Evaluating Chatbot User Experience

A Major Qualifying Project

Submitted to the faculty of

WORCESTER POLYTECHNIC INSTITUTE

In partial fulfillment of requirements for the degree of Bachelor of Science

By:

Alexandria Nogueras

Jonathan Coelho

Megan Seely

Sadie Dominguez

Date:

March 23, 2021

WPI Faculty Advisor:

Soussan Djamasbi

Acknowledgements

This Major Qualifying Project (MQP) is a continuation of a previous IQP, “Empathetic Conversational Agent Assisting with Title IX Cases and Digital Counseling”, by Brandon Persons (Persons, 2020) under Professor Soussan Djamasbi, with assistance from Ph.D. candidate Prateek Jain.

We would like to thank our advisors for their guidance throughout our project.

Abstract

This MQP was conducted to test the usability of ERIN, an empathetic chatbot, which was developed for undergraduate students to provide them with better access to mental health resources. We began by extending prior research that established the need for the ERIN chatbot on WPI's campus via interviews with key informants (Persons, 2020). We examined the need for ERIN by collecting information from WPI students via a survey. The analysis of data gathered from our survey study verified the need for ERIN on the WPI campus. Insights into ERIN's usability and user experience were gathered through a second study via Zoom interviews with 41 participants. Interviews revealed that ERIN performed better than the control website used in the study. No significant difference was found between the experience of using ERIN on smartphone or laptop devices. Based on the results, this MQP also provides recommendations for short-term and long-term projects for future iterations of the ERIN chatbot.

Table of Contents

1 Introduction

Studies suggest that mental wellness is a major issue on college campuses. For example, Mowbray et al. (2006) argue that "students may be more likely to have current or past trauma or family problems to deal with. Given all these trends, it should be no surprise that the number of college students with mental illness diagnoses has grown (p.226)." The analysis of a recent survey from the American College Health Association (ACHA) shows that students suffered from severe anxiety (about 60%) and depression (40%) (2019). These findings highlight a need for improved access to mental wellness resources on college campuses.

In an effort to improve access to mental wellness resources on college campuses, Empathetic Research IoT Network (ERIN) was developed via an Interdisciplinary Qualifying Project (Persons, 2020). The goal of the ERIN chatbot project was to create a user-friendly platform to help WPI students obtain resources about mental wellness and Title IX issues (Title ix, n.d). We are continuing this project by conducting user studies for Erin chatbot. We had three research objectives that we aimed to answer for our Major Qualifying Project:

1. Extend the verification for the need for ERIN, which was established by key informant interviews (Parsons, 2020), by gathering information from WPI undergraduate students.
2. Examine whether it would be beneficial to include Erin Chatbot in the Student Development and Counseling Center (SDCC) website.
3. Examine whether the experience of using ERIN is similar/different on laptops and smartphones.

In the following sections we provide a brief review of literature relevant to our project. We then explain how we conducted two studies to address our research questions and explain what we learned from the results of these studies.

2 Background

In this section, we briefly discuss the literature to provide the necessary background for our project.

2.1 Human Chatbot Interaction

The use of chatbots have become popular in businesses because they can reduce customer service cost and they can handle responding to multiple customers/users at the same time (Ranoliya, 2017). Many companies, such as Bank of America, Disney, and Dominos, have implemented chatbots to mitigate the need for users to search through lengthy FAQ sections (Budiu, 2018). Chatbots were implemented with the goal of creating a more natural flow of conversation rather than having the user interact with a web page that is neutral and expressionless (Park et al., 2018). The basic premise of human chatbot interaction is to enable a conversation between the chatbot and users through the use of conversational interfaces. The interactions between humans and chatbots allow the users to feel as if they are having a casual conversation which is more comfortable and natural for many people (Park et al., 2018).

2.1.1 Empathetic Chatbots

Empathetic chatbots are those that can engage in an empathetic dialogue (Morris et al., 2018. According to the Computers are Social Actors theory (CASA) (Nass et al., 1994) and Liu, B., & Sundar, S. S.'s research, users find a chatbot that shows sympathy, cognitive empathy, and affective empathy to be more supportive than a chatbot that only gives advice (2018). Furthermore, a study discovered that teenagers would prefer to speak with a chatbot about sensitive topics because they believe chatbots are good at listening, good at keeping secrets, give reliable advice, and are non-judgmental (Liu & Sundar, 2018).

The performance of an empathy chatbot depends on the implementation of Natural Language Processing (NLP). Chatbots use NLP to extract relevant information from the text that is entered by a user (Singh, 2017). Within NLP literature there are two major topics, intent and entities. The intent, broadly defined, refers to the intentions of the user. Intents can be defined within two categories. The first is casual intents, these are opening and closing statements as well as affirmative and negative statements. The second is business intents, these are the mappings within the chatbot that allows the chatbot to find the answer that the user is searching for (Singh, 2017). Entities are connected to business intents. Entities are used to find the desired answer that the user seeks (Singh, 2017). An example of an entity would be a user seeking the release date of a movie. The title of the movie would be what the user intends to use to find the release date for and the entity would be the release year. Entities and intent enable chatbots to be dynamic and to continually learn from users (Singh, 2017).

2.1.2 Chatbots and Mental Wellness

Chatbots are becoming more common in the area of mental wellness. The demand for wellness resources may be greater than the supply and it can be overwhelming to try to find these resources (Cameron et al., 2018). Additionally, chatbots can be a resource for those that are reluctant to get help due to stigmatization (Vaidyam et al., 2019). Since chatbots cannot form their own opinions or thoughts, people feel like they can talk to a chatbot without being judged (Lucas et al., 2014). For these reasons, multiple researchers have started to look into chatbots as a resource for mental wellness (Cameron et al., 2018; Vaidyam et al., 2019). While each team of researchers came up with different chatbots, all of the chatbots had the common theme of using empathy. That is, these chatbots rely heavily on the ability to be empathic towards the end user (Cameron et al., 2018; Vaidyam et al., 2019).

2.1.4 Chatbots and Mediums

Chatbots are available on a variety of devices and applications to help the user access information conveniently via mediums such as websites, messaging channels, mobile applications, and in-app widgets (Srikanth, 2020). Chatbots are most notably accessed via devices such as smartphones, laptops, and tablets. Moreover, in a study consisting of 16 first-time chatbot users interacting with eight chatbots over several sessions, the authors found that "six participants predominantly used the chatbots on their phone, one on her tablet, and nine on their laptop" (Patel, 2018). Additionally, previous research found that older users preferred using their laptop over a smartphone or tablet medium (Bröhl, et al.). Bröhl, et al., believe that this influence could be from smartphones and tablets being developed later than laptops or that those smaller screens are a potential issue for older participants (2018). However, there is also evidence that smartphones and laptops are considered to be equally convenient for performing certain tasks (Wilson, E. V., & Djamasbi, S., 2019).

