Assignment M4:
Prototype Evaluation Planning

Elizabeth Boyett
eboyett3@gatech.edu

Abstract — The application Microsoft Teams (now referred to as Teams) is a business communication platform. Teams offers a variety of services, including chat features, video meetings, polling, task tracking, file sharing, storage, and a calendar. The M assignments will focus on solely sending a message through the chat feature, in either the desktop or mobile applications. One design life cycle will be implemented to improve the task.

1 QUALITATIVE EVALUTION

The textual prototype outlined in Assignment M3 will be evaluated qualitatively. The textual prototype was an improved version of the current Teams interface, which can be reviewed in ***Figure 1***. Certain features and improvements that were defined in the needfinding were integrated into the current interface.

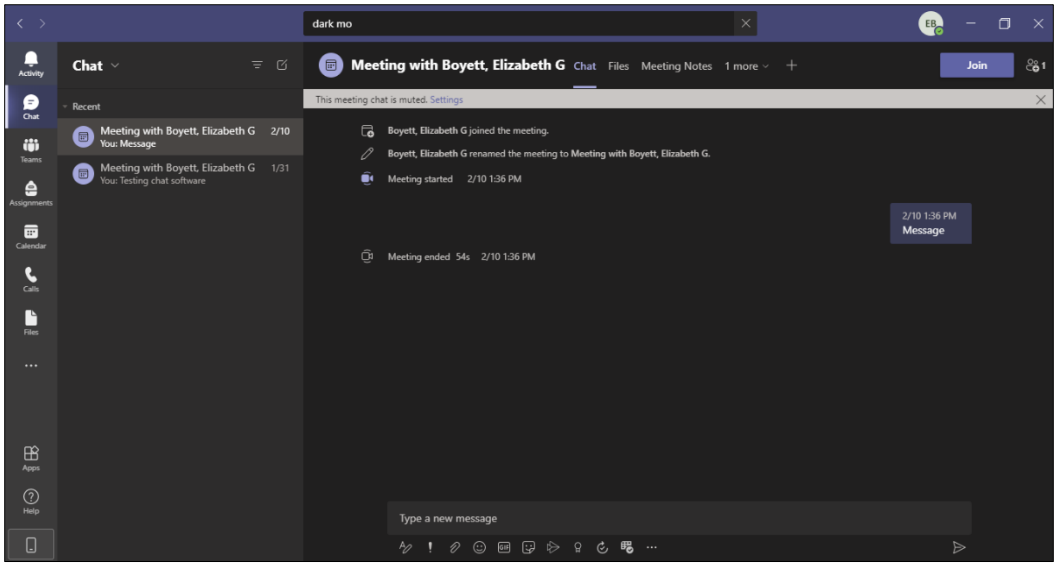

Figure 1 — Current Teams messaging interface in dark mode on the desktop application.

1.1 Evaluation plan

The interface will be evaluated with a survey. A survey was chosen because it works best for textual prototypes due to the asynchronous delivery and individual feedback that is received. The survey is planned to be distributed to business professionals who utilize Teams. The survey will be shared with Georgia Tech classmates, family, and friends through class forums and group chats. The goal is to have at least 25 respondents. The participants will take the survey where they see fit. It will be somewhere where they can access it, such as at home or while outside on their mobile devices. The survey results will be recorded by the PeerSurvey software and will be exported to Excel to be reviewed and analyzed.

The survey will summarize the textual prototype with the proposed changes, including the new features, reminders of shortcuts, and security in place to prevent the accidental messaging of personal passwords. The survey will then ask some open-ended questions All survey questions will not be required to be answered to allow the user to feel as if they are not forced to answer anything. The survey questions will also not be written with leading or loaded questions to avoid potential bias. Demographic information will also be gathered in the survey. The open-ended questions asked in the survey are:

1. What do you like about these features?
2. What do you dislike about these features?
3. The interface will remind you when a shortcut is available when you execute the long way with an automatically disappearing popup. For example, if the user right-clicks and selects paste, it will remind you that the user can use Ctrl + v instead. What do you think of this feature?
4. The interface will not allow messages to be sent accidentally with the user's Microsoft password while signing in. Has this ever happened to you and what do you think of this change?
5. What are your thoughts and goals when using the Teams message function?
6. How difficult do you think it would be to use this new interface?

