**A5: Final Evaluation with Users**

**Group 10: Magdalen, Maya, Soyun, Leslie**

**Describe Study**

**Pilot Test and Improvements Made**

During our study, we conducted two pilot tests to help refine our methods, aiming for more reliable results and to rehearse for our evaluations with our study participants. This allowed us to test our research approach with a small number of participants before the main study, enabling us to refine our research questions based on their feedback. The decision to conduct two pilot sessions was made to validate the effectiveness of changes implemented after the first session (Schade, 2015).

**In our First Pilot Test, we Developed the Following User Goals and Instructions**

**User goal one:** Using a set of flashcards

**Instructions**: You have your first midterm coming up in a few days and want to memorize some key terminologies for SENG 310. What would you do?

**User goal two**: Creating a study schedule

**Instructions**: Create a plan for SENG 310 midterm 1 using our study planner.

After the first test, we asked the users the following questions: Firstly, what is your preferred study method? (e.g., flashcards, mind maps, reviewing notes). Secondly, did you find the interface easy to navigate? Were there any areas where you felt stuck or confused? And lastly, as a Brightspace user, did you find the interface familiar? From this pilot, we learned that the preferred study method is to review notes. Participants found the interface easy to navigate for the most part, they appreciated having a sidebar with courses to choose from; however, they found the "start study session" button to be redundant in multiple places. Most participants found the interface familiar as Brightspace users. Suggestions for improvement included making the calendar's purpose more obvious and explaining the differences between the "start study session" and "Generate study plan" features.

**In our Second Pilot Test, we Learned The Following**

After remedying some issues on our interfaces, we conducted a second pilot study, conveying the instructions mentioned above to the user. We asked them the same questions and learned the following. This user also preferred reviewing notes. They found that the buttons made more sense, with less redundancy after changes were made. The Brightspace interface is similar and easy to use. Additionally, they found the interface easier to navigate. They mentioned making the calendar's purpose more obvious and explaining the differences between the "start study session" and "Generate study plan” section in our interface like the participant in pilot one.

**Improvements Made**

Having completed the two pilot sessions, we learned a lot about some of the usability issues with our interface. We improved the interface by addressing several key areas. Specifically, we fixed the calendar to make its purpose clearer. For instance, users had difficulty understanding the significance of different colours for classes, more specifically, whether they represented class due dates, exams, etc. Additionally, we edited the buttons to reduce redundancy based on user feedback.

**To address these suggestions, we decided to make the following changes:**

1. Fix the button for creating a study session.
2. Modify the calendar to be more readable, clarifying the purposes of different colors and what they indicate.
3. Make it more obvious to users when they already have a study plan generated.

**Who Were the Participants and How Many Participated**

For our study, we recruited five students from the University of Victoria, ensuring that each team member recruited at least one student. These participants ranged from second-year to 4th year undergraduates students. It was essential to involve University of Victoria students as our participants because our tool is integrated into the university's Brightspace system, and UVic students are our target users.

**Methods Used and Why**

We segmented our study methodology into quantitative and qualitative analyses, explained below.

**Quantitative Analysis**

For the quantitative analysis, we utilized two quantitative metrics collected for each task: the time taken by the user to complete the task and the number of clicks they made. These results were then converted into line graphs to facilitate comparison among different participants, we also utilized these quantitative metrics to calculate descriptive statistics, such as the mean number of clicks for each task.

Quantitative analysis was selected due to its capacity for objectively measuring human performance, enabling the comparison of different interaction phenomena. By quantitatively assessing the number of clicks made by users, we can detect potential confusion within the interface. For instance, if Participant One clicked approximately 50 times and took 5 minutes to complete a task, while the average number of clicks by other participants on the same task was 19, it suggests there may be a design issue warranting further investigation.

**Qualitative Analysis**

For the qualitative analysis, we employed a thematic analysis method alongside semi-structured user interviews. These interviews were used to gain a deeper understanding of the users' thoughts about the interface, its drawbacks, and potential improvements they believed could be made.