2.2 ERIN

ERIN, which stands for Empathetic Research IoT Network, was designed in the User Experience Decision Making (UXDM) lab at Worcester Polytechnic Institute (WPI). The goal of ERIN was to create a virtual assistant for college students that could provide support with a multitude of problems (e.g., "don't touch your face" for COVID crisis). ERIN chatbot, a specific feature of ERIN, was designed to provide an additional channel of access to mental wellness resources for undergraduate students (Persons, 2020). ERIN chatbot, which this report focuses on, was originally designed with an emotion detection feature to enhance the bot's user experience. The emotion detection feature, which is still being developed, was not included in our MQP project.

2.2.1 Title IX

Title IX refers to a law, passed in 1972, that made it illegal to discriminate against someone based on their sex (What is Title IX?, n.d). Title IX applies to institutions that receive federal funding from the Department of Education. On college campuses, Title IX is dedicated to preventing sexual misconduct, protecting students from discrimination based on sexual orientation, gender identity, or transgender status (Title ix, n.d).

2.2.2 Current Solution

Currently, all information that WPI offers regarding mental health or Title IX can be found on WPI's SDCC website across many pages. Within these pages, there are resources available on each topic. To find a specific resource, the user can read through the pages or use the navigation bar to search for it. The external links to resources on the website need to be consistently updated.

As determined by the previous IQP, the structure of the SDCC website can be improved to provide better access to resources that is provided by the center. When dealing with mental health or Title IX issues, it is important to minimize cognitive effort, due to the user's heightened emotions (Persons, 2020). Paying attention to colors is also important here because certain colors, such as red, can potentially trigger or heighten emotions (Persons, 2020). To minimize cognitive effort and the threat of heighten emotions, mental resources should be provided in an easy to access manner via a welcoming and calming environment (Persons, 2020).

2.3 User Experience Research

User Experience research is essential to User-Centered Design (Djamasbi and Strong, 2019). The feedback and insight that is gained from UX research helps to develop personalized product experience for target users. There are a variety of UX research methods that can be

implemented to successfully gather data. Some of the more well-known options include usability-experiment studies, moderated remote usability studies, A/B testing, focus groups, eye tracking, interviews, and surveys (Rohrer, 2014). UX research helps developers to make informed design decisions (Ghosh 2018).

In this project we conducted two separate UX studies to address our project's research questions. The first study was a campus-wide online survey, and the second study was a remote moderated user experience study including recorded interviews and surveys.

3 Study 1: Need Verification

The Need for Erin was originally established via key informant interviews (Persons 2020). To verify the need for ERIN from a different perspective, we designed a survey study to solicit feedback from WPI students. The first question in this survey study asked participants about resources available to WPI students. The next set of questions related to resources specific to the SDCC. Finally, participants were asked if they would like an extra medium of communication with the SDCC, and if they did, what form would they want the medium to be. At the end of the needs survey, participants were asked if they would like to help with the MQP in the future. If participants were willing to help, they were asked to provide their email so that we could contact them regarding our future user experience interviews. The questions asked in this survey can be found in Appendix A.

Participants for this study were recruited through WPI organizations, social media, and team member's personal networks. A total of 73 WPI students participated in our survey study.

The results of this study showed that 79% of participants knew at least some of the resources that was offered by SDCC. Most students (93%) wanted an extra medium of communication with the SDCC. Of those students, 58% of them preferred a chatbot as the extra form of communication. These results revealed that a chatbot, such as ERIN, is likely to be well received by students.

4 Study 2: Usability and User Experience

The goal of this study was to evaluate the user experience of the ERIN chatbot and determine improvements to be made in the future. To achieve this goal, we constructed our user research based on the previous study (Persons, 2020) to better understand ERIN's audience and to provide insight for improving user experience and usability of ERIN chatbot.

4.1 Study Design and Process

We designed our study protocol taking into account the need for qualitative and quantitative data. A moderated remote usability study was the only feasible option due to the COVID-19 pandemic. Moderating the study allowed our team to ask clarifying questions and to implement the think-out-loud method. Guided interviews and post-task surveys allowed us to gather both qualitative and quantitative data to discover positive and negative aspects of ERIN's user experience.

As in prior research (Persons, 2020) we used a set of predefined scenarios that participants were required to read before interacting with ERIN either via a laptop or smart phone. The experiment scenarios and device (laptop vs. smartphone) were assigned to each participant using a Latin Square design. This design enabled each scenario to be used the same amount of times in varying order. Due to the complexity of the study design having the SDCC website as a control, and the chatbot being tested on both smartphone devices and laptops, the Latin Square design helped to mitigate two sources of variability that could occur. The Latin Square method was used to first assign participants to the device (laptop or smartphone) through which they would complete the study on. Next, Latin Square was used to determine the order of mediums used (website and ERIN) for each participant. The participant would use either the SDCC website then the ERIN chatbot or vice versa. Finally, the Latin Square method was used

to assigned each participant to two different scenarios, one for the SDCC website and one for ERIN. This design ensured consistency, that each device was being tested the same amount, and randomization helped to mitigate order effects. Figure 1, visualizes our 2X2 (website/ERIN, smartphone/laptop) study design.

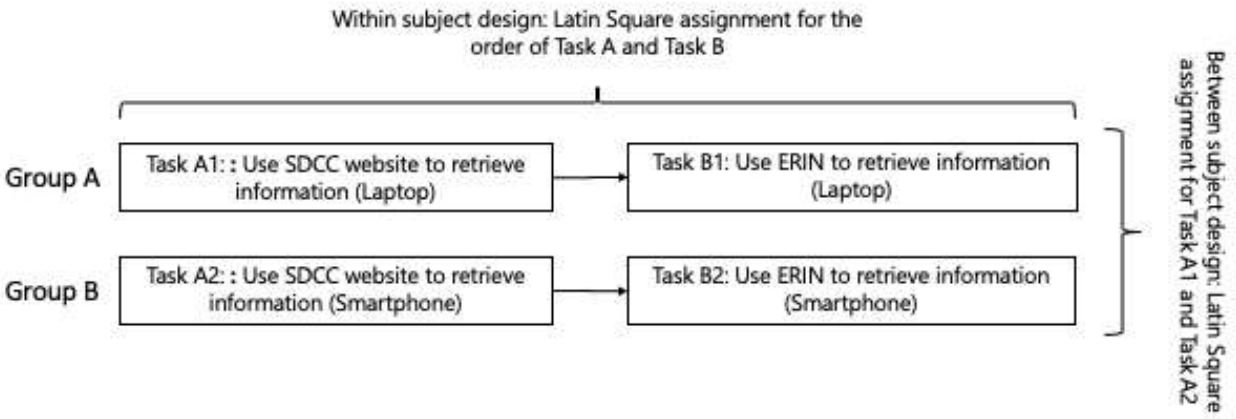


Figure 1: Study design

Throughout our usability study, we asked our participants to use the Think out Loud Method to describe their interactions with the SDCC website and the ERIN chatbot. This was helpful for both the note-taker to confirm the participant's interactions and the interviewer to guide their experience-diagnosis questions. Notes were taken on the utterances the participants made using the Think out Loud Method. This approach allowed us to better understand what the users struggled with and in turn what should be improved.

The usability study was conducted via Zoom calls. We started the study by welcoming participants and verbally reciting our IRB-approved introduction. This established the interviewer and interviewees' understanding of the interview process and at-will participation. The participants' names were altered to their participant ID number and participants were asked to turn their camera off to keep their identity hidden in the study recordings.

