

Executive Summary

Kritik is an educational platform that is dedicated to improving critical thinking through peer assessment and iterative processes [1]. Due to challenges with the platform's current interface, particularly for instructors creating and scheduling activities, the project's focus is to enhance and streamline usability for educators in university settings. The *Understand* and *Create* phases of the project consisted of gathering data on user interactions with the interface, analyzing the information to reveal shortcomings, and creating a low-fidelity prototype. The executive summary of the *Evaluate* phase details this phase's usability testing goals, methods, findings, and recommendations intended to further streamline workflows to help reinforce Kritik's mission of peer evaluation and continuous learning.

The usability testing focuses on iteratively diagnosing issues and identifying areas of improvement in the prototype across five usability testing participants from the target user group of university teaching teams. Participants were asked to complete a series of usability tasks to test for the five factors of usability criteria (learnability, efficiency, memorability, errors, and satisfaction). By examining qualitative and quantitative data of how users interact with the prototype and measuring their performance, three usability problems were discovered [2]. The three problems were: navigating through the grouped sections, identifying formatting icons, and setting activity durations when scheduling, with low, medium and high severity respectively.

After analyzing the data, this phase recommends grouping features and implementing labels more intuitively. Additionally, it is recommended to employ easy to read and understandable usages of formatting tool labels. Lastly, since participants found that their mental model of deadlines and calendars were incongruent with the range calendar provided, future iterations should be adjusted accordingly. In conclusion, repeated iterations of the aforementioned actions will provide additional benefits to Kritik's overall interface and further streamline user's interactive processes.

1.0 Introduction

Kritik is a platform created to help students build their critical thinking skills through peer evaluation and feedback. In the *Understand* phase, the initial walkthrough of Kritik revealed frustration while creating and scheduling an activity. Since instructors would be the primary users of this function, the target users were identified to be university teaching teams as these individuals often have other responsibilities that require their time. The HTA was created to map out all steps used in activity creation, and ten users who had not used Kritik before were asked to explore the platform. This helped locate parts of the design that users felt needed change. Then, the team began the ideation process in the *Create* phase using several techniques to produce three alternative designs. These were then combined to make one prototype which was later finalized for user testing.

The purpose of this report is to detail the *Evaluate* phase of the redesign of Kritik's platform. To assess the functionality of the prototype produced in the *Create* phase, members of university teaching teams, the target user group, were invited to complete a series of tasks. The users' performance of these tasks were measured using quantitative usability metrics including errors, time, and number of clicks per task. With this data, the team performed descriptive analysis to assess the efficiency, memorability, and errors of the redesign. Additionally, qualitative data was collected to provide insights on the learnability and overall user satisfaction with the new design. The evaluation procedure is explained in Section 2.0, the results and analysis of the evaluations are presented in Section 3.0, and the final conclusion of the project is covered in Section 4.0.

2.0 Evaluation Procedure

As a result of in-person observations of several usability assessing tasks, both quantitative and qualitative data was collected from five participants of various university teaching teams. To inform analysis, several usability tasks were performed by participants, supported by structured/unstructured interview and a post-observation questionnaire. In-depth details are provided in the following sections.

2.1 Usability Tasks

To ensure that the new design addresses the most important parts of the shortcomings found in the previous phase, the usability evaluation tests for six tasks. The tasks and the corresponding HTA subtasks are listed in Table 1.

Table 1: Tasks and corresponding HTA subtasks performed during evaluation

Tasks	Corresponding HTA Subtasks (see Appendix A)
Navigate to the Creation page from the Activities page	Subtask 3
In Overall Learning Objective(s) textbox (without using keyboard): <ul style="list-style-type: none"> ● Identify and select the icons: hyperlink and embed, then clear all formatting 	Subtask 3.3.2
Set the minimum number of words to 28	Subtasks 3.4.3
Schedule “Create” part of the activity to be released August 20 at 9:15am and due August 25 at 9:15pm	Subtask 4.2 (including Subtasks 4.2.1 and 4.2.2)
Set the number of evaluations to 5	Subtask 3.4.4
Save activity	Subtasks 3.5 and 4.5

2.2 Usability Criteria & Testing Methods

The following testing procedure was used to maintain a consistent process for user testing.

1. Obtain the user’s consent, and detail expectations for the tests.
2. Provide the user with the prototype and list of tasks, then have them complete the tasks on a separate device while time, number of errors, and number of clicks are recorded by the team members.
3. Prompt users to fill out the post observation questionnaire.

To understand and analyze how the user interacts with the prototype, the following usability criteria, corresponding data type, and the particular testing method is listed in Table 2.

Table 2: Usability criteria, data type, and testing method performed during evaluation

Usability Criteria	Data Type	Testing Method (Observe and document)
Errors	Quantitative	Number of errors per task and type of error
Learnability	Quantitative	Time taken per task
	Qualitative	Post-observation questionnaire short answer response
Memorability	Quantitative	A similar task is repeated at end of test: <ul style="list-style-type: none"> • Time taken for task, number of errors for task and type of error
Satisfaction	Quantitative	Post-observation questionnaire utilizing Likert scale to produce SUS results
	Qualitative	Post-observation questionnaire open ended questions
Efficiency	Quantitative	Number of clicks per task, time taken per task

3.0 Results & Analysis

This section aims to use the data collection results (see Appendix C) to analyze the data to find successful implementations and areas for improvement (see Figure 1).

Observations	P1	P2	P3	P4	P5
General					
Previous Kritik Experience					
Dropdown labels are confusing					
Liked the dropdown menu/layout					
Switching between activity Creation and Scheduling was simple					
Liked that scheduling and creation were on the same page					
Confusion about instruction					
Activity Creation					
Able to locate <i>Create Activity</i> button with no errors					
Identifying Icons in Formatting Toolbar					
Labels are too small					
Able to identify the name of each icon					
Unable to identify the purpose of the icon					
Options					
Liked the ability to use manual input					
Navigation back to options for memorability test with no errors					
Schedule					
Calendar doesn't work the way user expects					
Unable to complete scheduling task					
Confusion about how to reset the schedule					
Suggested a prompt on calendar for release and due date					
Saved calendar date					
Saving Work					
Able locate <i>Save Activity</i> with no errors					

Figure 1. Compiled Observations Rainbow Sheet

During each usability test, the number of errors made per task was recorded to evaluate the capability of the design to prevent errors. Users made an average of five errors per walkthrough, with a task having at most nine errors (see Appendix D). Due to miscommunication from the instructions, many participants had issues with task two leading them to deviate from the intended task and making knowledge-based mistakes. The combination of errors from the aforementioned miscommunication, the fidelity of the prototype, and unfamiliarity with the functions of hyperlink/embed/clear formatting (see Appendix C) implies that these mistakes could be avoided and were not interface design problems.