For the full extent of the survey content, layout, and questions, refer to *Appendix: Survey Outline*.

1.2 Requirements

The survey evaluation will help assess if the requirements defined in M2 were met because of the feedback that is received from the user base. The requirements were built according to the needs of the user base, so their feedback concerning the prototype is essential.

The **compliance** of the interface will be evaluated because users will be asked if they are satisfied with the proposed solution to the current security problem, or if they were even aware or have run into the security problem of accidentally sending their password over chat. The new **functionality** will be evaluated by the survey respondents, concerning if they like or dislike the proposed prototype. **Learnability** will also be assessed, as the users will be asked about their perceived difficulty in using this interface.

The evaluation of the prototype will also address some parts of the data inventory. Who the users **are** will be measured from the demographic information gathered. Their evaluation of the execution of the **main task** of sending a message, and various **subtasks** such as replying to a specific message and sending emojis, will also be evaluated in the features question.

2 EMPIRICAL EVALUATION:

The second prototype, the speech-to-text card prototype, will be evaluated through empirical means. The prototype from M3 can be reviewed in ***Figure 2***.

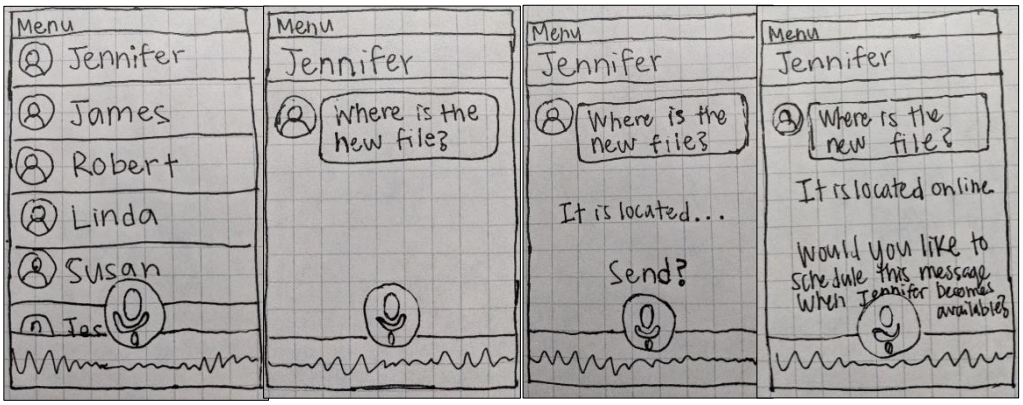

Figure 2 — Left: The main page of the interface Left Middle: The page when the user selects a conversation or requests the assistant to read messages. Right Middle: The page when the user starts to relay a message as a reply. Right: The page when the user replies to a user who is currently offline.

The prototype will be improved to be completely functional for the empirical evaluation by including more cards for all possible scenarios.

2.1 Experimental conditions

The empirical evaluation will compare two treatments, the current Teams interface, and the new proposed prototype. The **target problem** of the most efficient way to send a message will be measured, by measuring the number of seconds it takes to send a message in each interface. Other measurements, such as the number of clicks, are not measured because a speech-to-text interface will always require minimal amounts of clicks comparatively. Multiple categories of sending a message will be evaluated. They are:

- **S**end a message
- Send a message with a file attached
- Send a message marked urgent
- Send a message as a reply
- Send a message containing emojis
- Write a message and schedule it

The **experiment conditions** will include a within-subjects design where one group will experience both treatments. All subjects will test all categories in both treatments. The subjects will be recruited from different neutral internet ads, and a target of 500 subjects will be recruited for the experiment so it can produce statistically significant data. A third party will instruct the subjects, and a software program will be used to measure the speed to execute in the interface, to reduce social bias. Random assignment will occur by using a random number generator to assign each test subject to their first interface, which will help **control** for bias. Other **controls** will be that the test subjects will be instructed what to write in their message exactly. The messages will be the same across users and interfaces so bias will be controlled. This is to avoid the results getting skewed because some subjects write longer, more complex messages when typing compared to when verbally drafting them, or vice versa.