After completing each scenario, participants completed a Qualtrics survey which is discussed later in this paper.

4.1.1 Participants

The ERIN empathy chatbot is specifically targeted towards WPI undergraduate students seeking mental wellness resources. More specifically, the target users are 17-25 years old tech-savvy users who have access to the internet. Erin users could come from a variety of socio-economic backgrounds and have various degrees of social involvement on the WPI campus.

Participants were recruited for our user study through word of mouth and emails. Each team member reached out to their personal network, school organizations, and social media. Additionally, participants were recruited through various undergraduate email aliases. In total, 41 WPI undergraduate students participated in our user study.

4.1.2 Device and Medium

In this usability study, we examined the possible impact of the device (laptop vs. smartphone) and medium (chatbot vs. website) on user experience. Erin was used to conduct and observe our participants' interactions with an empathetic chatbot in either a smartphone or laptop environment (Figure 2). In the same manner, the SDCC website and ERIN Chatbot were used to compare participants’ information-seeking behavior and interactions with the two different technology mediums.

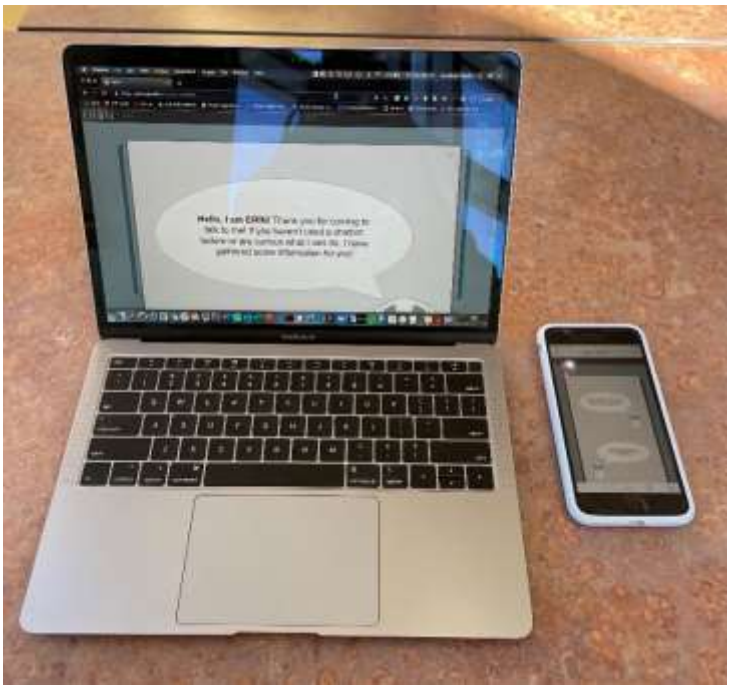


Figure 2: Laptop VS Smartphone Devices

Additionally, the SDCC website acted as a control for our WPI student audience. The mix of these stimuli and instruments enabled our within-subject and a between-subject study design (Figure 1).

4.1.3 Constructs

As previously stated, the goal of this MQP was to evaluate the user experience of the ERIN chatbot and determine improvements to be made in the future. In order to do so, we used similar dependent variables (constructs) outlined in Persons (2020), utilizing all of the constructs used in the previous study as well as some new constructs. The constructs used in our study were as follows:

Perceived Decision-Making Effort (PDME) refers to the amount of cognitive effort it takes users to complete a task. The higher the PDME, the more laboring the task (Wang and Benbasat, 2009).

Perceived Ease of Use (PEOU) is a UX concept referring to how easy it is to use a technology. Higher PEOU means that users find the technology easy to use (Adipat et al. 2011).

Perceived Usefulness (PU) measures how well a technology matches the needs of a user. Higher PEOU values indicate higher perceived usefulness (Adipat et al. 2011).

The goal of Mobile User Experience (MUX) is to holistically assess user experience of mobile devices (Wilson and Djamasbi, 2019).

System Usability Scale (SUS) inspired MUX. They fall under similar categories as they both measure the acceptability of technology. SUS however, was developed before mobile technologies became popular (Tullis and Albert, 2013).

Unified Theory of Acceptance and Use of Technology (UTAUT) is used to assess the acceptability of a technology. UTUAT consists of six constructs: performance expectancy (PE), effort expectancy (EE), social influence (SI), facilitating conditions (FC), hedonic motivation (HM), and behavioral intention (BI) (Viswanath, et al., 2012).

The survey given to participants asked questions regarding each of the above constructs. Each question had five choices ranging from “strongly disagree” to “strongly agree”. This survey, as well as the open-ended questions and user feedback through the think out-loud protocol, allowed us to measure user experience both qualitatively and quantitatively. The questions asked in the surveys can be found in Appendix B.

4.2 Study Results

Individual ratings of each usability metric were recorded per participant and then aggregated to find the average rating for each construct. Average SUS score was also calculated per person for each device and medium. Paired t-tests were conducted to compare the performance of the ERIN chatbot with the SDCC website as a control. Two sample t-tests were

conducted to compare user experience of ERIN chatbot on laptops and smartphones. Qualitative data collected during the interviews were sorted by repeat responses and common themes to report findings more easily. These results are reported in the following sections.

4.2.1 SDCC v ERIN (Paired T-Tests)

Table 1: Comparing SDCC v ERIN Experience (Paired T-Test) for Laptop

Experimental Conditions	SDCC		ERIN		Paired T-Test		
	Mean	SD	Mean	SD	t Stat	df	p-value
Decision Making Effort	3.16	1.15	1.66	0.62	5.27	20	0.00
Ease of Use	3.24	0.90	4.29	0.53	-4.64	20	0.00
Usefulness	2.89	0.87	4.17	0.70	-5.76	20	0.00
Performance Expectancy	2.59	0.68	3.49	0.61	-4.80	20	0.00
Effort Expectancy	3.45	0.98	4.44	0.55	-4.32	20	0.00
Social Influence	3.21	0.75	3.37	0.60	-0.84	20	0.41
Facilitating Conditions	4.29	0.68	4.40	0.56	-1.11	20	0.28
Hedonic Motivation	2.27	0.83	3.79	0.56	-8.31	20	0.00
Behavioral Intention	2.70	0.86	3.16	0.63	-2.94	20	0.01
SUS (System Usability Score)	3.28	0.84	4.31	0.44	-4.65	20	0.00
MUX (Mobile User Experience)	3.48	0.76	4.15	0.50	-4.15	20	0.00

As seen in Table 3, the results of paired t-test show that the following variables were significantly different for: Decision Making Effort, Ease of Use, Usefulness, Performance Expectancy, Effort Expectancy, Hedonic Motivation, Behavioral Intention, SUS, MUX. Additionally, the average SUS score for the SDCC website on a laptop was 57, which indicated

need for improvement while the average SUS score for ERIN on a laptop was 83 which indicated 'good' performance (Bangor, 2009).