Two methods were used to gain insight into the learnability of the design, the first being time taken per task. This was analyzed to see how quickly participants were able to adapt to the new platform. The average time per task was 46.7 seconds (see Appendix D), which can be mostly

attributed to task two as again many participants had difficulty understanding the task (see Figure 2).

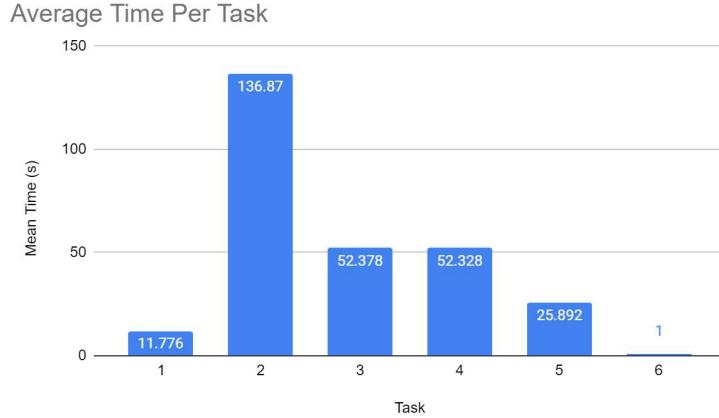


Figure 2. Bar graph of average times per task

Secondly, users were asked the question, “How did you feel about the user interface as a new user?”. All responses included positive words such as clear, intuitive, or easy, and the usage of Vader sentiment analysis confirmed an average of 0.4718 positive compound score for all responses (see Figure 3). Since Vader compound scores are bounded between [-1, +1] (strong negative to strong positive), a 0.4718 means a fairly positive overall sentiment [3].

```

fairly clear in terms of layout.
compound: 0.3818, neg: 0.0, neu: 0.658, pos: 0.342,

I found it quite easy to use, and not challenging
compound: 0.4116, neg: 0.124, neu: 0.602, pos: 0.274,

I don't usually like technological things, but this process was intuitive and clear.
compound: 0.4301, neg: 0.104, neu: 0.669, pos: 0.227,

Other than the "options" drop down, it was pretty intuitive.
compound: 0.2732, neg: 0.158, neu: 0.602, pos: 0.241,

I like it better than the existing platform because it feels like I know where things are more easily.
compound: 0.8622, neg: 0.0, neu: 0.551, pos: 0.449,

```

Figure 3. Vader Sentiment analysis of “How did you feel about the user interface as a new user?”

To measure the memorability of the new design, task five and task three test the same usability feature. In order to determine any improvements in the time taken in the second round of that task, a paired sample test was conducted to compare the two tasks. It was found that at the 0.1

significance level (90% confidence), the times for task five are less than task three. Since the 0.1 to 0.01 levels of significance are generally accepted in statistical analysis, the team decided to accept the 0.1 level due to the small sample size of participants [4].

To measure users' satisfaction with the prototype, the team performed sentiment analysis on the short answer responses in the post-observation questionnaire. Additionally, the System Usability Scale (SUS) score was calculated using the data collected in the post-observation SUS Likert questions. The design scored a standardized average of 74.5 (see Appendix D), which is above the average usability of 68 [5]. Thus, it is a satisfactory design.

Finally, the efficiency of the design was assessed by calculating time based efficiency using the formula below (see Figure 4). Given that these tasks involve scrolling through the interface, locating the task, and completing it, a time based efficiency of 0.0832 tasks/second was found. However, this is a low task per time value overall and is a metric to improve upon in future iterations. Since time based efficiency is just one of the many areas of improvement in the design, three main issues will be the focus of discussion in the next section.

$$\text{Time Based Efficiency} = \frac{\sum_{j=1}^R \sum_{i=1}^N \frac{n_{ij}}{t_{ij}}}{NR}$$

N = The total number of tasks

R = The total number of users

n_{ij} = The results of task 'i' by user 'j'. If the user has completed the task, then $N_{ij} = 1$, if not, then $N_{ij} = 0$

t_{ij} = The time spent by user 'j' to complete task 'i'. If the task is not completed, time is measured until the user quits the task.

Figure 4. Formula used to calculate time based efficiency [6][7]

3.1 Usability problems

Using the collected participant data, three usability problems (denoted UP) were identified and ranked by severity in Table 3.

1. UP1 was identified by the comments made by participants. All five users addressed the changes made to the general organization of the Create/Schedule page, where the content of the two pages were combined and organized into a dropdown menu. However, the general consensus of the participants was that the headings for the dropdown were not indicative of the section's content. All test users were able to figure it out by the end of the walkthrough without guidance, commenting that they would know where to look once they familiarize themselves with the layout indicating low persistence and impact.
2. UP2 located in the formatting toolbar had legibility issues, as the labels added were not large enough. $\frac{2}{5}$ users were able to complete the task without further explanation resulting in a medium impact rank, as many commented that they did not know the name or function of each icon.
3. UP3 addressed how the calendar did not behave how users expected it to. $\frac{3}{5}$ users were unable to complete this task without intervention indicating high impact. Instructors commented that they do not think about assignments as a duration of time, but were able to connect it to the systems used in flight and hotel bookings.

Table 3: Usability Problem Table

ID	Task	Persistence [low, med, high]	Impact [low, med, high]	Frequency [common, rare]	Severity
UP1	Navigating the grouped sections	Low	Low	Common	Low
UP2	Identifying tools for formatting	Med	Med	Common	Med
UP3	Range Calendar	Low	High	Common	High

3.2 Recommendations

From the evaluation of the project done in this phase, both shortcomings and successes were discovered that will help to inform future iterations of the project. From the users' responses, it was found that transitioning from creating to scheduling assignments was straightforward with both on one page (see Appendix C). However, the organization of the dropdowns was confusing to some participants, so changes to the groupings of features and utilizing more intuitive headings should be considered in the future. A second recommendation is to make the formatting

tools more accessible [2]. The labels helped the participants find the tools, but were hard to read and did not explain the function or use of each tool (see Appendix C). In the future, the addition of descriptions or examples of the tools functions would be beneficial and further utilize the knowledge in the world principle [2]. Finally, the biggest issue users faced was using the range calendar to set deadlines. One participant explained how instructors do not usually think of assignments as a duration as they just release them at the beginning, then collect and grade them at the end, demonstrating that this design did not match target users' mental models. Thus, it is valuable to create a design that keeps these successes and shortcomings in mind and provides better support to their mental models for future iterations of the project [2].