Thematic analysis was used to gain a deeper understanding of the qualitative data derived from the interviews. This method involved breaking down and organizing qualitative data by tagging individual observations and quotations with appropriate codes, facilitating the discovery of themes. The thematic analysis allowed us us to generate meaningful findings and interpretations about our design from our qualitative data (Rosala, 2022). Additionally, it helped us organize our data. For example, under the theme of Interface Navigation, we identified that a user found the feature "Generate/continue study plan" confusing, expressing difficulty in determining which menu they were in.

**Data Sources Collected and how we Analyzed the Data**

Our data was sourced from our research and interactions with users, especially from our pilot studies, which informed the development of interview questions for the final user evaluation. We analyzed the data by visualizing it using graphs (Graphs 1 through 4) and in a thematic analysis map (Table 1).

To collect our research data, we utilized the following tables, where we collected our quantitative data at the top and qualitative data at the bottom. Each table represents a different participant. In addition, task one is using a set of flashcards, task two is figuring and creating a new study schedule.

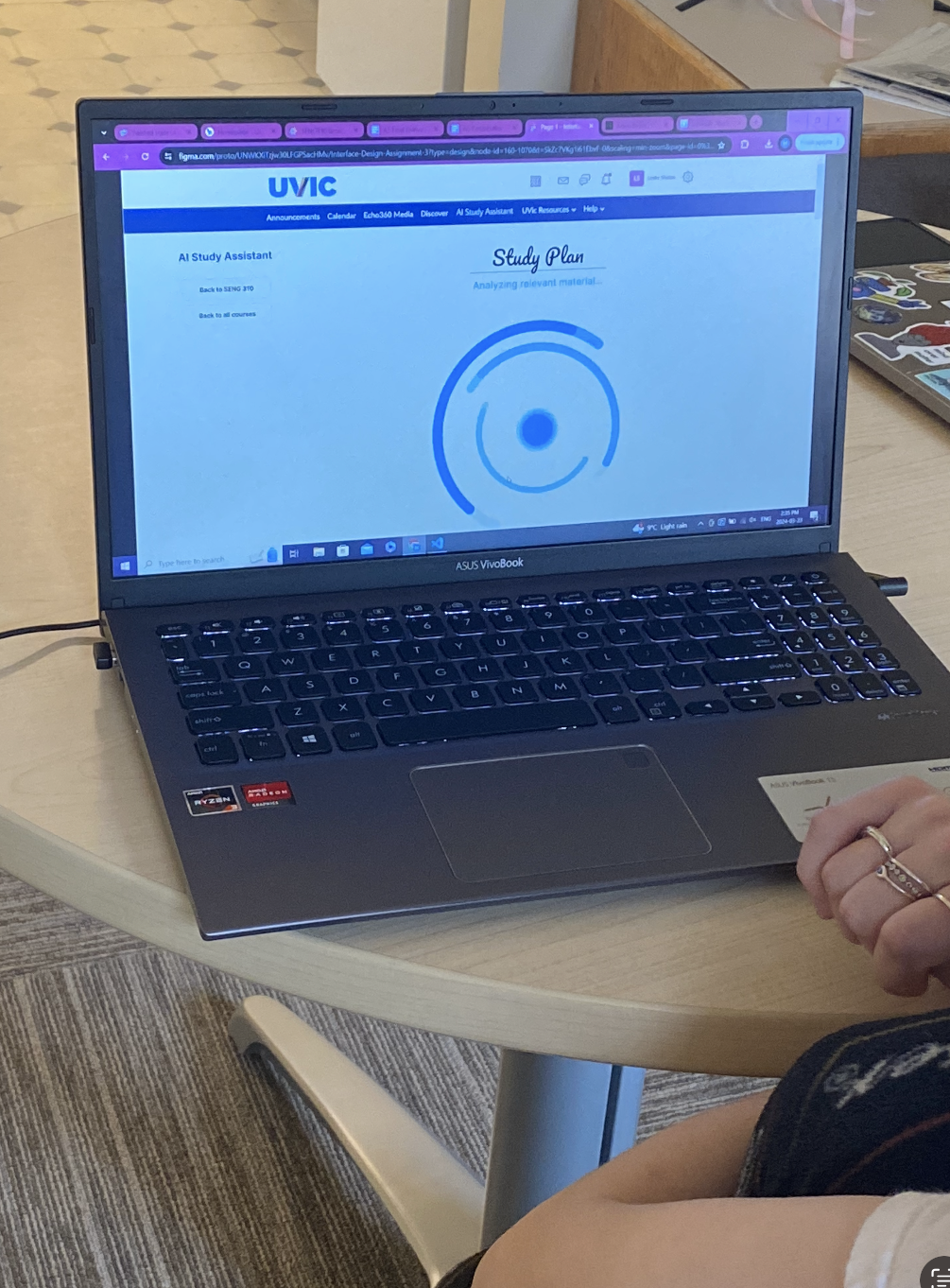
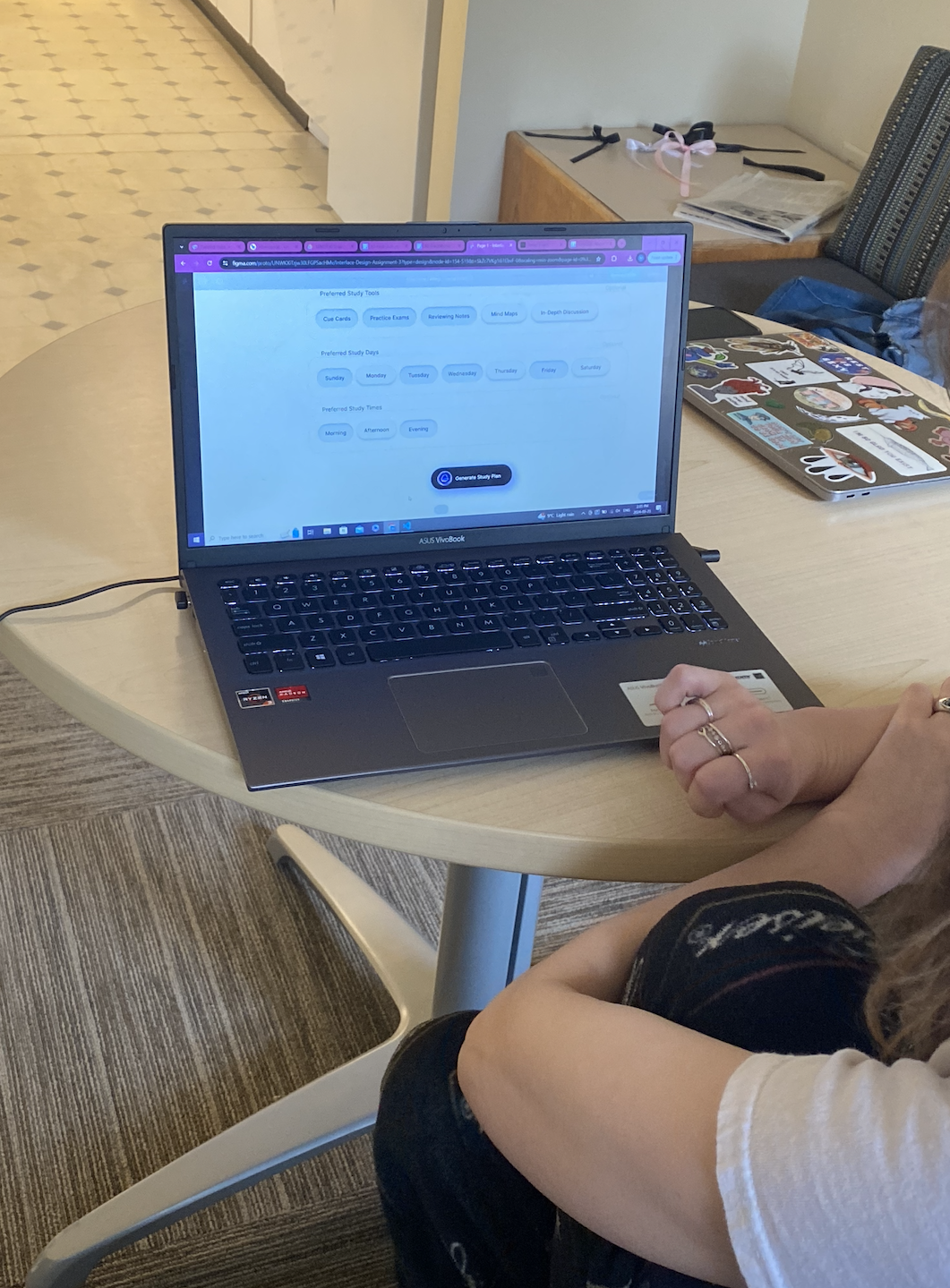
| **Participant #1** | **Time recorded** | **Number of clicks** |
| --- | --- | --- |
| 1st task | 2 min 12 seconds | 17 |
| 2nd task | 1 min 8 seconds | 13 |
| Interview Questions:   * What is your preferred study method? Ie flashcards, mindmaps, reviewing notes   + Reviewing notes * Did you find the interface easy to navigate? Were there any areas where you felt stuck or confused?   + How to navigate from page to page was confusing for her.   + Make steps to create a study plan more obvious * As a brightspace user, did you find the interface familiar?   + It is familiar, but says the university may need to teach students about the new feature even if they are familiar with BrightSpace. | | |

