

# Visual Analytics Opportunities and Usability Assessment of a Resident Clinical Competency Assessment Dashboard

Scott Vennemeyer, BS\*

Department of Biostatistics,  
Health Informatics, and Data  
Sciences

University of Cincinnati  
College of Medicine

Ting Zhang, MDes†

Department of Biostatistics,  
Health Informatics, and Data  
Sciences

University of Cincinnati  
College of Medicine

Eric J. Warm, MD

Department of Internal  
Medicine,

University of Cincinnati  
College of Medicine

Danny T.Y Wu, PhD

Department of Biostatistics,  
Health Informatics, and Data  
Sciences

University of Cincinnati  
College of Medicine

## ABSTRACT

We conducted a usability assessment of a resident clinical competency dashboard among eleven Clinical Competency Committee members using the System Usability Scale (SUS) and open-ended questions. Although the average SUS score was 53.47, indicating below-average usability, qualitative feedback highlighted strengths in data integration and interface layout but identified technical issues and requested features for improvement. Findings will inform future development efforts for the dashboard.

**Index terms:** Visual Analytics, Medical Education, User-Centered Evaluation, Residency

## 1. INTRODUCTION

Clinical Competency Committees (CCCs) are groups that use a variety of clinical and educational data to make formative and summative assessments [1], [2]. These include decisions about a resident physician's progress, growth, and readiness to practice unsupervised. Although there are many ways to support training and educational efforts, one of the most important ways is by utilizing the data collected by programs to create dashboards. Dashboards and other visual analytics approaches help medical educators to track, understand, and monitor a resident's progress close to real time. Over the past several years, our team has undergone an iterative design process to develop a dashboard that can support the CCC's evaluation process. Specifically, the CCC at our institution uses the dashboard to visually explore both qualitative and quantitative feedback from frontline assessors.

Our prior work, published in Applied Clinical Informatics [3] conducted qualitative analysis of interviews with CCC members to generate four design recommendations: 1) CCCs designing dashboards should integrate quantitative and qualitative feedback into their design, 2) dashboards should create multiple views to display data based on user roles, 3) programs should work with designers to create a usable, interpretable dashboard, and 4) teams must develop a strong informatics pipeline to manage the system. We also used specific feedback from this

work to decide where to place the visualizations and to design a new system that would better serve the needs of the CCC.

In this paper, we are seeking to improve the usability of the initial prototype of the dashboard by assessing the usability and understanding the current and potential visual analytics capabilities of the dashboard. This type of usability assessment is critical to ensuring that the end users of the tool (CCC members) have the best experience possible.

## 2. METHODS

### 2.1 Setting

The Internal Medicine (IM) Residency Program at the University of Cincinnati College of Medicine (UCCOM) utilizes a CCC to provide critical educational feedback to residents as they go through their time in the program. The CCC had ~28 members at the time of this paper and included the residency program director, associate program directors, educational faculty (attending physicians and hospitalists), and chief residents. Not every member of the CCC attends meetings or has used the dashboard, but all were invited to participate in this survey.

### 2.2 Dashboard Design

The current dashboard has been developed using the Flask library [4] for Python with custom visualizations created using D3 (data driven documents) [5]. The data is hosted on a MySQL [6] database server. A screenshot of the current dashboard is included below (Figure 1)

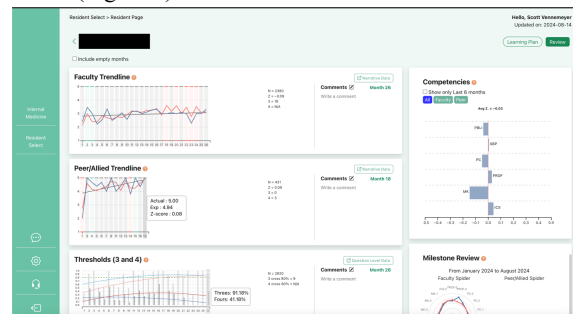
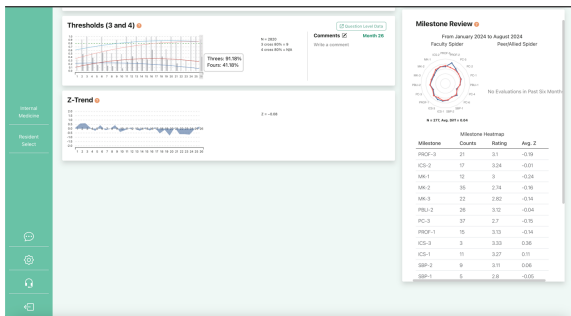


Figure 1a. Resident Page on the Dashboard Interface. Top of page.