4.0 Conclusion

In the *Evaluate* phase, the main objective was the assessment of the usability prototype produced in the *Create* phase. To accurately assess whether the prototype met the five usability criteria, several methods of data collection and analysis were implemented. Quantitatively, to measure errors, learnability, and efficiency, each task was timed, and the number of clicks and errors recorded. Qualitatively, a post-observational questionnaire containing the System Usability Scale (SUS) and five open-ended questions was provided to measure satisfaction and learnability. A memorability test, which assessed the user's ability to recall the location and function of an input, was also implemented. Using the collected data, the team was able to identify three usability problems of different severity levels. Specifically, issues were found in the area surrounding dropdown headings, formatting toolbar label legibility and functions, and expectations regarding the range calendar. Despite these issues, the prototype was generally well-received, with overall positive sentiment and an above average SUS score of 74.5. The next steps aim to continue the iterative process starting with further understanding successes and shortcomings of this design, and brainstorming ways to address them. Thus, starting the next iteration of the project and restarting the Design cycle to approach an optimal solution.

6.0 References

- [1] “Kritik App,” Kritik, Available: <https://ca.kritik.io/>. (Accessed Feb. 23, 2024).
- [2] J. D. Lee, C. D. Wickens, Y. Liu, L. N. Boyle, *Designing for People: An Introduction to Human Factors Engineering*, 3rd ed., CreateSpace, 2017. [Online]. Available: https://www.researchgate.net/publication/319402797_Designing_for_People_An_introduction_to_human_factors_engineering. (Accessed Feb. 23, 2024).
- [3] Y. Ma, “NLP: How does NLTK.Vader Calculate Sentiment?,” *Medium*, Feb. 05, 2020. https://www.google.com/url?q=https://medium.com/@mystery0116/nlp-how-does-nltk-vader-calculate-sentiment-6c32d0f5046b&sa=D&source=docs&ust=1712869794496090&usg=AOvVaw2xdKsddySec03N-qM2_-TK (accessed Apr. 11, 2024).
- [4] “Redirecting,” *Google.com*, 2024. https://www.google.com/url?q=https://www.westga.edu/academics/research/vrc/assets/docs/confidence_intervals_notes.pdf&sa=D&source=docs&ust=1712869794497476&usg=AOvVaw27pjAkTxjIx17BWZnI6SIC (accessed Apr. 11, 2024).
- [5] M. Soegaard, “System Usability Scale for Data-Driven UX,” *The Interaction Design Foundation*, Nov. 21, 2023. <https://www.google.com/url?q=https://www.interaction-design.org/literature/article/system-usability-scale&sa=D&source=docs&ust=1712869794501031&usg=AOvVaw0EwlKp6kM6UVdM3vILfhGA> (accessed Apr. 11, 2024).
- [6] “Usability Testing Metrics,” *Guides - Learn user testing & UX research | UXtweak*. <https://www.google.com/url?q=https://www.uxtweak.com/usability-testing/metrics/&sa=D&source=docs&ust=1712869794498053&usg=AOvVaw2sJmTzrEsOB2WPoODt6g8-> (accessed Apr. 11, 2024).
- [7] “7 Essential Usability Metrics and How to Use Them,” *www.eleken.co*. https://www.google.com/url?q=https://www.eleken.co/blog-posts/usability-metrics&sa=D&source=docs&ust=1712869794466327&usg=AOvVaw3VTJSoSxnfE_uv_CmiGQvXh (accessed Apr. 11, 2024).

7.0 Appendix

Appendix A: Phase 1 Recap

This section contains visuals for the *Create* phase.

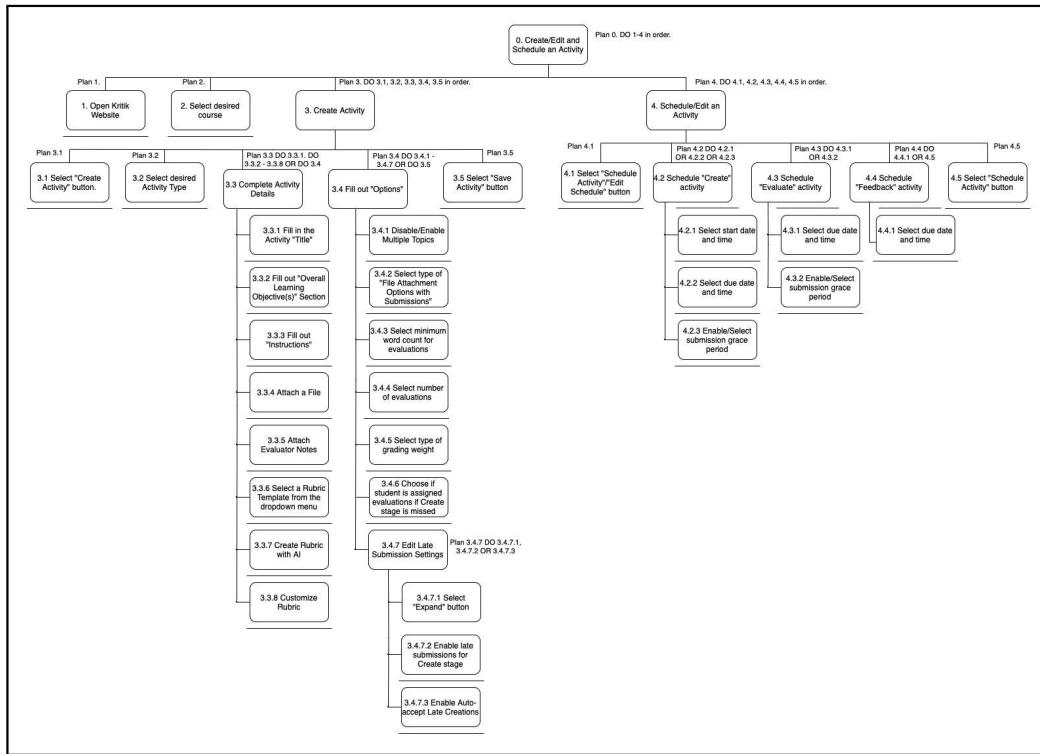


Figure A1. HTA in graphical form. An image with higher resolution can be found [here](#).