| **Participant #2** | **Time Recorded** | **Number of Clicks** |
| --- | --- | --- |
| 1st task | 1 min 21 seconds | 13 |
| 2nd task | 1 min 15 seconds | 18 |
| Interview Questions:   * What is your preferred study method? Ie flashcards, mindmaps, reviewing notes   + Reviewing notes, writing down key points and summarizing   + Memorizing key points   + Varies if its a short answer test or multiple choice * Did you find the interface easy to navigate? Were there any areas where you felt stuck or confused?   + Pretty easy to navigate   + Not stuck or confused in a few areas, but took a minute to get their bearings when first looking at it * As a brightspace user, did you find the interface familiar?   + Yes, looked a lot like brightspace and liked the familiarity | | |

| **Participant #3** | **Time Recorded** | **Number of Clicks** |
| --- | --- | --- |
| 1st task | 3min 15 seconds | 11 |
| 2nd task | 1 min 10 seconds | 14 |
| Interview Questions:   * What is your preferred study method? Ie flashcards, mindmaps, reviewing notes   + Flashcards, Rewrite notes, Youtube videos, Practice exams, textbook/HW questions * Did you find the interface easy to navigate? Were there any areas where you felt stuck or confused?   + Generate/continue study plan confusing. What is it doing? Regenerating study plan confusing. Hard to know which menu I was in.   + Generally quite clean, not too many places to go, not overwhelmed * As a brightspace user, did you find the interface familiar?   + Yes | | |

| **Participant #4** | **Time Recorded** | **Number of Clicks** |
| --- | --- | --- |
| 1st task | 1 min 15 seconds | 14 |
| 2nd task | 1 min 39 seconds | 25 |
| Interview Questions:   * What is your preferred study method? Ie flashcards, mindmaps, reviewing notes   + Reviewing notes   + Focusing on definitions + how they’re used   + Trying to make my own definition by explaining to someone else * Did you find the interface easy to navigate? Were there any areas where you felt stuck or confused?   + Easy to navigate but sources are limited   + Found the concept of ‘AI generated planner’ a bit confusing as it doesn’t know what my personal schedule outside of the school   + Calendar needs more improvement- confused what shades and borders mean   + Larger font for cue cards * As a brightspace user, did you find the interface familiar?   + yes | | |