2.2 Hypothesis

The **null hypothesis** of the empirical evaluation is the speed to send a message is equal in both interfaces. The **alternative hypothesis** is the speed to send a message is not equal in both interfaces.

2.3 Experimental method

As mentioned earlier, a **within-subjects** design will be utilized in the experiment. Subjects will be **assigned randomly** the order in which they utilize the interface. **Subjects will complete** executing the tasks in the categories above, and the data of time to execute will be measured in seconds. They will **generate categorical ratio data** for both interfaces.

Analysis of the data will occur after the experiment is conducted. The data has two treatments and is ratio data. The data of the time to execute will be averaged for each interface across all the test subjects. The standard deviation will also be calculated for each of these. The sample size will also be measured by the number of subjects who participated in the evaluation, so around 500. A **student's t-test** will be used to analyze the data, to find if there is statistical significance in the difference in the execution time between the two treatments. It is assumed the data distribution is normal to utilize the student's t-test analysis.

2.4 Lurking variables

Some lurking variables may confound the data. The lurking variable concerning the treatment order was minimized using random assignment to the order in which each interface is tested. The expertise of the subject may also confound the data, they may be expert users in the Teams interface, and novice users in the prototype because they have never used it before. The experiment still wants to recruit and test Teams users because that is the target demographic. So, a demo will be given on both interfaces before the experiment to equalize the knowledge as much as possible, but the lurking variable will still be present.

Another lurking variable may be the experience with speech-to-text interfaces in general. The subjects recruited will be US citizens. US citizens are much more likely to own Apple phones according to the Global Stat Counter, with 61% of US citizens owning an Apple phone compared to 27% globally (2021). Apple has a virtual assistant named Siri. Though other phones may have virtual assistants, like Cortana or Google Assistant, Siri has been on the market longer and is utilized more. Thus, the subjects tested may be more familiar and may test better with a text-to-speech interface, than when compared to a global user base. This would not be representative of Teams actual users, because it is used globally.

3 PREDICTIVE EVALUATION

The simplified paper prototype introduced in M3 will have a predictive evaluation performed. Refer to ***Figure 3*** for the paper prototype. A GOMS model will be created to analyze using the prototype in different situations and/or contexts.