Appendix B: Final Prototype

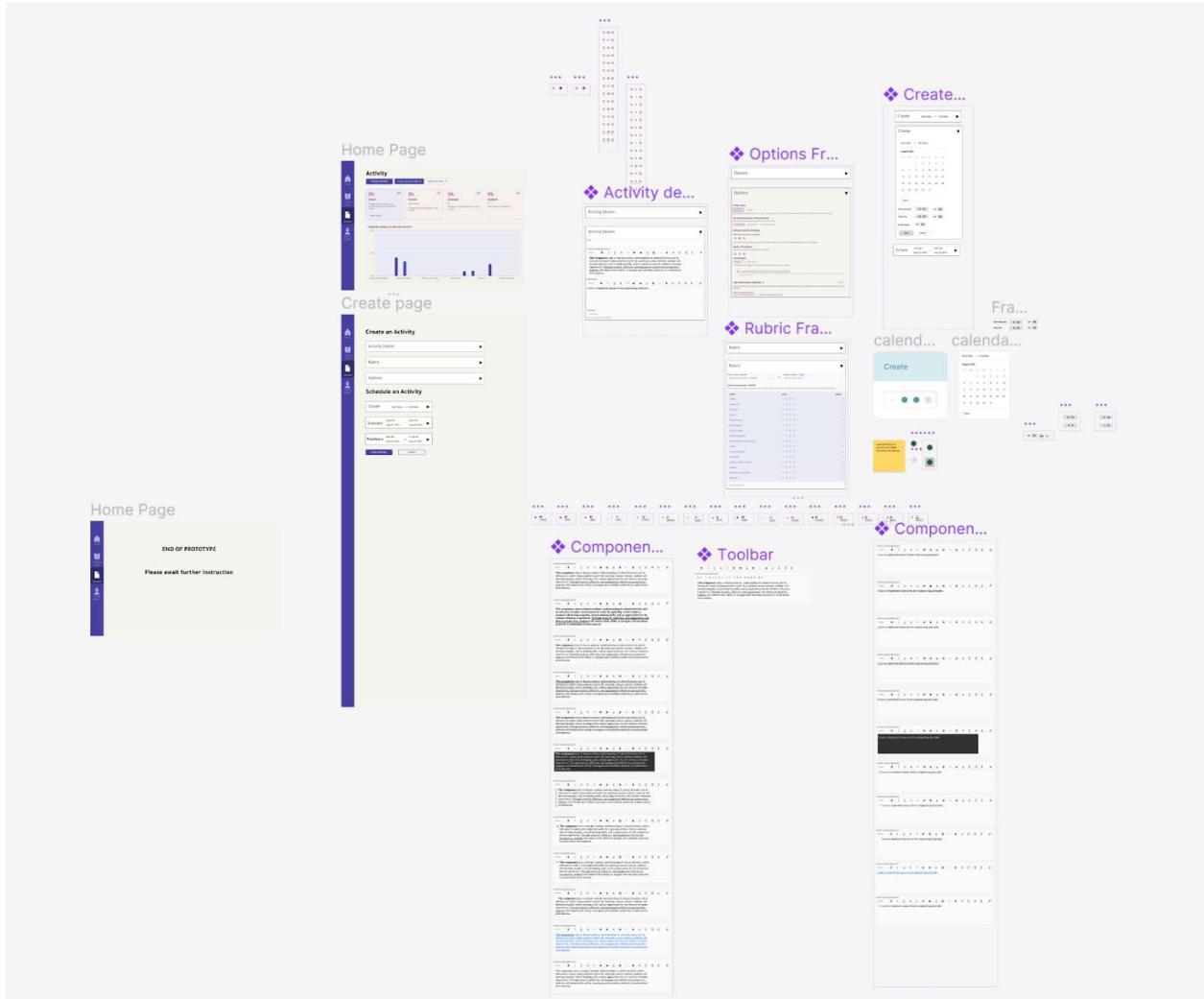


Figure B1. Figma Prototype Overview

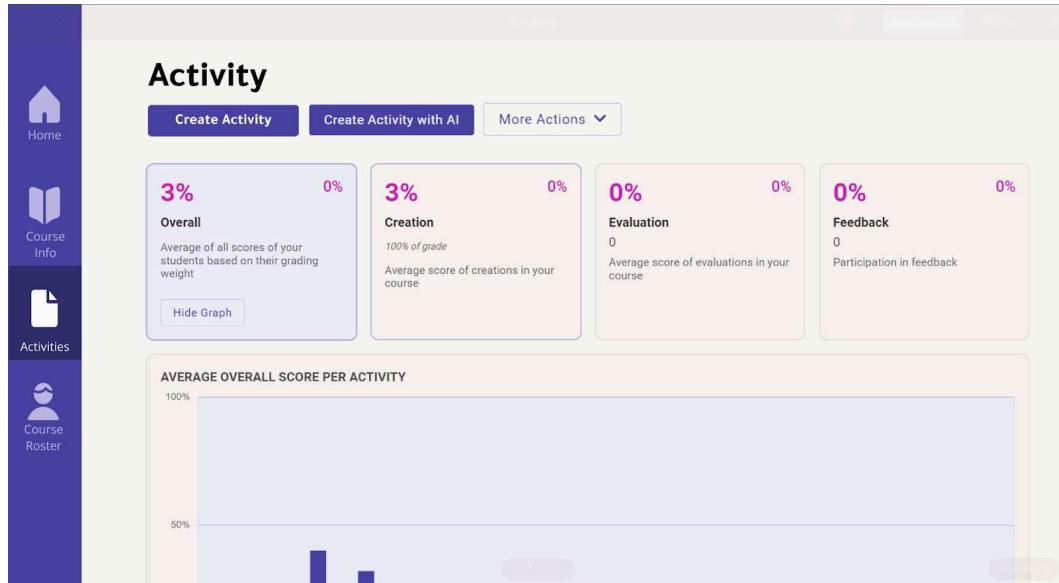


Figure B2. Figma Home Page

The screenshot shows the Figma Create Activity Page. On the left is a vertical sidebar with icons for Home, Course Info, Activities, and Course Roster. The main area is divided into sections:

Create an Activity

- Activity Details
- Rubric
- Options

Schedule an Activity

- Create:** Start Date: 9:45 PM → End Date: 9:45 PM
- Evaluate:** Aug 25, 2023 → Aug 28, 2023
- Feedback:** 9:45 PM Aug 28, 2023 → 11:00 PM Aug 29, 2023

At the bottom are "Save Activity" and "Cancel" buttons.