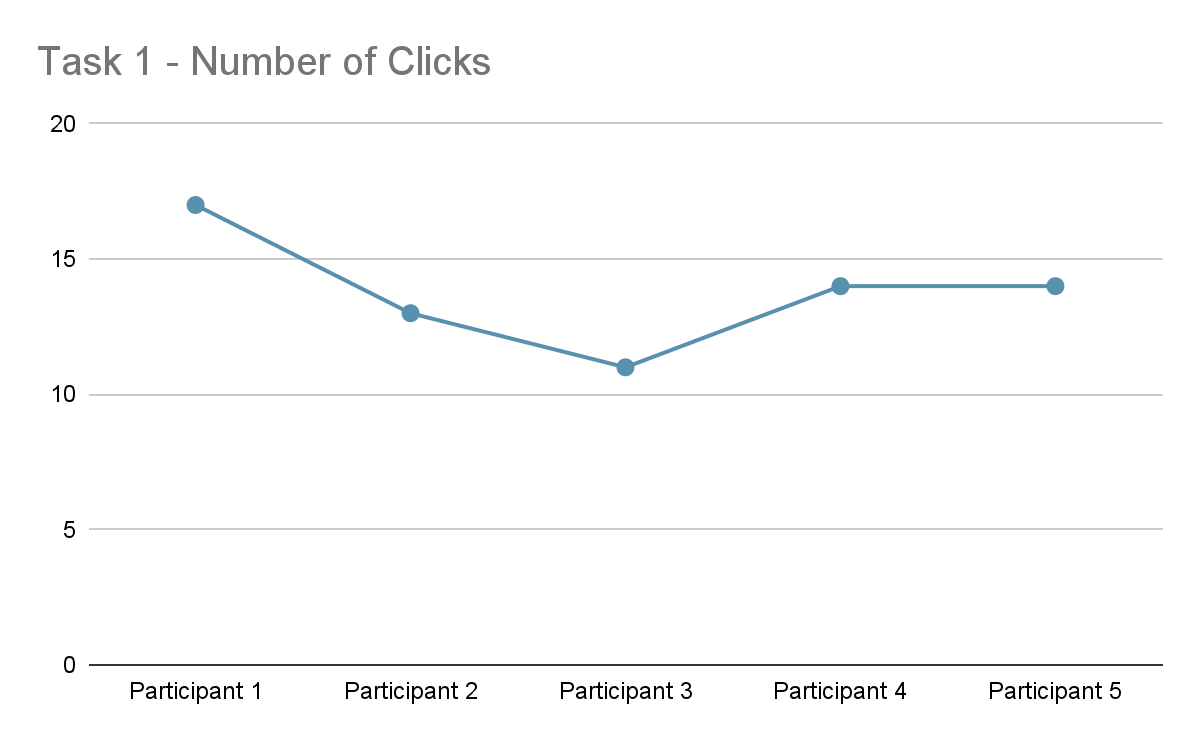
| **Participant #5** | **Time Recorded** | **Number of Clicks** |
| --- | --- | --- |
| 1st task | 2 min 20 seconds | 14 |
| 2nd task | 1 min 20 seconds | 17 |
| Interview Questions:   * What is your preferred study method? Ie flashcards, mindmaps, reviewing notes   + Reviewing notes   + Practice tests * Did you find the interface easy to navigate? Were there any areas where you felt stuck or confused?   + Found most of the part straightforward to navigate although it took a few clicks to find the function I was looking for. Some words were ambiguous. * As a brightspace user, did you find the interface familiar?   + Yes, has similar layouts | | |