\* e-mail: vennemsm@mail.uc.edu



**Figure 1b.** Resident Page on the Dashboard Interface. Bottom of page

Several custom visualizations have been created to help users understand the resident's progress. These visualizations include trendlines showing the average actual (blue lines) and expected (red line, from a regression model) entrustment ratings per month. Also on this chart is a black best fit line and colored bars to show when the z-score is above or below a predefined threshold. Next, there are spider charts to show average score (both actual in blue and expected in red) by competency as well as a table to show this information another way. Finally, there is a combination bar and line chart to show the percentage of ratings at/or above levels 3 and 4, a z-trend graph to show the average z-score of the ratings per month, and a competencies bar chart to show the overall z-score by competency being assessed. It is worth noting that each visualization allows for drill down analysis by connecting to additional data sources like narrative assessments and actual question level data by month to provide greater context to the scores.

### 2.3 Survey Creation

A survey was administered to collect usability information and to understand current areas of strength and potential opportunities for improvement. The participants were recruited from the CCC through convenience sampling. The survey was created in and deployed using Qualtrics. There were three sections of the survey: First, we collected demographic information to understand more about those CCC members that responded to the survey. These included questions designed to understand job titles, length of time on the CCC, whether the participant has used both the new and original dashboards, and the level of experience that the participant has with the dashboard. Second, we asked each user a set of standardized questions containing the System Usability Scale (SUS). This set of questions has been utilized in multiple areas of health and clinical informatics to understand the usability of software and interactive tools. We asked a slightly modified version of the questions as described in a similar work from our lab [7]. Finally, we wanted to understand the thoughts behind the usability scores. To do this, we asked each participant to provide three areas of strength for the dashboard and three opportunities for improvement.

### 2.4 Data Analysis

Survey results were analyzed in two ways. First, the SUS questions were scored according to the guidelines provided by Brooke [8] and Bangor et al [9]. No additional modifications were made to the scoring algorithm. Second, we performed thematic analysis to categorize the feedback provided by CCC members in the "Opportunities for Improvement" and "Areas of Strength" questions. Using the responses to these questions, we generated two findings for each question. Representative quotes were then

selected to illustrate the types of responses that led us to each finding.

## 3. RESULTS

### 3.1 Participant Characteristics

A total of 11 CCC members that utilize the dashboard filled out the survey. The users included Chief Residents, Hospitalists and Attending Physicians, and Residency Program leaders. On average, the participants have served on the clinical competency committee for approximately 3 years and only 64% (n=7) have used both the original excel based dashboard and the current dashboard being assessed in this work. 45% (n=5) of users described themselves as "Beginner Users, knowing only the basics of the tool". Out of the remaining 55% of participants (n=6), 2 described themselves as advanced users, 3 as intermediate users, and 1 described themselves as having no experience with the system at all and was therefore removed from the analysis, bringing the total number of participants down to 10.

### 3.2 Survey Results

Table 1 lists the full SUS ratings from each participant. Three participants left some questions blank when filling out the survey. To correct for this, the missing values were filled in with the average score for that question. Only one question had two missing SUS score values. The final combined SUS score was 53.47 (range: 35 – 67.5, std: 10.74). An average score of 68 or above is considered good usability, meaning that that our dashboard did not meet expectations. This was surprising for a variety of reasons but led us to be particularly interested in the qualitative feedback collected in the last two questions of the survey.

### 3.3 Item Analysis

Each question in the SUS instrument is scored differently to account for the positive or negative framing of the question. All odd numbered questions, for example, are framed such that a response of 5, strongly agree, is the desired answer. Even numbered questions are framed so that 1, strongly disagree, is the desired answer. Of the odd numbered questions, the lowest score (least desirable answer) in this study was for question 5, "I found that the various functions in this system were well integrated", with an average score of 2.91. For the even numbered questions, the highest score (least desirable answer) was for question 6, "I thought there was too much inconsistency in this system" with a value of 3.73.