Figure B3. Figma Create Activity Page

Activity Details

Title*

Overall Learning Objective(s)

This assignment aims to deepen students' understanding of cultural diversity and its relevance in today's interconnected world. By exploring various cultures, students will develop empathy, critical thinking skills, and an appreciation for the richness of human experiences. Through research, reflection, and engagement with diverse perspectives, students will enhance their ability to navigate and contribute positively to multicultural environments.

Instructions

Learn to implement human factors engineering principles

Resources

Attach a File

Figure B4. Figma Formatting Toolbar

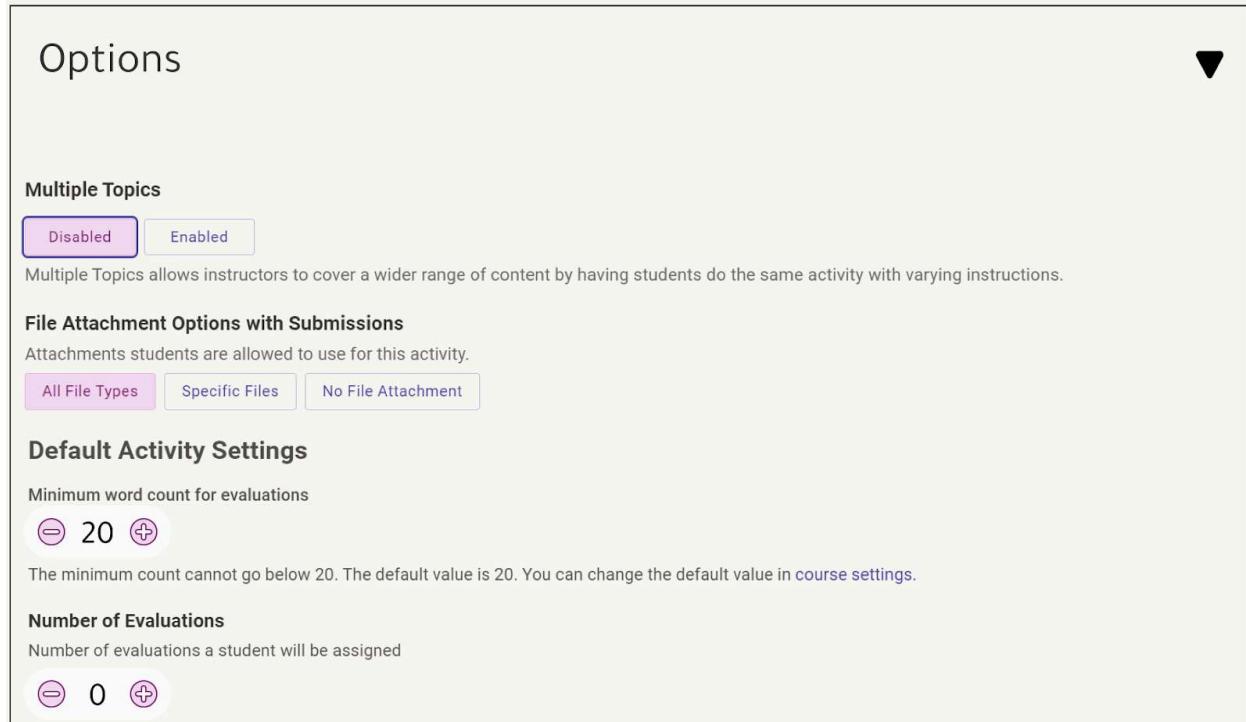


Figure B5. Figma Options Dropdown

Start Date → End Date

August 2024

Sun	Mon	Tue	Wed	Thu	Fri	Sat
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31		

Reset

Time Released 12 : 00 AM PM

Time Due 12 : 00 AM PM

Grace Period Off On

Save **Cancel**

Figure B6. Figma Scheduling Dropdown

Appendix C: Data Collection for Usability Testing

Please provide any comments you had about the process of creating an activity

5 responses

the directions about formatting etc were a bit confusing

I found that it was very easy to create an activity and set deadline and due dates. It was a bit confusing what the meaning of setting minimum word count was.

I am a Faculty of Music professor, and I didn't know the meaning of the words hyperlink, embed etc. Once they were explained to me, the rest of the prototype was very easy to use.

things are hidden behind "options" that maybe shouldn't have been?

It was intuitive in terms of finding the button to create an activity. There was no challenge in doing so.

Figure C1. Open Ended Questionnaire - Response to Activity Creation

Please provide any comments you had on scheduling a task

5 responses

this was straightforward

I found it tricky to set the end date, I didn't realize that you needed to set the end date at the same time as the start date and wasn't clear that I needed to reset this.

Once I got started, it was very intuitive to move through the tasks.

I like the calendar, but the due date doesn't behave the way I expected

It was generally intuitive in doing so. Could have probably added a prompt on which date corresponded to create the task and which was the due date, maybe right above each selected date, but otherwise it was easy to guess which was which.

Figure C2. Open Ended Questionnaire - Response to Activity Scheduling

Please provide comments on the process of switching between creating and scheduling a task.

5 responses

this was fine

I like that it was easy to navigate between the two tasks (ie I could set creation settings as well as set the scheduling times in the same interface without changing pages)

Switching between creating and scheduling the task and back again was very smooth.

It's nice it's on the same page

I would have preferred not an expanding menu because I wasn't sure which options were located in which menu, but once I figured it out then it was easy to find within the individual menu.

Figure C3. Open Ended Questionnaire - Response to design flow

Please provide any additional comments about your overall experience.

5 responses

written directions for the sequence would have been helpful

Overall, I liked the format of the interface and found it very simple and easy to utilize

Team 8 did a great job with this system!

I like the flexibility of entering information by keyboard, vs. clicking buttons

It was generally intuitive in finding where things are. I didn't have much difficulty in finding where things were even when they were not in the most intuitive location.

Figure C4. Open Ended Questionnaire - Response to Overall experience

How did you feel about the user interface as a new user?

5 responses

fairly clear in terms of layout.

I found it quite easy to use, and not challenging

I don't usually like technological things, but this process was intuitive and clear.

Other than the "options" drop down, it was pretty intuitive.

I like it better than the existing platform because it feels like I know where things are more easily.