Table 2: Comparing SDCC v ERIN Experience (Paired T-Test) for Smartphone

Experimental Conditions	SDCC		ERIN		Paired T-Test		
	Mean	SD	Mean	SD	t Stat	df	p-value
Decision Making Effort	3.65	0.93	2.13	1.18	4.98	19	0.00
Ease of Use	2.71	0.87	4.38	0.75	-7.05	19	0.00
Usefulness	2.38	0.91	3.88	1.24	-4.16	19	0.00
Performance Expectancy	2.2	0.68	3.13	0.98	-4.17	19	0.00
Effort Expectancy	3.28	1.02	4.33	0.88	-3.43	19	0.00
Social Influence	3.17	0.88	3.30	1.00	-0.64	19	0.53
Facilitating Conditions	3.96	0.73	4.20	0.63	-1.55	19	0.13
Hedonic Motivation	1.85	0.68	3.75	0.83	-8.62	19	0.00
Behavioral Intention	2.43	0.88	3.02	0.86	-2.56	19	0.02
SUS (System Usability Score)	3.13	0.83	4.17	0.69	-4.57	19	0.00
MUX (Mobile User Experience)	3.28	0.82	4.08	0.82	-3.11	19	0.01

As seen in Table 2, the results of the paired t-test show that all the variables were significantly different, other than Social Influence (SI) and Facilitating Conditions (FC). In addition, the average SUS score for the SDCC website on a smartphone was 53, which indicated an opportunity for improving the experience of SDCC website on a smartphone. The average SUS score for ERIN on a smartphone was 79, which again indicated 'good' performance (Bangor, 2009).

The findings reported in Table 1 and Table 2 show similar results comparing the UX of SDCC website vs. ERIN regardless the device that participants were used. Therefore, we conducted a paired- test for the combined data (Table 3).

Table 3: Comparing SDCC v ERIN Experience (Paired T-Test) Regardless of Device Used

Experimental Conditions	**SDCC**		**ERIN**		**Paired T-Test**		
	Mean	**SD**	**Mean**	**SD**	**t Stat**	**df**	**p-value**
Decision Making Effort	3.40	1.07	1.89	0.95	7.34	40	0.00
Ease of Use	2.98	0.91	4.33	0.64	-8.00	40	0.00
Usefulness	2.64	0.91	4.03	1.00	-6.69	40	0.00
Performance Expectancy	2.40	0.70	3.32	0.82	-6.39	40	0.00
Effort Expectancy	3.37	0.99	4.38	0.72	-5.44	40	0.00
Social Influence	3.19	0.80	3.33	0.81	-1.06	40	0.30
Facilitating Conditions	4.13	0.72	4.30	0.59	-1.92	40	0.06
Hedonic Motivation	2.06	0.78	3.77	0.70	-11.85	40	0.00
Behavioral Intention	2.57	0.87	3.09	0.75	-3.83	40	0.00
SUS (System Usability Score)	3.20	0.83	4.24	0.57	-6.61	40	0.00
MUX (Mobile User Experience)	3.39	0.79	4.11	0.67	-4.94	40	0.00

As seen in Table 3, the results of paired t-test show that the following variables were significantly different for: Decision Making Effort, Ease of Use, Usefulness, Performance Expectancy, Effort Expectancy, Hedonic Motivation, Behavioral Intention, SUS, and MUX. Additionally, the average SUS score for the SDCC website was 55 indicating room for improvement and the average SUS score for ERIN was 81 which indicates ‘good’ performance (Bangor, 2009).

4.2.2 Laptop v Smartphone (Two sample T-Tests)

Table 4: Comparing ERIN's Experience on Laptop v Smartphone

Experimental Conditions	Laptop		Smartphone		T-Test		
	Mean	**SD**	**Mean**	**SD**	**t Stat**	**df**	**p-value**
Decision Making Effort	3.16	1.15	3.65	1.91	-1.50	39	0.14
Ease of Use	3.24	0.90	2.71	0.87	1.91	39	0.06
Usefulness	2.89	0.87	2.38	0.91	1.82	39	0.08
Performance Expectancy	2.59	0.68	2.2	0.68	1.82	39	0.08
Effort Expectancy	3.45	0.98	3.28	1.02	0.57	39	0.57
Social Influence	3.21	0.75	3.17	0.88	0.57	39	0.88
Facilitating Conditions	4.29	0.68	3.99	0.73	1.41	39	0.17
Hedonic Motivation	2.27	0.83	1.85	0.68	1.77	39	0.08
Behavioral Intention	2.70	0.86	2.43	0.88	0.95	39	0.34
SUS (System Usability Score)	3.28	0.84	3.13	0.83	0.56	39	2.02
MUX (Mobile User Experience)	3.48	0.76	3.28	0.82	0.81	39	0.42

As shown in the Table 4, only perceived ease of use was almost significantly different (p=0.06), none of the other results reached significant level. Additionally, the average SUS score for the ERIN on a laptop was 83 and the average SUS score for ERIN on a smartphone device was 79 which are considered as relatively good scores (Bangor, 2009).

4.2.3 Qualitative Data

After gathering qualitative data from participants' think-out-loud comments and observations, each participant was given a point for each theme they mentioned. These points were aggregated and averaged to find patterns in participants' interactions with either ERIN or

the SDCC website. Percentages were measured for all participants and then separated by smartphone versus laptop.

Table 5: ERIN Qualitative Data (User Utterances: Think-Out-Loud Process and/or Experimenters' Observation)

ERIN Averages				
Topic	**Definition**	**Laptop**	**Smartphone**	**Total**
Tutorial Picture	Participants thought tutorial picture was either confusing, small, or non-interactive. See Appendix C.	14.29%	35.00%	24.39 %
Message Spam	After the first input, participants thought that ERIN spammed messages, needs more concision.	14.29%	20.00%	17.07 %
Font	Participants thought the fonts were too small on smartphone, or that it was unclear between some letters like L and I.	4.76%	15.00%	9.76 %
Submission	Participants were not able to decipher how to submit their choices for resources. Click the choice and press the arrow button.	42.86	50.00%	46.36 %
Scrolling	Participants were not able to figure out how to scroll on the laptop or they thought the organization of the choices should be laid out in a vertical scrolling to feel more normal.	38.10%	15.00%	26.83 %
Simple/Easy /Intuitive	Participants thought the chatbot interface was intuitive to use, felt natural, and natural and familiar to other chatbots.	57.14%	50.00%	53.66 %
Empathetic	Participants felt the chatbot was empathetic and represented a sufficient amount of compassion.	19.05%	35.00%	26.83 %
Anonymous	Participants enjoyed the anonymity of the chatbot.	4.76%	15.00%	9.76 %
Quick/Fast	Participants like it that the chatbot took to a short time to provide resources.	28.57%	20.00%	24.39 %
Fun	Participants found that the chatbot experience was pleasant.	4.76%	5.00%	4.88 %
Content	Participant liked the mental wellness content supplied by the chatbot.	47.62%	55.00%	51.22 %
Feedback	Participants liked feedback from the chatbot. For example, "I am searching my database to determine what resources will be best for helping you."	9.52%	25.00%	17.07 %
Cute	Participants thought that the mascot for ERIN was "cute".	28.57%	55.00%	41 %
Color Scheme	Participants thought the color scheme was appropriate for the chatbot.	4.76%	10.00%	7.32 %
Accessible	Participants felt that the chatbot could be accessed on different devices.	4.76%	0.00%	2.44 %
Pressure	Participants felt pressured to word their interactions with ERIN in a way that ERIN could understand.	4.76%	10.00%	7.32 %
Too Positive	Participants felt the optimism displayed by ERIN seemed inappropriate or too positive.	9.52%	5.00%	7.32 %
Welcome Good	Participants enjoyed the onboarding process before actually using the chatbot.	42.86%	40.00%	41.46 %
Welcome Poor	Participants thought that the welcome screens directions and interactions was not smooth or interactive enough.	23.81%	40.00%	31.71 %

Based on the data displayed in Table 5, we discuss negative experiences for topics with values greater or equal 20%: Tutorial Picture, Submission, Scrolling, Simple/Easy/Intuitive, Empathetic, Quick/Fast, Content, Cute, Welcome Good, and Welcome Poor.

