^{1,2}, Danny T.Y. Wu, PhD¹

University of Cincinnati College of Medicine, Cincinnati, OH ¹

Cincinnati Children's Hospital Medical Center, Cincinnati, OH ²

ABSTRACT

This study aimed to evaluate a dashboard designed by the Internal Medicine Residency Program Clinical Competency Committee at the University of Cincinnati by conducting semi-structured interviews and a card-sorting activity. We found that users of the system were well equipped to share actionable insights about the current system and that they provided useful commentary that will inform the design of a new system that is more dynamic and that will fit the needs of the current users more effectively.

Keywords: user-centered design, health IT, visual analytics

1 INTRODUCTION

In 1999, the Accreditation Council for Graduate Medical Education (ACGME) introduced 6 core competencies for residency training assessment¹. In 2012, the standards were updated with the goal to improve education outcomes and reduce the burden of current structure and process-based approach². Additionally, the Internal Medicine residency training program at the University of Cincinnati College of Medicine (UC CoM) uses data collected using these standards for both summative (can the residents be promoted or graduate) and formative (to help improve skills) purposes. To facilitate the competency assessment and reporting process, the program has developed a spreadsheet-based dashboard, updated monthly, (Figure 1) for the clinical competency committee (CCC) to review and verify the milestones of the trainees. Because of the ever-growing amount of educational data, the current dashboard must be updated by creating an informatics solution that will maximize the value of the data for CCC decision-making. Before a new dashboard can be created, however, a thorough analysis of the old system must be conducted to understand the needs of the end users and stakeholders.

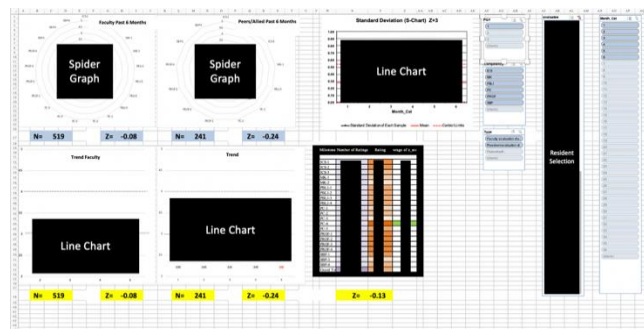


Figure 1: Home page of excel based dashboard, confidential information removed

There has been a trend in medical education literature to use data-driven approaches to support resident training and the decisions of the CCC. For example, Friedman et al in 2016 developed a resident dashboard providing both quantitative and qualitative data. (PMID: 27037226) Others utilized electronic health records and quality improvement methods to create timely reports for CCC

meetings. (PMID: 28638518, 31428276, 31636830) Recently, Thoma et al employed design-based methods to develop a thematic framework to inform the CCC needs and guide the dashboard implementation. (PMID: 32215140) Our study adds to this trend by evaluating an existing resident dashboard with a focus on usability and residents' needs.

As a whole, this series of projects aims to answer the following research questions: 1) what are the information needs of the CCC and the residents? 2) what are the gaps and design opportunities in these two viewpoints? and 3) how can we redesign the current dashboard to support all their needs? This abstract focuses on analyzing the results of a card sorting activity with the goal of learning how participants utilize an existing system in order to formulate design recommendations for our next step in process.

2 METHODS

This preliminary study was conducted with members of the Internal Medicine Residency program CCC at UC CoM. A total of 11 physicians participated. 8 of them were Attending Physicians and the remaining 3 were Chief Residents in the program. Two of the participants are excluded from this analysis because they did not complete the card-sorting activity (see section 2.2). The participants were chosen using convenience sampling. This study was reviewed and approved by the UC Institutional Review Board (IRB# 2019-1418).

2.1 Semi-structured interviews

In order to determine how the participants use the dashboard on a regular basis, a series of semi-structured interviews were conducted. Due to the COVID-19 pandemic, interviews were conducted through online video calls and screen sharing. These interviews lasted approximately 45 minutes and were recorded and transcribed verbatim by the interviewer (first two authors). The interviewers reviewed each other's transcriptions as a quality check to guarantee the accuracy and completeness of the transcriptions. The privacy of each participant was kept by de-identifying them and referring to them as "P00"- "P10" in all interviews. No specific patient information was collected. Twelve questions were asked in total. For the purpose of this study, the analysis is limited to the card sorting portion of the interview, lasting 10-20 minutes of the total interview time. The remainder of the questions will be analyzed in future work.

2.2 Card Sorting

During the qualitative interview, the participants were asked to participate in a card sorting activity. This was designed to help our team understand how the users viewed sections of the dashboard and how they were used. A total of nine cards were presented to each participant using an online platform that allowed "drag and drop" style interaction. Each card depicted a separate table, chart, or feature of the current dashboard. First, participants were asked to rank each of the cards by "value". In other words, how valuable they found the information presented by that aspect of the dashboard. After this, the initial ranking was covered from view and

participants were asked to rank the cards again by frequency of use. After this, the first round of cards was uncovered and the participants were asked to discuss any discrepancies between the lists. Quotes were pulled from the transcription of the interviews that describe how the participants came to their decision.

2.3 Data Analysis

The ranking assigned to each card, for each of the two questions, was recorded by participant and averaged. The average score of each card by question was then compared to determine an overall combined ranking by value of the cards and by frequency of use. Data analysis was conducted using Microsoft Excel.

3 RESULTS

The results of the card sorting activity are displayed in Table 1. We identified three clusters of scores when comparing value and frequency scores. The top ranked items are highlighted in yellow. These included the spider graph and line charts that can be seen in figure 1, as well as an additional "comments by month" section that is on another page of the dashboard. Next, the items in orange were in the middle tier. The red items were in the bottom tier and were consistently ranked the lowest.