Figure C4. Open Ended Questionnaire - Response to Overall experience

Table C1. System Usability Scale (SUS) - Response to Overall experience

Question	Participant				
	A	B	C	D	E
I think that I would like to use this website frequently.	4	5	5	3	4
I found this website unnecessarily complex.	4	2	1	3	2
I thought this website was easy to use.	4	4	4	4	4
I think that I would need assistance to be able to use this website.	4	3	5	3	1
I found the various functions in this website were well integrated.	3	5	5	4	4
I thought there was too much inconsistency in this website	3	2	1	1	1
I would imagine that most people would learn to use this website very quickly.	4	5	5	5	4
I found this website very cumbersome/awkward to use.	4	1	1	2	2
I felt very confident using this website.	3	4	5	5	4
I needed to learn a lot of things before I could get going with this website.	2	3	3	2	1

Participant: 1
 -Consent Form

Task #	Time(s)	Errors Made	Number of Clicks	Comments
1	7.76	0	1	
2	95.8	2	7	Unsure of what they were being asked to do. hard time finding icons (embed) hyperlink was found really fast, formatting cleared quickly was stuck on the objectives box page for a little bit
3	38.9	0	10	went to options right away, used +/- buttons
4	52.5	0	7	
5	47.13	0	8	Used +/- buttons
6	4.2	0	1	
Total	4:06.29 (min:sec)	2	34	

debrief: Do you have feedback on the task you just completed?

Written instructions rather than spoken

What components on the screen lead you to think that way? How could we make it more clear?

Text is too small

Figure C5. Participant 1 Quantitative Data

Participant: 2

Consent Form

Task #	Time(s)	Errors Made	Number of Clicks	Comments
1	18.0	0	1	
2	110.3	3	4	Likes toolbar menu labels, scrolled around and then was confused clicked activity details hovered for a while to find the right icon
3	98.1	1	9	Intuitive labels
4	60.6	6	8	Assuming automatic save, due date not intuitive, confusion with the end date clicked a lot of the calendar
5	18.5	0	5	
6	3.6	0	1	
Total	5:15.3 (min:sec)	10	28	

debrief: Do you have feedback on the task you just completed?

straightforward , didn't take too long to find what to

What components on the screen lead you to think that way? How could we make it more clear?

Figure C6. Participant 2 Quantitative Data

Participant: 3

-Consent Form

Task #	Time(s)	Errors Made	Number of Clicks	Comments
1	10.02	0	1	
2	282.05	9	15	Had trouble with finding the right section to use and finding/understanding the icons (more scrolling around the site rather than clicking)
3	54.89	1	10	opened rubric by accident when trying to find the right section
4	33.14	0	5	Used the manual option
5	20.96	0	5	
6	6.02	0	1	
Total	6:47.10 (min:sec)	10		

debrief: Do you have feedback on the task you just completed?

What components on the screen lead you to think that way? How could we make it more clear?

Figure C7. Participant 3 Quantitative Data

Participant: 4

-Consent Form

Task #	Time(s)	Errors Made	Number of Clicks	Comments
1	17.8	0	1	scrolled a little around page, no misclicks just exploration
2	124.3	0	3	
3	52.9	0	2	
4	85.5	1	5	scrolling issue with time , pressed save no issues with the dates
5	28.7	0	3	
6	6.41	0	1	
Total	5:26.0 (min:sec)	1	15	

debrief: Do you have feedback on the task you just completed?

What components on the screen lead you to think that way? How could we make it more clear?

Huh !? When default were in options it's weird wasn't sure which one was correct, not expected

due date of scheduling looks like booking hotels, scheduling assignments not thought of that way not connected like that, send out submission set due date, seems like students who designed it

Due dates not usually thought of as a durations for instructors

Figure C8. Participant 4 Quantitative Data

Participant: 5

Consent Form

Task #	Time(s)	Errors Made	Num of Clicks	Comments
1	5.3	0	1	
2	71.9	2	5	Confused about task, missed embed
3	17.1	0	9	Used +/- buttons
4	29.9	0	6	Used manual inputs
5	14.1	0	5	Used +/- buttons
6	1.63	0	1	
Total	2:33.0	2	27	

debrief: Do you have feedback on the task you just completed?

What components on the screen lead you to think that way? How could we make it more clear?

Figure C9. Participant 5 Quantitative Data

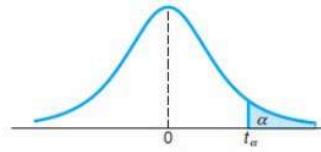
Appendix D: Data Analysis

This Section contains documents in regard to the data analysis.

<u>Paired sample</u>			
	Task 3	Task 5	difference (task5-task3)
user 1	38.9	47.13	8.23
user 2	98.1	18.5	-79.6
user 3	54.89	20.96	-33.93
user 4	52.9	28.7	-24.2
user 5	17.1	14.1	-3

$t_{stat} = \frac{\bar{d} - d_0}{s_d / \sqrt{n}}$	$n=5$	$H_0: \bar{d} \geq 0$
		$H_A: \bar{d} < 0$
$= \frac{-26.5 - 0}{30.458259 / \sqrt{5}}$	$t^* = 1.583$	$t_{stat} < -t_{v=4, \alpha=0.1}^*$
$= -1.9455$	\therefore at 0.1 significance level, reject the null hypothesis	