26.83% (Table 5) of all participants during the think-out-loud process mentioned associating a negative experience with the horizontal scrolling feature for resource choices. More specifically, 38.10% of laptop participants said and showed they did not enjoy the horizontal scrolling behavior compared to 15.00% on smartphone. Particular patterns with the horizontal scrolling were that participants thought the horizontal movement was ‘unnatural’ or didn’t allow for all the options to be displayed. The laptop variant of ERIN received comments that the tab that would appear to scroll was unintuitive and hidden.

The onboard page or welcome screen showed and received multiple opportunities to perform better. 24.39% of all participants mentioned a negative experience with the tutorial picture and specifically, 35.00% of smartphone participants struggled with the tutorial picture compared to 14.29% on laptops. Some patterns of comments made were that the tutorial picture was too small or too ‘busy to read’. Additionally, 31.71% of all participants mentioned a negative aspect of the onboarding process. Patterns of comments here were that the onboard pop-up was ‘not interactive’, ‘lengthy’, and ‘bland’.

Finally, 46.34% of all participants struggled with the method to submit input into ERIN. Equally as high between smartphone (50.00%) and laptop (42.86%) users struggled with the submission process. Some observations made were the participants spent a significant amount of time either pressing enter after they selected their resource choices or not understand the arrow was a button.

While the differences were smaller than our stated 20% cut off point, we believe it would be important to also report that 9.76% of all participants mentioned that they struggled with the font ERIN used. Some comments received were that the font was too small on smartphone, or that it was confusing with similar-looking letters or numbers.

Based on the data displayed in Table 5, we discuss positive experiences for topics with values greater or equal 20%. The first positive comment that was repeated was from the 53.66% of users who referred to ERIN as simple, easy or intuitive. Next, 26.83% of participants commented that they found ERIN's empathetic responses to be sufficient. 24.39% of users commented that they enjoyed how quickly they were able to interact with ERIN and find the resources they were looking for. Additionally, participants responded positively to the content that ERIN provided for them (51.22%). Lastly, participants thought ERIN was cute (41%) and had an informative welcome screen (41.46%).

Table 6: SDCC Website Qualitative Data (User Utterances: Think-Out-Loud Process and/or Experimenters' Observation)

Website Averages				
Topic	**Definition**	**Laptop**	**Smartphone**	**Total**
Navigation	Participants mentioned or exhibited challenges with website navigation.	42.86%	45.00%	43.90 %
Too Many Paragraphs	Participants mentioned that content of the website comprised of too many paragraphs.	71.43%	75.00%	73.17 %
Poor Highlight	Participants mentioned specifically that there wasn't proper visual differentiation between links and plain text.	19.05%	35.00%	26.83 %
Good Highlight	Participants mentioned that they enjoyed the highlighting of links and text on the website or like the idea of it more than plain text.	14.29%	10.00%	12.20 %
Content Organization	Participants said that the organization of the content on the website could be improved.	57.14%	45.00%	51.22 %
Useful Content	Participants thought that the content itself was very useful and helpful.	42.86%	45.00%	43.90 %
Right Mental Space	Participants felt that the "right mental mindset" would be needed to appropriately analyze the websites information.	14.29%	20.00%	17.07 %

Based on the results reported in Table 6, we believe some noteworthy patterns to review are: Navigation, Too Many Paragraphs, Poor Highlight, Poor Content Organization, and Useful Content (total value of 20%>=).

5 Discussion

The Objective of this study was to 1) verify the need for ERIN, 2) compare the experience of finding resources on the SDCC website vs. querying the ERIN chatbot, and 3) investigate possible differences in user reactions to ERIN when it is accessed via laptops and smartphones.

To achieve our research goals, we conducted two studies. In the first study we administered a survey to solicit feedback from undergraduate students at WPI, which are ERIN's target market. The analysis of this survey verified the results of key informant interviews (Persons, 2020) which established the need for an empathy chatbot on campus.

The second study was conducted to address our second and third research objectives, which are explained in the following sections.

5.1 ERIN v SDCC

The results of our second study showed that experience of finding resources via ERIN vs. the SDCC website was significantly different as evidenced by values obtained for the following constructs: PDME, PEOU, PU, PE, EE, BI, HM, SUS, and MUX. We attribute the significance of Perceived Decision-Making Effort (PDME) to the fact that over half of users, 53.66% (Table 5), thought that the chatbot interface format of ERIN was intuitive to use, natural, and was familiar to pre-existing chatbots. Additionally, 24.39% (Table 5) of users, as evidenced by their unsolicited remarks during the think-out-loud process, enjoyed the short time it took for ERIN to gather and supply resources based on their utterance. These results support previous research that show that users often find conversational user interfaces natural and casual (Park et al., 2018).

In the case of the SDCC website, 43.90% (Table 6) of users often mentioned or exhibited challenges with navigating the website. Most users, 73.17% (Table 6), also mentioned that the

site consisted of too many paragraphs and, therefore, thought the information was presented in a manner that was difficult to understand. Overall, 51.22% (Table 6) of users said that the organization of the site's content could be improved to provide a better overall user experience and usability.

For PEOU, PE and EE, we attributed the significance differences in user reactions between ERIN and SDCC website to all of the above reasons with the additional note that some users, 17.07% (Table 5), liked the continuous updates and feedback messages from ERIN as it was gathering resources. For example, "I am searching my database to determine what resources will be best for helping you." For PU, 26.83% (Table 5) of users appreciated ERIN's empathy. This feedback obtained during the think-out-loud portion of the study is consistent with previous research by Liu & Sundar (2018). The differences in BI between ERIN and SDCC website can be explained by positive qualitative feedback by a majority of users (51.22%) about the wellness content related to or supplied by ERIN as well as the comment (by 53.66% of participants) on the intuitiveness and general familiarity of the chatbot (Table 5).

SI was not significant when comparing the averages of ERIN to the SDCC website. One possible explanation is that ERIN is not available to the public. Future studies should re-examine SI once ERIN is launched.