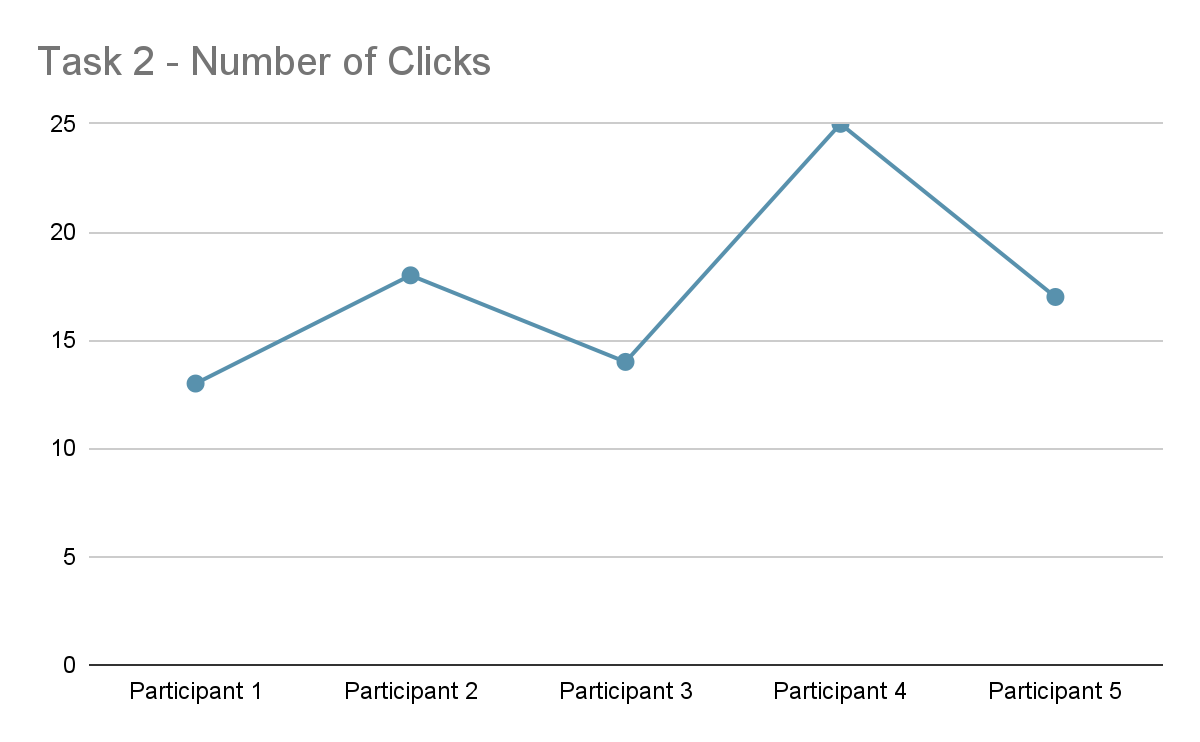
**Figure 1.** Image of a participant interacting with the system.

**Results**

After conducting the final study with five recruited participants, two quantitative values were collected for each task, and the results were converted into line graphs to facilitate a comparison between different participants.