Figure D1. Paired sample time done to compare the times for task 3 and task 5

**Table A.4 Critical Values of the t -Distribution**

v	α						
	0.40	0.30	0.20	0.15	0.10	0.05	0.025
1	0.325	0.727	1.376	1.963	3.078	6.314	12.706
2	0.289	0.617	1.061	1.386	1.886	2.920	4.303
3	0.277	0.584	0.978	1.250	1.638	2.353	3.182
4	0.271	0.569	0.941	1.190	1.533	2.132	2.776
5	0.267	0.559	0.920	1.156	1.476	2.015	2.571
6	0.265	0.553	0.906	1.134	1.440	1.943	2.447
7	0.263	0.549	0.896	1.119	1.415	1.895	2.365
8	0.262	0.546	0.889	1.108	1.397	1.860	2.306
9	0.261	0.543	0.883	1.100	1.383	1.833	2.262
10	0.260	0.542	0.879	1.093	1.372	1.812	2.228
11	0.260	0.540	0.876	1.088	1.363	1.796	2.201
12	0.259	0.539	0.873	1.083	1.356	1.782	2.179
13	0.259	0.538	0.870	1.079	1.350	1.771	2.160
14	0.258	0.537	0.868	1.076	1.345	1.761	2.145
15	0.258	0.536	0.866	1.074	1.341	1.753	2.131
16	0.258	0.535	0.865	1.071	1.337	1.746	2.120
17	0.257	0.534	0.863	1.069	1.333	1.740	2.110
18	0.257	0.534	0.862	1.067	1.330	1.734	2.101
19	0.257	0.533	0.861	1.066	1.328	1.729	2.093
20	0.257	0.533	0.860	1.064	1.325	1.725	2.086
21	0.257	0.532	0.859	1.063	1.323	1.721	2.080
22	0.256	0.532	0.858	1.061	1.321	1.717	2.074
23	0.256	0.532	0.858	1.060	1.319	1.714	2.069
24	0.256	0.531	0.857	1.059	1.318	1.711	2.064
25	0.256	0.531	0.856	1.058	1.316	1.708	2.060
26	0.256	0.531	0.856	1.058	1.315	1.706	2.056
27	0.256	0.531	0.855	1.057	1.314	1.703	2.052
28	0.256	0.530	0.855	1.056	1.313	1.701	2.048
29	0.256	0.530	0.854	1.055	1.311	1.699	2.045
30	0.256	0.530	0.854	1.055	1.310	1.697	2.042
40	0.255	0.529	0.851	1.050	1.303	1.684	2.021
60	0.254	0.527	0.848	1.045	1.296	1.671	2.000
120	0.254	0.526	0.845	1.041	1.289	1.658	1.980
∞	0.253	0.524	0.842	1.036	1.282	1.645	1.960

Figure D2. Critical values for t-distribution

	A	B	C	D	E	F	G	H	I	J
1		user 1	user 2	user 3	user 4	user 5			# of participants	# of tasks
2	Task 1		7.76	18	10.02	17.8	5.3	t1j		5
3	Task 2		95.8	110.3	282.05	124.3	71.9	t2j		6
4	Task 3		38.9	98.1	54.89	52.9	17.1	t3j		
5	Task 4		52.5	60.6	33.14	85.5	29.9	t4j	Time based efficiency	
6	Task 5		47.13	18.5	20.96	28.7	14.1	t5j		0.08322291695
7	Task 6		4.2	3.6	6.02	6.41	1.63	t6j		
8										
9	1/tij		0.1288659794	0.05555555556	0.0998003992	0.05617977528	0.1886792453			
10			0.01043841336	0.009066183137	0.003545470661	0.008045052293	0.01390820584			
11			0.02570694087	0.01019367992	0.01821825469	0.01890359168	0.05847953216			
12			0.01904761905	0.01650165017	0.03017501509	0.01169590643	0.03344481605			
13			0.02121790791	0.05405405405	0.04770992366	0.03484320557	0.07092198582			
14			0.2380952381	0.2777777778	0.1661129568	0.1560062402	0.6134969325			
15	all nij = 1 since all tasks completed									

Figure D3. Time based efficiency excel calculations

Time	Participant	1	2	3	4	5	Mean Time	Overall Average
Task		1	2	3	4	5		
1	7.76	18	10.02	17.8	5.3	11.776		46.707333333
2	95.8	110.3	282.05	124.3	71.9	136.87		
3	38.9	98.1	54.89	52.9	17.1	52.378		
4	52.5	60.6	33.14	85.5	29.9	52.328		
5	47.1	18.6	20.96	28.7	14.1	25.892		
6						1		
Errors								
tasks/Participants	1	2	3	4	5	% participants that made error during task		
1	0	0	0	0	0	0		
2	2	3	9	0	2	0.8	errors/task	
3	0	1	1	0	0	0.4		5
4	0	6	0	1	0	0.4		
5	0	0	0	0	0	0		
6	0	0	0	0	0	0		
	2	10	10	1	2			

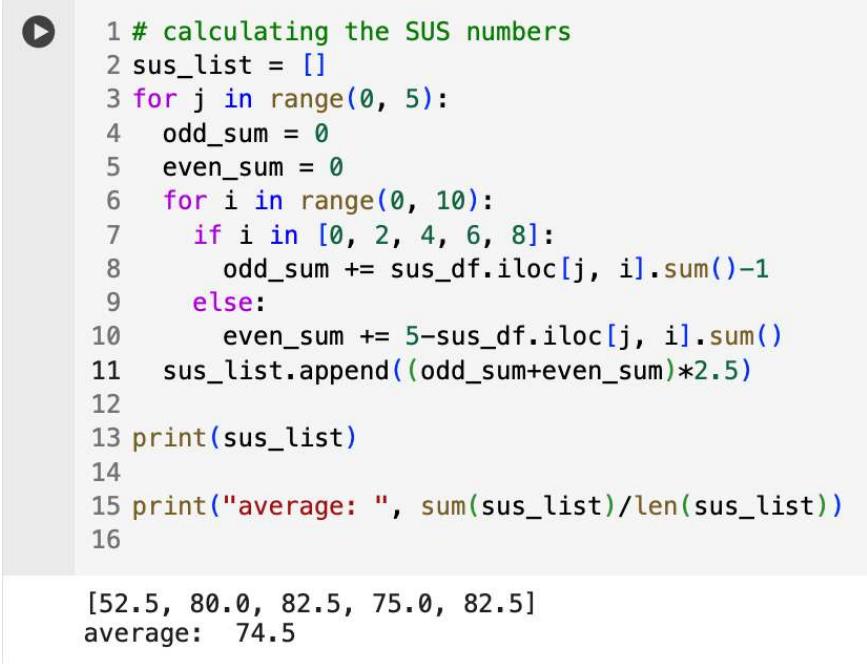
Figure D4. Time and error excel calculations

17	Num Clicks Participant						
18	Task	1	2	3	4	5	Median Clicks
19	1	1	1	1	1	1	1
20	2	7	4	15	3	5	5
21	3	10	9	10	2	9	9
22	4	7	8	5	5	6	6
23	5	8	5	5	3	5	5
24							
25							
26							
27	Task	Mean Time					
28	1	11.776					
29	2	136.87					
30	3	52.378					
31	4	52.328					
32	5	25.892					
33	6	1					