Card	Average Perceived Value	Rank by Value	Standard Deviation	Average Frequency of Use	Rank by Frequency	Standard Deviation
Line Graph	1.31	1	0.46	1.88	1	1.36
Spider Graph	2.83	2	2.46	3.31	3	2.07
Comments by Month	3.00	3	1.27	2.69	2	2.19
Ratings Count Chart	5.38	4	1.79	6.00	6	1.65
Lowest OPA's Past 6 Months	5.63	5	1.16	5.81	5	2.07
Milestone Review	5.88	6	2.01	4.88	4	1.91
Main Heatmap	6.44	7	0.93	6.44	8	1.71
OPA's > 25	6.88	8	1.53	6.19	7	1.81
SD Chart	7.81	9	1.98	7.81	9	1.98

Table 1: Card Sorting – Average Rank and Relative score for ranking by value and by frequency. Clusters of scores are colored

The most beneficial result of conducting a card sorting analysis is that it allowed us to directly obtain feedback from the end users about components of the dashboard. For example, the highest rank for both value and frequency was the "line graph". Participants remarked that "it gives us a good sense of trajectory. It gives us a really quick sense, both with the trendline itself and also the z score that's underneath it, of how they're performing compared to expected" (P00) and that it allowed them to "view the whole gestalt" (P05) as compared to other aspects of the dashboard.

Feedback about our lowest placed components was useful as well. A consistent finding from almost every participant was that the Standard Deviation (SD) chart, meant to show "hidden variation" in the scoring provided by the evaluators, was often confusing and could be hard to interpret. Because there are many residents to review and a lot of information to interpret, the consensus, best argued by P08, seemed to be that the information on this chart could be useful but it was not consistently used and was ranked lower because "sometimes we look at it, and sometimes we don't" (P08). P06 remarked, however, that "I actually find myself looking more frequently at the standard deviation... I just have a tendency to overall scan the page and that just happens to be there in terms of frequency". This is a useful distinction that was captured by conducting two card sorting activities (once by value and once by frequency of use).

For the middle tier items, it was often useful to hear how the participants use them in order to discover why they were not ranked more significantly. For example, P00 mentioned that "[Observable Practice Action's reviewed less than 25 times chart] is a little bit less helpful in terms of making big summative decisions, but we certainly use it when trying to develop feedback." By correlating feedback such as this, which is somewhat inconsistent with an overall value score of 8 and a frequency score of 7, we are able to learn that this table may not need to have a prominent place on a

newly designed dashboard, but should still be accessible so that larger decisions can be made about the progress of a resident.

4 DISCUSSION

In this study, we conducted card sorting activities combined with qualitative interviews to prompt participants to share their own ideas about how a system could be designed to benefit them by learning about how they use an existing model. Through these complementary methods, we were able to gain actionable insights into how the end users of a system respond to features and how we can avoid current pitfalls in our own future design processes.

4.1 Implications

Conducting a card sorting analysis also yields information beyond the tasks given. Users such as P06 reflected upon their ranking and noted that "if we had something where the dashboard would be more of sort of a longitudinal tool, where we could be able to quickly assess those graphs, not going back in or not just looking at sort of this dashboard for this particular month, but for this particular learner over the past three years...". These types of findings are very useful after the evaluation phase and during the design phase. We are currently consolidating the workflows of the participants and organizing their pain points using an affinity diagram. These additional analyses will lead to more detailed design implications and recommendations, which will be published in the full manuscript.

On a different note, for practitioners that wish to conduct their own card sorting experiment to redesign a current system, we suggest that initial interviews be conducted with a stakeholder in the project if no prior information about the dashboard is known by the research team. This will allow for the team to choose elements of the dashboard that are extremely relevant to be chosen as "cards" rather than choosing items on the dashboard that are irrelevant.

4.2 Limitations

Card sorting, although useful, can limit the feedback that is received from participants. By forcing participants to rank a card that represents an aspect of the dashboard, they may be forced to boil down the function of that card to less than it realistically is. Luckily, the qualitative interview portion allowed us to hear their feedback about the cards so that we could correct for any variation in placement between participants for this reason. Finally, two of our participants did not participate in the card sorting interview. P01 was offered and took a job at another institution and was unable to participate in a follow-up card-sorting interview. P04 does not spend time utilizing the dashboard because they are responsible for preparing the data for use by running it through a regression model and, therefore, did not participate in card sorting.

4.3 Future Work

This project has several future directions. Using inductive coding, our research team will analyze the rest of the qualitative interviews. We will also interview current residents to hear their feedback so that a new dashboard can be designed with their needs in mind. As we design the new dashboard, we will look to create a system that is flexible and automatic, allowing for multiple users, different views, and drill-down analysis. We will also conduct thorough usability testing to validate our product.

4.4 Conclusion

We conducted qualitative interviews and card sorting with physicians from the UC CoM Department of Internal Medicine to evaluate a static competency assessment dashboard. Based on our findings, we will design a new dashboard that addresses the

needs of the committee and that allows for residents to learn from and improve based on their scores.

REFERENCES

[1] Swing SR, Clyman SG, Holmboe ES, Williams RG. Advancing resident assessment in graduate medical education. *J Grad Med Educ.* 2009;1(2):278-286. doi:10.4300/JGME-D-09-00010.1

[2] Nasca TJ, Philibert I, Brigham T, Flynn TC. The next GME accreditation system--rationale and benefits. *N Engl J Med.* 2012;366(11):1051-1056. doi:10.1056/NEJMSr1200117