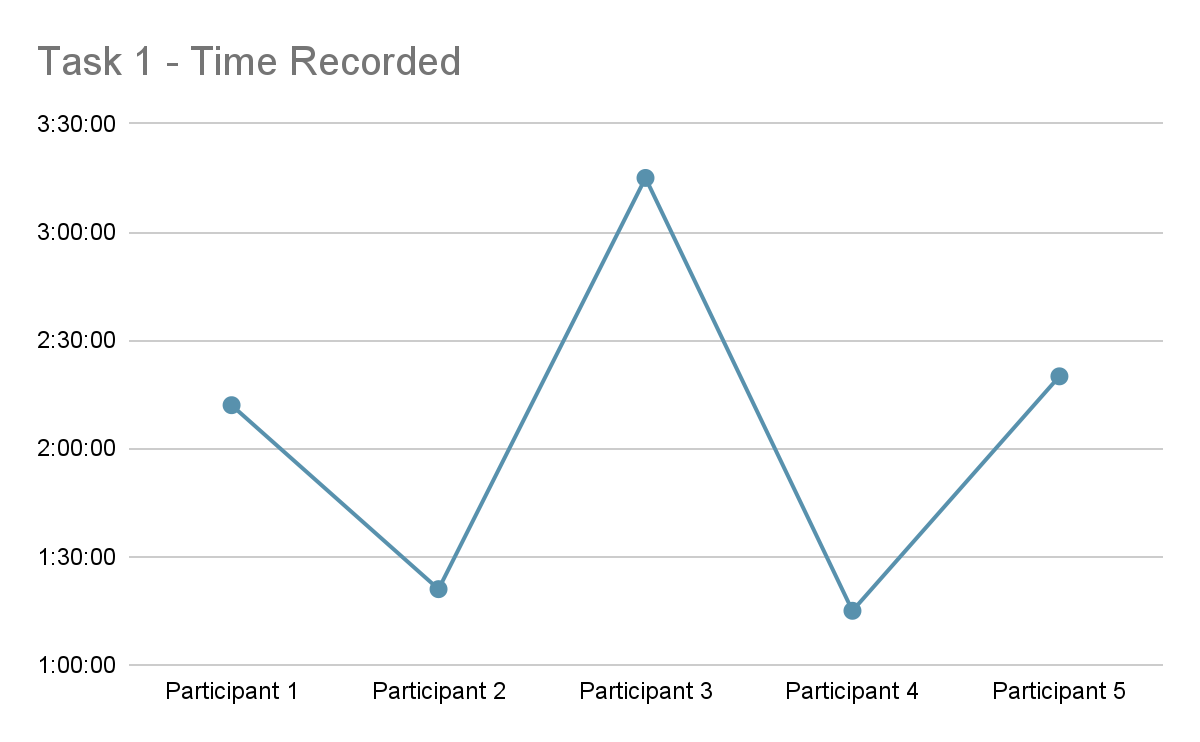


**Graph 1.** Line graph depicting the number of clicks collected for Task 1 across five participants.



**Graph 2.** Line graph depicting the number of clicks collected for Task 2 across five participants.

Graph 1 illustrates the number of clicks collected for the first task, which involved the goal 'Using a set of flashcards'. The individual values collected were as follows: [17, 13, 11, 14, 14] (unit in clicks). The average mean number of clicks was calculated to be 13.8 clicks. Considering that the task can typically be completed with around 10 clicks, we concluded that the participants were able to perform the task efficiently. Graph 2 displays the number of clicks for the second task, focusing on 'Figuring out the next study schedule'. The individual values were [13, 18, 14, 25, 17] (unit in clicks). Given that the task was expected to be completed within 14 clicks, we identified participant 4’s value as an outlier and therefore excluded it from the average calculation. The average number of clicks was found to be 15.5, which met the learnability goal.



**Graph 3.** Line graph depicting the time taken by each participant to complete Task 1 across five participants.



**Graph 4.** Line graph depicting the time taken by each participant to complete Task 2 across five participants.

Graphs 3 and 4 both depict the time taken by each participant to complete tasks 1 and 2, respectively. The individual recorded times for task 1 were [2:15, 1:21, 3:15, 1:15, 2:20], and for task 2 were [1:08, 1:15, 1:10, 1:39, 1:20] (unit in min:sec). Upon conversion into graphs 3 and 4, it became apparent that graph 3 had a value indicating a longer duration than the others, suggesting participant 3 as an outlier. After removing this outlier, the average time taken for the remaining four participants to complete task 1 was found to be 1 minute and 47 seconds. Regarding Graph 4, though it appeared to have an outlier value, this was due to the y-axis scaling. As the difference between the participant who took the longest and shortest time to complete the task was only 31 seconds, we included all five values in the average calculation, resulting in 1 minute and 18 seconds. We attributed the differences among participants to some experimenters allowing participants to navigate through the interface before the task, enabling them to familiarize themselves.