### 3.4 Findings

The full list of findings and representative quotes are included in table 2 below. In the "Areas of Strength" responses, we found that users felt that the data was well integrated in the dashboard and that the tool was easy to interact with. Users appreciated that data was easily accessible, could be viewed in many ways, and connects numerical assessments to the corresponding narrative comments that describe why assessors gave their scores. They felt that the dashboard would be very helpful for residents and would make it much easier for them by allowing them to view all their data on one platform rather than in many different locations (as is currently the case). Second, the users felt that the UI was laid out well and was intuitive to use in most cases. Although participants were less descriptive here, they felt that many graphs made sense,

the visualizations were laid out well, and that the design makes it easy to identify trends in the data.

Responses for the “Opportunities for Improvement” generally centered around two areas. First, the users scored many sections of the SUS lower because of technical issues that prevent the system from reaching its full potential. Unfortunately, several of the issues they described are not specifically caused by our dashboard but rather are inherent to the data that fuels the dashboard. For example, several participants noted that there is a delay in how quickly evaluations show up in the dashboard, making it difficult to base clinical decisions on the data. This delay is due to the monthly export process from the data source, MedHub, which occurs at the end of each month to give time for the expected scores to be generated. The committee does not use the dashboard until 3 weeks later, however, so more evaluations have been completed in the interim. This leads to some lack of trust in the data because it is perceived as “out of date” or “behind”. Another example of an inherent issue with the data comes from blank evaluations. Some participants noticed that occasionally, they will try to view an evaluation but are not able to drill down because some of the responses were blank. This translates to a poor usability score for our system but is an issue with the underlying data.

Second, users provided several requests for features that they felt would make the system a better fit for their needs. These requests included features like enlarging graphs to make it easier to view information, adding options to hide resident names to protect privacy, logging in without having to log on to a virtual private network. This was because the current dashboard is behind a firewall of the medical school, but most residents would access the tool from the hospital, which is on a different network outside of the firewall.

## 4. DISCUSSION

### 4.1 Key Findings

Overall, the dashboard was not given favorable usability ratings despite some positive comments describing the visualizations and layout of the system. Many of the usability issues seem to stem from underlying technical and data issues that are preventing the system from reaching its full capacity in the eyes of some users. While we recognize that there are many opportunities for improvement, we believe that the efforts to create useful and user-friendly visualizations were largely effective. This is evident from the user responses to the survey because the users noted that the dashboard was well laid out and easy to navigate even while describing the technical issues that caused them to provide lower SUS scores.

There are many possible ways to address the concerns shared by the users. One important observation is that no training has been given to CCC members to teach them how to effectively utilize the system. A training session may help clear up some of the confusion and usability issues by helping the users to understand where the data comes from and how it is visualized on the dashboard.

### 4.2 Limitations

This work has several limitations. First, there is a small sample size of participants. With only 11 respondents, we cannot be sure that all opinions have been shared and that the themes have achieved saturation. However, there are only 28 members of the current CCC, meaning that 40% of the members responded to

the survey. This is enough, in our estimation, to gather data and understand a variety of thoughts and opinions about our dashboard.

Next, some participants did not fully complete the survey or did not answer the thematic questions usefully. Specifically, P5 had not used the dashboard but still filled out the survey. The thematic responses were not useful because they all said, “I would need to use the system to answer this question”. This led to the removal of P5 from the rest of the analysis in this work. Another example is the few questions that were not answered. This is likely due to a technical error during the survey creation that would allow the respondents to continue without answering those questions. The solution to this problem was described above and allowed for the calculation of the SUS scores while trying to maintain the integrity of the data.

Finally, this work collected highly specific feedback about our dashboard that is not necessarily generalizable to other institutions. While we understand that these may be significant limitations to the work, we believe that sharing our results in this venue would allow for others to understand how visual analytics solutions can be assessed for usability and improved based on feedback from stakeholders.

## 5. CONCLUSION

We conducted user-centered evaluation of a clinical competency assessment dashboard to understand how visual analytics affect the dashboard’s usability and to identify areas of strength and opportunities for improvement. Although the usability scores were less than desirable, the visualizations were noted as a strength. The opportunities for improvement are also addressable and provide excellent future directions for the development of our work.