When comparing HM, there was a significant difference between the averages for ERIN and the SDCC website. During the think-out-loud portion of the experiment, participants mentioned that they felt ERIN was cute (41.46%, Table 5) and empathetic (26.83%, Table 5). This may have been the reason why their overall experience with ERIN was more enjoyable (as evidenced by higher HM values for ERIN). Additionally, 51.22% (Table 6) of users felt that the

SDCC website was poorly organized, which could be a reason why participants did not enjoy using the SDCC website (as evidenced by the HM values) as much as they enjoyed using ERIN.

For FC, there was not a clear significant difference between the averages for ERIN and the SDCC website. FC refers to the perceived strength of organizational and technical infrastructure to support the use a system. The non-significant results in this case could be attributed to the fact that both the SDCC website and ERIN are provided by WPI network.

According to the SUS scores, there was a clear significant difference between the averages for ERIN and the SDCC website. ERIN's SUS score was 81 which is categorized as 'good', rating according to Bangor while the SDCC website had a SUS score of 55 which indicated opportunity for improvement (2009). These scores showed a clear difference between the two mediums, with ERIN providing a better user experience.

When comparing the averages for MUX, we were able to determine that there was a significant difference, indicating that users found ERIN (compared to SDCC website) to provide a better mobile experience. Similarly, 2.44% (Table 5) participants said that they found ERIN more accessible than the SDCC website. These results are interesting because they were independent of the of the device used (laptop vs. smartphone). While mobile experience can be influenced by both the device or medium (Wilson and Djamasbi 2019), our results indicate that mobile experience in our study was related to the medium (website vs. chatbot).

5.1.1 Insights

Our ERIN v SDCC study showed that users had a clear preference for the ERIN chatbot over the SDCC website. ERIN was more intuitive and easier to use than the SDCC website. Additionally, users found ERIN easier to navigate overall, and were able to find resources faster than they would on the SDCC website. ERIN proved to be more inviting and enjoyable to use

than the SDCC website. Most notability, users were able to perceive and responded positively to ERIN's empathic responses.

5.2 Smartphone v Laptop

We compared differences in user experience of ERIN between devices via the same construct that we reported in the previous section. The averages in user reactions between smartphone devices and laptop devices for PDME, PEOU, PU, and HM all proved to be non-significant. However, all four of these constructs were showing a slight trend towards laptop devices. We attribute this trend, even though non-significant, to the 9.76% (Table 5) of users who commented that the font size of ERIN was too small on smartphone. Additionally, 35.00% (Table 5) of smartphone users mentioned issues with the tutorial picture in ERIN's welcome pop-up compared to only 14.29% (Table 5) of laptop users making similar mentions, refer to Appendix C.

The averages between smartphone devices and laptop devices for PE, EE, and FC also proved to be non-significant. The results suggest that ERIN performs equally well on both devices. Participants (57.14% using laptop and 50% using smartphone) provided unprompted positive remarks indicating that ERIN was simple, intuitive, and easy to use. These findings are consistent with previous research that suggests laptops and smartphones can provide equally well mobile experiences (Wilson and Djamasbi 2019).

The averages between laptop and smartphone devices for SI and BI also proved to be non-significant. This lack of significance can be attributed to user indifference towards the device used.

The average between laptop and smartphone devices for SUS proved to be non-significant as well. The SUS scores for ERIN using laptop and smartphone were similar; ERIN on laptop scored an 83 and on smartphone scored a 79. Both are categorized as providing relatively ‘good’ experience under the grading template for empirical SUS scores (Bangor, 2009), indicating that ERIN was well received on both devices. Similarly, MUX t-test results were non-significant between laptop and smartphone. These non-significance results could suggest that a laptop and smartphone devices are closely associated in a student’s life (Wilson and Djamasbi, S., 2019).

5.2.1 Insights

Our smartphone v laptop study analysis showed that devices had little influence on user experience with ERIN. General feedback regarding ERIN’s performance trended positively towards laptop devices. Additionally, we were able to gather feedback on user preferences that differ between smartphone and laptop. Most notable was the need to improve the responsive design of ERIN that simply transferred the laptop interface onto a smartphone screen. Some design elements that performed well on a laptop, did not transfer optimally on a smartphone device. Examples of these elements include the font size of ERIN on a smartphone device, as well as the tutorial picture on the welcome screen, which can be found in Appendix C.

5.3 Limitations

In our first study we were not able to gather feedback in-person allowing for a more natural conversation. While surveys allow for a faster data collection, they inherently provide a narrow set of responses. Future studies can extend our findings by conducting individual interviews or focus groups. Despite its limitations, the results of our first study confirmed the

results of a previous study that was obtained from interviews with key informants (Persons, 2020).

Our second study also presented limitations. Since the study was conducted via Zoom (due to COVID crisis), we could not perform in-person UX research techniques such as eye tracking. Eye tracking could have been beneficial to more holistically assessing ERIN's user experience. Additionally, the virtual climate limited our ability to observe user reactions. The duration of this project, specifically the time we had to conduct user studies, limited the number of participants and therefore limited the amount of verbal feedback we were able to receive regarding ERIN's performance. The duration of the study also prevented us from being able to iterate on the ERIN platform to include short term user feedback, this will be discussed in the next section.

5.4 Future Projects

One of the benefits of Study 2 was being able to identify limitations that exist within the current ERIN chatbot. These limitations can be addressed in both short and long term projects.

A few ways to address limitations in the short-term include fixing the horizontal scrolling feature, revamping the onboarding process, modifying font size and the submission approach. To sum up all the concrete feedback, we have recommendations to incorporate into the next iteration of ERIN. The first is to restructure the onboarding process into a clickthrough format with an interactive tutorial picture and better font choice to include more unique characters (e.g. clear differentiation between a lowercase L and an uppercase I). This would avoid the long scrolling of the onboard page, as well as create an immersive experience to maintain user engagement. Another recommendation is to change the horizontal scrolling of the resource choices to a vertical scrolling or drop-down menu to better display all the choices. This would be more

inclusive for both smartphone and laptop interfaces. The next recommendation would be to change the method of submission to a more passive style like a click or physical ‘enter’ click on a keyboard. Finally, optimize the smartphone variant of ERIN for a more natural experience.

In the long term, we recommend that more smartphone user testing be performed on ERIN. This study was the first time that ERIN was tested with a larger pool of students, hence the study was able to reveal nuances that provided insight for refining ERIN in the next development cycle.

Additionally, as previously reported, 17.07% (Table 5) of users experienced malfunctions while using ERIN. One particular user experienced a malfunction after typing a long story into ERIN, rather than a short description. We recommend that in the future ERIN is improved to expand its ability to respond to such situations. By training ERIN to recognize various and longer user utterances, ERIN will perform better in different situations. We believe that expanding user utterances will improve the user experience with ERIN. This could be accomplished by either brick-and-mortar coding or through machine learning.

Once the above recommendations are implemented, we believe that the services ERIN chatbot provides could be expanded to benefit larger populations. ERIN could provide resources for employees and customer in various organizations from universities to companies to hospitals. The first institution that ERIN would suit well would be the WPI SDCC. We believe that ERIN would function well as an expansion to the SDCC website, helping students to find information quickly. ERIN could also perform well in high stress work situations such as hospitals. For example, employees who experience burnout in high stress environments could use ERIN to find resources for stress reduction. Our results suggest that ERIN has the potential to serve as an

expansive and powerful platform to make a positive impact not only on college campuses, but also in other organizations.