Following the completion of the tasks, each participant was asked three questions to allow for a more in-depth analysis and reflection on our interface. The collected answers were categorized into three main themes: Preferred Study Method, Interface Navigation, and Study Plan. Although all participants noted that they found the interface familiar to BrightSpace, some provided feedback on the navigation or study plan functions, suggesting they were not as straightforward as expected. The collected data were analyzed using thematic coding methods.

| Topic | Theme | Code |
| --- | --- | --- |
| AI Study Planner | Preferred Study Method | - Reviewing notes  - Memorizing key points  - Using flash cards  - Doing practice exams |
| Interface Navigation | - “Navigating from page to page was confusing”  - “Took a minute to get the bearings when first looking at it”  - “Generate/continue study plan was confusing and it was hard to know which menu I was in”  - “It was easy to navigate but I felt like the options for the users are very limited” |
| Study Plan | - Make steps to create a study plan more obvious  - “The concept of AI generating everything for me is confusing when it doesn’t even know what my schedule outside the school is like” |

**Table 1.** Summary of thematic analysis derived from post-interview questions and qualitative data.

**Discussion**

In our analysis of the usability study for the AI Study Assistant prototype, we've pinpointed key areas for design improvement to enhance user experience. The data revealed that tasks such as using flashcards and scheduling study plans took more time and clicks than expected, highlighting a need for greater efficiency. Despite the interface's familiarity, thanks to its resemblance to the UVic Brightspace platform, users faced difficulties with navigation and the clarity of specific features like the study planner. This feedback underscores the importance of improving "Efficiency of Use" and "Visibility of System Status" to make the interface more intuitive and user-friendly.

To address these concerns, we plan to refine the interface with clearer labels and instructions, especially around the study planner's functionality. This includes rewording and repositioning buttons to remove ambiguity around their purpose, thereby enhancing the user's control and understanding. Additionally, we aim to augment the study plan calendar's design to display all relevant schedules in one view, integrating class, lab, and all created study plan schedules into the calendar. This will also involve improving the interface to clearly indicate customizable study times, ensuring users can easily adjust their schedules to fit personal preferences.

By focusing on these key improvements, informed by our usability study, we aim to create a more efficient, user-centric design for our AI Study Assistant, ultimately supporting students in their academic endeavors more effectively.

**Study Limitations and Reflections**

Our study yielded some great results for us, and gave us lots of useful feedback to continue developing and improving our interface design. However, our study did have a few limitations, which ultimately affected the quality of results we were able to obtain from our study.

One of the major constraints that affected our results was the amount of time we had to work on both our prototype and develop our study itself. It was a bit challenging to have our prototype itself developed to a point where we could successfully run a study with it, as our vision involved numerous features working within it, which meant that a large time investment was needed in order to get everything ready for usability tests. However, despite the time constraints and the large amount of work that was necessary to get our prototype working, we were still able to run successful tests with the limited features we were able to implement and learn about how to improve our interface as a whole.

We also found that our sample size was a bit limiting, but given the time constraints, there was not much that could be done to remedy this issue. From the few participants we were able to recruit and run a usability test with, we were able to gain valuable feedback and information from their experiences, especially as we were able to recruit participants from a variety of different faculties and fields of study. After our first usability tests back in the second assignments, we figured that having participants from a variety of fields was very useful, as assessments are often quite varied between faculties. Thus, we learned that keeping our sample size as diverse as possible was key to ensuring that our interface is useful for as many students as possible.

Despite the limitations, our study was still largely successful, and has given us the information we needed to continue developing our prototype, and recognize problem areas that we as designers would not have been able to see ourselves.

**References**

Rosala, M. (2022, August 17). How to Analyze Qualitative Data from UX Research: Thematic Analysis. Retrieved from Nielsen Norman Group website: <https://www.nngroup.com/articles/thematic-analysis/>

Schade, A. (2015, April 5). Pilot Testing: Getting It Right (Before) the First Time. Retrieved from Nielsen Norman Group website: <https://www.nngroup.com/articles/pilot-testing/>