Penn State University Park
IST521 Human-Computer Interaction
Fall 2020

LionSpeed Interface 2.0: A Walking Speed Device

Jessica M. González-Vargas (jmg900@psu.edu)
Courtney Cole (cmc6503@psu.edu)

Submission Date: December 15th, 2020

1. ABSTRACT

Measuring walking speed, which some medical professionals refer to as “the sixth vital sign,” is an uncommon but insightful means to providing information about the overall patient health that can provide another layer of patient information to compliment other vital signs, such as blood pressure and body temperature. Although walking speed shows promise as a standard vital sign measurement according to some professionals, standardization in measuring walking speed is still far away due to biases and methods to conduct a walking speed test which can lead to inaccurate results. To combat this issue, the LionSpeed device was created as an automated device that standardizes how to measure walking speed. While the device itself can standardize the measurement procedure, providing an interface for healthcare practitioners to collect and review information is necessary for making this a feasible method of providing accurate and trackable information about patient health and recovery. Therefore, we developed a prototype of an interface to complement the LionSpeed device. This interface provides the user with a record-based platform where they can create, add, or update their patient’s record based on the walking speed, fatigue, breathlessness, pain, and other physical characteristics. A previous usability study was conducted with Interface 1 for the first iteration (N=12) and improvements were made upon that interface. To streamline the second iteration of the interface, we conducted another usability study with Interface 2 (N=10) for the second iteration. Specifically, we focused on Interface 2 in this report, where we gathered data through verbal protocol analyses (VPAs) and keystroke logs that evaluate the users’ interactions with the LionSpeed interface. Results from this study provided us with ways to improve the LionSpeed interface and the usability of the interface, such as tackling issues with the ease of use and improving overall interaction with knowing what buttons to click, depending on the task. Future work would be dedicated into improving the interface prototype and making a fully-programmed version that works with the LionSpeed device itself, which then can be tested in a clinical setting.

2. INTRODUCTION

In the medical field, walking speed has been identified as an additional vital sign that can be used to indicate the overall health of a patient, similar to heart rate, body temperature, blood pressure, and pulse rate (Lusardi, 2012). Specifically, walking speed can be used to predict medical conditions and gauge patient recovery (Lusardi, 2012), and can indicate risk of falls (Guimaraes & Isaacs, 1980), a patient’s functional dependency (Woo, Ho, & Yu, 1999), and recovery time upon having a stroke (Brach & VanSwearingen, 2002). Walking speed has also been used in predicting peripheral artery disease (McDermott et al., 2011), a disease that causes decreased blood flow to legs, stomach, arms and head (Kullo & Rooke, 2016), as well as hip fractures (Ingemarsson, Frändin, Mellström, & Möller, 2003).

Traditionally, to measure walking speed, patients are asked to walk a certain distance; first at a normal pace, and then as quickly as they can (Middleton, Fulk, Beets, Herter, & Fritz, 2016). Such distances range from 6 to 15 meters, and a timing device such as a stopwatch can be used to record the time it take to walk a specified distance, which is then converted into the walking speed value (Dobkin, 2006). However, most practices with measuring walking speed are flawed in the sense that results can be inaccurate due to lack of reliability of measurements, as it can be difficult for various people to measure walking speed the exact same way, and setting up areas to accurately record walking speed can be difficult (Kong et al., 2016). This is problematic, as the ability to track a patient’s walking speed and identify reductions in walking speed has been shown to provide critical insights regarding the patient's health (Odonkor et al., 2013), but this cannot be done without a systematic and automated method that allows medical practitioners to measure walking speed.

Around Fall 2019, an engineering capstone project team designed and constructed a prototype of an automated walking speed sensor device called "LionSpeed." This device uses Lidar technology to determine the walking speed of a patient. While the device's functionality is sufficient for operation, there is still a need to understand how the device would be implemented in clinics, and what the optimal experience of interacting with the device would look like. To understand this better, our team was approached by Dr. Everett Hills and Dr. Sven Bilen to identify how the current LionSpeed sensor can be implemented in clinical practice, and how this product can be effectively launched into the market. Specifically, using techniques from the risk-driven spiral model (Pew & Mavor, 2007), we created multiple iterations of a prototype of an interface while running usability studies and consulting with stakeholders throughout the process. Therefore, the goal of this paper is to introduce the process leading up to the development of designing a user interface for medical practitioners who will be measuring walking speed with the already-existing walking speed measurement device, LionSpeed.