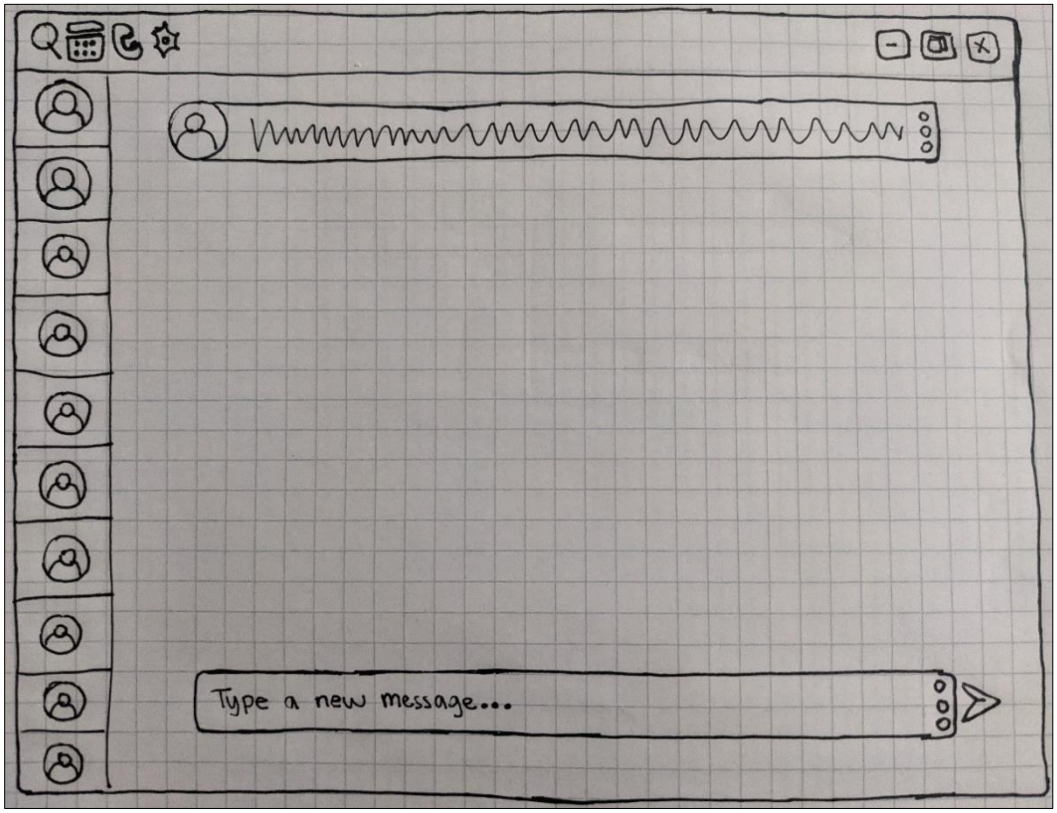

Figure 3 — Paper prototype of the simplified interface.

3.1 Task evaluation

The **specific task** that will be evaluated will be sending a message. This is a known task by the user, and the efficiency of executing this task will be assessed with the GOMS model. The **user's goal** will be to communicate specific information with a colleague. This information may be informal, or formal. It may include very complex information such as a file containing relevant information. The method of communication may also vary, with a message being marked as important, urgent, or as a reply.

3.2 Operators

The **operators** available to the user are:

- Open Teams
- Select draft box
- Draft a message
- Attach a file with a shortcut
- Add an emoji
- Add a gif
- Choose time to schedule message with shortcut
- Make message urgent with shortcut
- Bold, italicize or underline certain parts of a message with formatting shortcut options
- Pin a message
- Search previous messages
- Select a previous message
- Delete a previous message
- Reply to a previous message
- Send the message

The time to execute these operators may vary depending on the context and will be evaluated once the GOMS model is constructed.

3.3 Evaluation

The prototype will be evaluated by a single goal that the user knows how to do in advance, which is communicating with a colleague via a message. The efficiency of an expert user to perform that known task will be assessed.

4 PREPARING TO EXECUTE

The qualitative and predictive evaluations will be executed in the next assignment. The textual prototype will be evaluated through qualitative means by distribution and analysis of a survey. The paper prototype will be evaluated predictively through the creation of a GOMS model.

These two were chosen because the card prototype was not ready for empirical evaluation. To follow all the steps and experimental conditions, the interface needed to be more developed and working. The prototype and current interface also required software to track the users executing sending a message. Tracking software was required to get the most accurate measurements in the empirical

evaluation. This was not available at the time and would be too difficult to develop over a two-week period. It is also harder to recruit and execute an evaluation in person due to the ongoing coronavirus pandemic. I can recruit participants virtually for the qualitative evaluation, and I can individually create a GOMs model of the paper prototype.

5 REFERENCES

1. Mobile operating system market share worldwide. (2021, February). Retrieved March 14, 2021, from https://gs.statcounter.com/os-market-share/mobile/worldwide

6 APPENDIX: SURVEY OUTLINE

A prototype of the Teams interface is created. The current Teams interface can be referred to in ***Figure 1***.

The prototype has additional features/changes integrated into the interface. They are:

- Under the previous message three-dot menu
    - *Pin message*, to pin the message for easy access
    - *Download*, with a prompt to choose the download location
    - *Reply to Message*, will copy the message to current message as reply format
- Under the current emoji menu
    - *Add custom emoji*, Option to add custom emojis and gifs
- New clock icon near format icon
    - *Schedule a message*, allows the user to schedule a message for a future time
- Under the specific private chat options menu
    - *Delete*, to delete a message

1. What do you like about these features?
2. What do you dislike about these features?
3. The interface will remind you when a shortcut is available when you execute the long way with an automatically disappearing popup. For example, if the user right-clicks and selects paste, it will remind you that the user can use Ctrl + v instead. What do you think of this feature?
4. The interface will not allow messages to be sent accidentally with the user's Microsoft password while signing in. Has this ever happened to you and what do you think of this change?
5. What are your thoughts and goals when using the Teams message function?

6. How difficult do you think it would be to use this new interface?
7. What is your age?
8. What is your gender?
9. Are you employed?
10. What is your expertise with Teams?