6 Conclusion

We conducted two studies to evaluate ERIN Chatbot which was developed in the UXDM lab at WPI (Persons 2020). In our first study we gathered information about the need for a wellness chatbot (such as ERIN) to supplement the SDCC website at WPI. In our second study, we conducted 41 user interviews using the SDCC website as a control for ERIN. We found significance differences in user experience between ERIN and the SDCC website, which showed ERIN performed better than the SDCC website. When testing ERIN's experience between laptop and smartphone devices, we found the data to be non-significant indicating that these two devices had no impact on ERIN's user experience and were equally well-perceived by our participants. Our qualitative data, collected during interviews, helped strengthen the results of our quantitative analysis.

Our study was able to provide valuable insight into chatbot interactions in mental wellness by providing information about possible impacts of medium and device on user experience of information retrieval. Additionally, qualitative data in our study gathered through the think-out-loud protocol and experimenters' observations provided concrete suggestions for improving the user experience of the ERIN chatbot. We are confident that a future expansion of ERIN could help improve access to WPI resources other than those provided by SDCC, and would also prove to be useful in organizations outside the WPI community.

References

Adipat, B., Zhang, D., & Zhou, L. (2011). The effects of tree-view based presentation adaptation on mobile web browsing. Mis Quarterly, 99-121.

American College Health Association. (2019). Fall 2019 Reference Group Executive Summary. https://www.acha.org/documents/ncha/NCHA-III_Fall_2019_Reference_Group_Executive_Summary.pdf

Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. Journal of usability studies, 4(3), 114-123.

Bröhl, C., Rasche, P., Jablonski, J., Theis, S., Wille, M., & Mertens, A. (2018, July). Desktop PC, tablet PC, or smartphone? An analysis of use preferences in daily activities for different technology generations of a worldwide sample. In International Conference on Human Aspects of IT for the Aged Population (pp. 3-20). Springer, Cham.

Budiu, Raluca. "Between-Subjects vs. Within-Subjects Study Design." *Nielsen Norman Group*, 13 May 2018, https://www.nngroup.com/articles/between-within-subjects/.

Budiu, Raluca. "The User Experience of Chatbots." *Nielsen Norman Group*, 25 November 2018, https://www.nngroup.com/articles/chatbots/.

Ranoliya, B. R., Raghuwanshi, N., & Singh, S. (2017, September). Chatbot for university related FAQs. In 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI) (pp. 1525-1530). IEEE.

Cameron, G., Cameron, D., Megaw, G., Bond, R., Mulvenna, M., O'Neill, S., ... & McTear, M. (2018, July). Best practices for designing chatbots in mental healthcare–A case study on iHelpr. In Proceedings of the 32nd International BCS Human Computer Interaction Conference 32 (pp. 1-5).

Dholakia, P. (2019, December 04). Emotional ai: Empathy in chatbots. Retrieved March 10, 2021, from https://medium.com/kevit-technologies/emotional-ai-empathy-in-chatbots-b573df34afbf

Djamasbi, S., & Strong, D. (2019). User Experience-driven Innovation—Theory and Practice: Introduction to Special Issue. *AIS Transactions on Human-Computer Interaction, 11*(4), 208-214. https://doi.org/10.17705/1thci.00120
DOI: 10.17705/1thci.00120

Djamasbi, S., & Strong, D. (2019). User Experience-driven Innovation in Smart and Connected Worlds. AIS Transactions on Human-Computer Interaction, 11(4), 215-231.

Ghosh, A. (2018, September 24). UX Research is Essential to Product Success. Retrieved February 06,2021, from https://www.uxmatters.com/mt/archives/2018/09/ux-research-is-essential-to-product-success.php#:~:text=into%20thin%20air.-,User%20research%20provides%20an%20essential%20foundation%20for%20design%20strategy.,who%20would%20use%20your%20product.

Jain P., Djamasbi S., Wyatt J. (2019) Creating Value with Proto-Research Persona Development. In: Nah FH., Siau K. (eds) HCI in Business, Government and Organizations. Information Systems and Analytics. HCII 2019. Lecture Notes in Computer Science, vol 11589. Springer, Cham. https://doi.org/10.1007/978-3-030-22338-0_6

Lin, Z., Xu, P., Winata, G. I., Siddique, F. B., Liu, Z., Shin, J., & Fung, P. (2020). *CAiRE: An Empathetic Neural Chatbot.* Hong Kong: Center of Artificial Intelligence Research. Retrieved May 21, 2020, from https://arxiv.org/pdf/1907.12108.pdf

Liu, B., & Sundar, S. S. (2018). Should Machines Express Sympathy and Empathy? Experiments with a Health Advice Chatbot. Cyberpsychology, behavior and social networking, 21(10), 625–636. https://doi.org/10.1089/cyber.2018.0110

Lucas, G. M., Gratch, J., King, A., & Morency, L. (2014). It's only a computer: Virtual humans increase willingness to disclose. Computers in Human Behavior, 37, 94-100. doi:https://doi.org/10.1016/j.chb.2014.04.043

Miaskiewicz, T., & Kozar, K. A. (2011, September). Personas and user-centered design: How can personas benefit product design processes? *Design Studies*, *32*(5), 417-430. doi:10.1016

Morris R, Kouddous K, Kshirsagar R, Schueller S

Towards an Artificially Empathic Conversational Agent for Mental Health Applications: System Design and User Perceptions

J Med Internet Res 2018;20(6):e10148

URL: https://www.jmir.org/2018/6/e10148

DOI: 10.2196/10148

Mowbray, C. T., Megivern, D., Mandiberg, J. M., Strauss, S., Stein, C. H., Collins, K., . . . Lett, R. (2006). Campus mental health services: Recommendations for change. American Journal of Orthopsychiatry, 76(2), 226-237. doi:10.1037/0002-9432.76.2.226

Nass, C., Steuer, J., & Tauber, E. R. (1994, April). Computers are social actors. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 72-78).

Nielsen, J. (1993, November). Iterative user-interface design. *Computer, 26*(11), 32-41. doi: 10.1109/2.241424

Park, M., Aiken, M., & Salvador, L. R. (2018). How do Humans Interact with Chatbots?: An Analysis of Transcripts. *International Journal of Management and Information Technology*, 3338-3350. doi:10.24297/IJMIT.V14I0.7921

Persons, B. (2020). ERIN Chatbot - User Experience and Decision Making. : Worcester Polytechnic Institute.

Pruitt, J., & Grudin, J. (2003). Personas: Practice and Theory. *DUX '03: Proceedings of the 2003 Conference on Designing for User Experiences*, (pp. 1-15). doi:10.1145/997078.997089

Rohrer, Christian. "When to Use Which User-Experience Research Methods." *Nielsen Norman Group*, 12 Oct. 2014, www.nngroup.com/articles/which-ux-research-methods/.