Figure D5. Number of clicks excel calculations and mean times

1	# this is the dataframe with only the numerical parts of the questionnaire aka									
2	# the parts specifically for SUS scale related questions									
3										
4	sus_df = q_df.select_dtypes(include=['number'])									
5	sus_df									
6	I think that I would like to use this website frequently. I found this website unnecessary complex.									
7	I thought this website was easy to use.									
8	I think that I would need assistance to be able to use this website.									
9	I found the various functions in this website were well integrated.									
10	I thought there was too much inconsistency in this website.									
11	I would imagine that most people would learn to use this website very quickly.									
12	I found this website very cumbersome/awkward to use.									
13	I felt very confident using this website.									
14	I needed to learn a lot of things before I could get going with this website.									
0	4	4	4	4	3	3	4	4	3	2
1	5	2	4	3	5	2	5	1	4	3
2	5	1	4	5	5	1	5	1	5	3
3	3	3	4	3	4	1	5	2	5	2
4	4	2	4	1	4	1	4	2	4	1

Figure D6. Table of Likert responses from the SUS part of the post-observation questionnaire.



```

1 # calculating the SUS numbers
2 sus_list = []
3 for j in range(0, 5):
4     odd_sum = 0
5     even_sum = 0
6     for i in range(0, 10):
7         if i in [0, 2, 4, 6, 8]:
8             odd_sum += sus_df.iloc[j, i].sum()-1
9         else:
10             even_sum += 5-sus_df.iloc[j, i].sum()
11     sus_list.append((odd_sum+even_sum)*2.5)
12
13 print(sus_list)
14
15 print("average: ", sum(sus_list)/len(sus_list))
16

```

[52.5, 80.0, 82.5, 75.0, 82.5]
average: 74.5

Figure D7. The Python code used to generate the SUS values and the overall average.

```

100%|██████████| 5/5 [00:00<00:00, 524.37it/s]
the directions about formatting etc were a bit confusing
compound: -0.2263, neg: 0.213, neu: 0.787, pos: 0.0,
I found that it was very easy to create an activity and set deadline and due dates. It was a bit confusing what the meaning of setting minimum word count was.
compound: 0.5657, neg: 0.057, neu: 0.777, pos: 0.166,
I am a Faculty of Music professor, and I didn't know the meaning of the words hyperlink, embed etc. Once they were explained to me, the rest of the prototype was very easy to use.
compound: 0.4927, neg: 0.0, neu: 0.907, pos: 0.093,
things are hidden behind "options" that maybe shouldn't have been?
compound: 0.0, neg: 0.0, neu: 1.0, pos: 0.0,
It was intuitive in terms of finding the button to create an activity. There was no challenge in doing so.
compound: 0.0516, neg: 0.097, neu: 0.752, pos: 0.15,

```

Figure D8. Vader sentiment analysis for the question “Please provide any comments you had about the process of creating an activity”.

```

100%|██████████| 5/5 [00:00<00:00, 353.77it/s]
this was straightforward
compound: 0.0, neg: 0.0, neu: 1.0, pos: 0.0,
I found it tricky to set the end date, I didn't realize that you needed to set the end date at the same time as the start date and wasn't clear that I needed to reset this.
compound: -0.4184, neg: 0.106, neu: 0.894, pos: 0.0,
Once I got started, it was very intuitive to move through the tasks.
compound: 0.0, neg: 0.0, neu: 1.0, pos: 0.0,
I like the calendar, but the due date doesn't behave the way I expected
compound: 0.1981, neg: 0.0, neu: 0.863, pos: 0.137,
It was generally intuitive in doing so. Could have probably added a prompt on which date corresponded to create the task and which was the due date, maybe right above each selected date, but otherwise it was easy to guess which was which.
compound: 0.6597, neg: 0.0, neu: 0.881, pos: 0.119,

```

Figure D9. Vader sentiment analysis for the question “Please provide any comments you had on scheduling a task”.

100%|██████████| 5/5 [00:00<00:00, 576.89it/s]
 this was fine
 compound: 0.2023, neg: 0.0, neu: 0.526, pos: 0.474,
 I like that it was easy to navigate between the two tasks (ie I could set creation settings as well as set the scheduling times in the same interface without changing pages)
 compound: 0.8225, neg: 0.0, neu: 0.73, pos: 0.27,
 Switching between creating and scheduling the task and back again was very smooth.
 compound: 0.296, neg: 0.0, neu: 0.845, pos: 0.155,
 It's nice it's on the same page
 compound: 0.4215, neg: 0.0, neu: 0.682, pos: 0.318,
 I would have preferred not an expanding menu because I wasn't sure which options were located in which menu, but once I figured it out then it was easy to find within the individual menu.
 compound: 0.3218, neg: 0.042, neu: 0.849, pos: 0.109,

Figure D10. Vader sentiment analysis for the question “Please provide comments on the process of switching between creating and scheduling a task”.

100%|██████████| 5/5 [00:00<00:00, 389.34it/s]
 written directions for the sequence would have been helpful
 compound: 0.4215, neg: 0.0, neu: 0.741, pos: 0.259,
 Overall, I liked the format of the interface and found it very simple and easy to utilize
 compound: 0.7152, neg: 0.0, neu: 0.701, pos: 0.299,
 Team 8 did a great job with this system!
 compound: 0.6588, neg: 0.0, neu: 0.577, pos: 0.423,
 I like the flexibility of entering information by keyboard, vs. clicking buttons
 compound: 0.5994, neg: 0.0, neu: 0.647, pos: 0.353,
 It was generally intuitive in finding where things are. I didn't have much difficulty in finding where things were even when they were not in the most intuitive location.
 compound: 0.2584, neg: 0.0, neu: 0.93, pos: 0.07,

Figure D11. Vader sentiment analysis for the question “Please provide any additional comments about your overall experience”.

100%|██████████| 5/5 [00:00<00:00, 1891.54it/s]
 fairly clear in terms of layout.
 compound: 0.3818, neg: 0.0, neu: 0.658, pos: 0.342,
 I found it quite easy to use, and not challenging
 compound: 0.4116, neg: 0.124, neu: 0.602, pos: 0.274,
 I don't usually like technological things, but this process was intuitive and clear.
 compound: 0.4301, neg: 0.104, neu: 0.669, pos: 0.227,
 Other than the "options" drop down, it was pretty intuitive.
 compound: 0.2732, neg: 0.158, neu: 0.602, pos: 0.241,
 I like it better than the existing platform because it feels like I know where things are more easily.
 compound: 0.8622, neg: 0.0, neu: 0.551, pos: 0.449,

Figure D12. Vader sentiment analysis for the question “How did you feel about the user interface as a new user?”.