## 6. TABLES

Table 1. SUS Scores by Participant

	SUS Score	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
P1	52.50	3	2	4	4	3	5	4	3	3	2
P2	65.00	4	2	4	3	4	4	4	2	4	3
P3	35.00	2	3	3	3	2	5	3	4	2	3
P4	52.50	4	3	4	4	3	3	4	2	2	4
P5	52.50	5	3	3	3	3	3	3	3	3	4
P6	55.00	4	3	3	3	2	3	3	2	3	2
P7	43.75	3	1	2	3	2	3	2.75*	3	2.75*	5
P8	67.50	4	2	4	2	4	4	4	2	3	2
P9	67.22	4	2	4	3.11*	3	3	4	2	4	2
P10	57.14	5	3.71*	5	3.71*	3	4	4	3.71*	3	2
P11	40.00	3	3	4	4	3	4	2	4	3	4
Avg.	53.47	3.73	2.52	3.64	3.26	2.91	3.73	3.43	2.79	2.98	3.00

\* Calculated score to replace missing value.

Table 2. Representative Quotes

Areas of Strength			
Finding	Participant	Representative Quotes	
Data is well integrated and easy to interact with	P2	“Having all of this information integrated into one site will make things a lot easier for the [clinical competency] committee as we are reviewing residents”	“...Mentors can give feedback and set goals directly on the ResDash site.”
	P4	“I like the feature that allows you to click a data point and see the narrative data”	
	P9	“Ease of ability to search for residents”	“Everything is in one place, don't have to open multiple spreadsheets”
The user interface is laid out well and easy to navigate	P3	“Easy to navigate”	“Good layout”
	P7	“Visual layout of multiple elements [is an area of strength]”	
	P8	“Some of the graphs really make sense”	
Opportunities for Improvement			
Technical issues prevent the system from reaching its full capacity	P1	“There is always something not working as expected. Either questionable data or an issue with a display. I don't fully trust the system.”	“The churn time when changing views can be cumbersome”
	P4	“There is a large amount of null data or blank evaluations. I wonder if there is a way to remove this?”	
	P9	“some data for some learners being pulled into incorrect sections, for example long block faculty evals pulling into peer data and comments”	“some residents missing data or other residents missing entirely”
	P3	“I'm not sure that all of the data is coming through accurately, which impacts usability significantly”	
Feature requests that would enhance usability and user adoption of the system	P7	“Ability to enlarge graphs visually to display information”	“An option to hover and see the name of the rotation on each data point.”
	P2	“It would be nice if residents were able to log on to ResDash remotely without the UC VPN”	
	P10	“Syncing/up-to-date with medhub evaluations”	

[2] A. Ekpenyong, J. S. Padmore, and K. E. Hauer, “The Purpose, Structure, and Process of Clinical Competency Committees: Guidance for Members and Program Directors,” *J Grad Med Educ*, vol. 13, no. 2 Suppl, pp. 45–50, Apr. 2021, doi: 10.4300/JGME-D-20-00841.1.

[3] S. Vennemeyer *et al.*, “User-Centered Evaluation and Design Recommendations for an Internal Medicine Resident Competency Assessment Dashboard,” *Appl Clin Inform*, vol. 14, no. 5, pp. 996–1007, Oct. 2023, doi: 10.1055/s-0043-1777103.

[4] M. Grinberg, *Flask web development*, First edition. Sebastopol, CA: O’Reilly, 2014.

[5] “D3 by Observable | The JavaScript library for bespoke data visualization.” Accessed: Sep. 16, 2024. [Online]. Available: <https://d3js.org/>

[6] “MySQL.” Accessed: Sep. 16, 2024. [Online]. Available: <https://www.mysql.com/>

[7] D. T. Y. Wu *et al.*, “Usability Testing of an Interactive Dashboard for Surgical Quality Improvement in a Large Congenital Heart Center,” *Appl Clin Inform*, vol. 10, no. 05, pp. 859–869, Oct. 2019, doi: 10.1055/s-0039-1698466.

[8] J. Brooke, “SUS - A quick and dirty usability scale,” p. 7.

[9] A. Bangor, “Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale,” vol. 4, no. 3, p. 10, 2009.

**ACKNOWLEDGEMENTS**

We would like to acknowledge Ms. Christy Li for her development efforts for the dashboard and for support calculating SUS scores.

**REFERENCES**

[1] K. E. Hauer, L. Edgar, S. O. Hogan, B. Kinneer, and E. Warm, “The Science of Effective Group Process: Lessons for Clinical Competency Committees,” *J Grad Med Educ*, vol. 13, no. 2 Suppl, pp. 59–64, Apr. 2021, doi: 10.4300/JGME-D-20-00827.1.