Signh, B. R. (2017, January 29). *Chat Bots — Designing Intents and Entities for your NLP Models*. Retrieved May 20, 2020, from Medium: https://medium.com/@brijrajsingh/chat-bots-designing-intents-and-entities-for-your-nlp-models-35c385b7730d

Tullis, T., & Albert, B. (2013). *Measuring the user experience: Collecting, analyzing, and presenting usability metrics*. Amsterdam: Elsevier/Morgan Kaufmann.

Venkatesh, V., Morris, M., Davis, G., & Davis, F. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, *27*(3), 425-478. doi:10.2307/30036540

Wang, Weiquan, and Benbasat, Izak (2009), "Interactive decision aids for consumer decision making in e-commerce: the influence of perceived strategy restrictiveness", MIS Quarterly, 33, 293-320.

Wilson, E. V., & Djamasbi, S. (2019). Measuring Mobile User Experience Instruments for Research and Practice. Communications of the Association for Information Systems, 44, pp-pp. https://doi.org/10.17705/1CAIS.04408

What is Human-Centered Design? (n.d.). Retrieved May 21, 2020, from Design Kit: https://www.designkit.org/human-centered-design

What is Title IX? (n.d.). Retrieved December 14, 2020, from https://titleix.harvard.edu/what-title-ix

Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., & Torous, J. B. (2019). Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape. Canadian journal of psychiatry. Revue canadienne de psychiatrie, 64(7), 456–464. https://doi.org/10.1177/0706743719828977

Appendix A: Need Discovery Survey

Questions	Scale
What resources are you aware of on the WPI campus?	Free response
Are you aware of the services that the SDCC provides?	Yes, No, Somewhat (Please Explain)
Do you know what services they provide?	Yes, No, Somewhat (Please Explain)
If you had the option to have an extra channel of communication, which would you prefer?	Chatbot, Email Alias, Hotline, None
Would you be willing to help with a project in the future?	No, Yes (leave your email)

Appendix B: Survey Questions

Scale: Strongly Disagree, Disagree, Neither Disagree or Agree, Agree, Strongly Agree
Thinking about the task you just completed with <medium>, please rate the following:

Survey Questions
Decision Making Effort • The task of obtaining resources using <the medium> took too much time. • Selecting resources using <the medium> required too much effort. • Selecting resources using <the medium> was too complex.
Ease of Use • I found <the medium> was easy to learn. • I found <the medium> was easy to use. • I found <the medium> was displayed in a way that was clear and understandable.
Usefulness • I found <the medium> was useful for completing the task. • I could find information I wanted quickly. • I found <the medium> was displayed in a way that was useful in searching for information.
Performance Expectancy • I find <the medium> useful in my daily life. • Using <the medium> helps me accomplish things more quickly. • Using <the medium> increases my productivity.
Effort Expectancy • Learning how to use <the medium> is easy for me. • My interaction with <the medium> is clear and understandable. • I find <the medium> easy to use. • It is easy for me to become skillful at using <the medium>.
Social Influence • People who are important to me think that I should use <the medium>

- People who influence my behavior think that I should use <the medium>.
- People whose opinions that I value prefer that I use <the medium>.

Facilitating Conditions

- I have the resources necessary to use <the medium>.
- I have the knowledge necessary to use <the medium>.
- <the medium> is compatible with other technologies I use.
- I can get help from others when I have difficulties using <the medium>.

Hedonic Motivation

- Using <the medium> is fun.
- Using <the medium> is enjoyable.
- Using <the medium> is very entertaining.

Behavioral Intention

- I intend to continue using <the medium> in the future.
- I will always try to use <the medium> in my daily life.
- I plan to continue to use this <the medium> frequently.

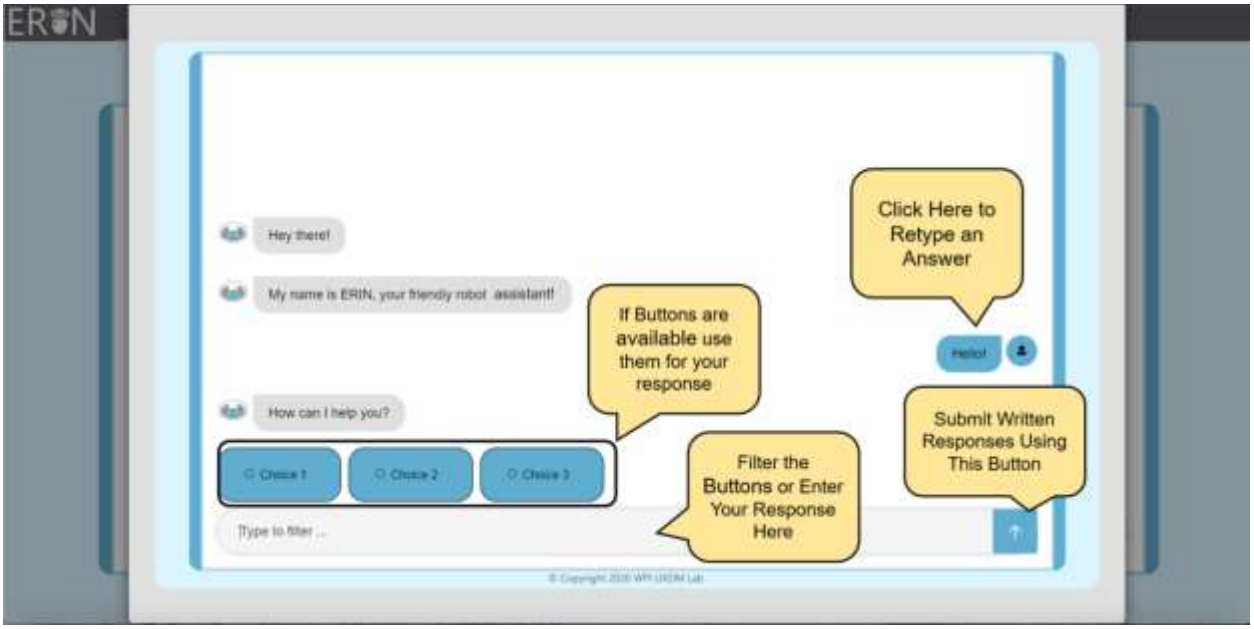
SUS (System Usability Score)

- I think that I would like to use <the medium>.
- I found the <the medium> unnecessarily complex.
- I thought <the medium> was easy to use.
- I think that I would need the support of a technical person to be able to use <the medium>.
- I found the various functions in <the medium> were well integrated.
- I thought there was too much inconsistency in <the medium>.
- I would imagine that most people would learn to use <the medium> very quickly.
- I found <the medium> very cumbersome to use.
- I felt very confident using <the medium>.
- I needed to learn a lot of things before I could get going with <the medium>.

MUX (Mobile User Experience)

- I felt using <the medium> to obtain resources at WPI would slow me down.
- Using <the medium> to obtain resources at WPI made me feel disconnected.
- <the medium> provided a good view of information.
- Clicking on links or buttons was easy to accomplish while using <the medium>.
- I would be able to use <the medium> on the go to obtain resources at WPI.
- I think <the medium> used to obtain resources at WPI would be easy to carry with me.

Appendix C: Tutorial Picture



Appendix D: ERIN

This screenshot was taken of the ERIN interface.