3. LITERATURE REVIEW

In this section, we review the literature on walking speed with respect to various types of vital sign measurements, including walking speed, and expand this to clinical walking tests. From there, we look at walking speed gait devices and how they compare. We also review such studies that use digital applications to measure walking speed from a human-computer interaction standpoint and describe how our work builds on the gaps in making measuring walking speed an easy and accurate vital sign measurement.

3.1 Vital Sign Measurements and Walking Speed

In medical practice, vital signs are used to provide information about a person's health, typically using blood pressure, respiratory rate, body temperature, and pulse rate as common measurements (Chester & Rudolph, 2011). Recently, prior studies have shown promise for establishing pain and walking speed as standard vital signs in medical practice (Fritz & Lusardi, 2009; Lusardi, 2012; Middleton, Fritz, & Lusardi, 2015), however implementation for walking speed specifically has been slow due to issues accuracy in measurement (Kong et al., 2016). This is problematic, as walking speed is dependent on the synchronous functioning of physiological systems, such as the cardiovascular, respiratory, nervous and musculoskeletal systems (Ferrucci et al., 2000), where detecting issues with walking speed could help identify underlying conditions that other vital signs may not. Furthermore, depending on the age and health status of a patient, changes in walking speed are usually an indication of reduction in functional ability (Brach & VanSwearingen, 2002), and tracking increases or decreases in walking can help monitor whether patients' conditions are improving or declining. Specifically, walking speed has been used in patients recovering from strokes (Bijleveld-Uitman, van de Port, & Kwakkel, 2013; Witte & Carlsson, 1997), peripheral artery disease (McDermott et al., 2011), back pain (Müller, Ertelt, & Blickhan, 2015), hip fractures (Ingemarsson et al., 2003), and spinal cord injuries (van Hedel, 2009), emphasizing its potential impact for widespread use in medical practice.

Independent of age, walking speed has also been used for early diagnosis of medical conditions, specifically linking slower walking speeds during individuals' mid-life years to declining health and an increased risk of mortality (Elbaz et al., 2013). Furthermore, meta-analysis shows that several studies point to a connection between slower walking speeds and cognitive decline, as well as inability to function (Cooper et al., 2010). In addition to identifying possible medical conditions, walking speed has been shown to explain *to what extent* a patient has recovered from a stroke through their own classification system (Bowden, Balasubramanian, Behrman, & Kautz, 2008). This aligns with findings from short rehabilitation settings, where walking speed was successfully implemented in determining improvements of patients' conditions from the time of admission to time of discharge (Barthuly, Bohannon, & Gorack, 2012). As a result of these findings, walking speed shows promise as a standard vital sign in medical practice, but only if a measurement technique via a physical device and easy-to-use interface that promotes accuracy in measuring walking speed can be implemented.

3.2 Clinical Walking Tests

In medical literature, walking tests are split between the categories of endurance, speed, and quality. Particularly, the six-minute walking test is the most common test used in clinical practice for measuring endurance, where the patient's evaluation is based on how much distance they can walk in a six-minute period (Faggiano, D'Aloia, Gualeni, Brentana, & Cas, 2004). These tests are typically performed on a 30-meter track and the distance is determined by the number of laps walked (Faggiano et al., 2004). Although this method is easy to perform and interpret for measuring and monitoring functional limitations (Faggiano et al., 2004) and has been expanded to successfully covering conditions such as chronic obstructive pulmonary disease, lung cancer, coronary artery disease, diffuse parenchymal lung disease, non-cystic fibrosis bronchiectasis, and adults with fear of falling (Bohannon & Crouch, 2017), constraints due to the distance needing to be covered can make this method difficult to implement in smaller areas. In exchange, clinical walking speed tests, which require less physical space and take less time, can be implemented to measure a patient's walking speed and still convey information on the patient's overall health (Graham, Ostir, Fisher, & Ottenbacher, 2008).

Although the previously mentioned methods of performing clinical walking tests output a numerical value to characterize patients' walking abilities, many practitioners still prefer to conduct qualitative observations of their patients' gaits, which are used to classify different walking patterns during the gait cycle (Wang, Lin, Yang, & Ho, 2012) and are defined as stance, push-off, swing, and heel-strike phases (Godha & Lachapelle, 2008). Specifically, length and width of a stride, stride consistency, sway (balance), and double-support duration (time where the patient stands with both feet on the ground without moving) have been used to characterize gait qualities (Maki, 1997). Abnormal gaits have been shown to indicate malfunctions in multiple bodily systems (Imms & Edholm, 1981), as well as higher risks of falls that can lead to more severe injuries (Maki, 1997). For these reasons, patients' gaits tend to be used over numerical measures, as the qualitative observations provide information on the *quality* of how patients walk. This is problematic, as providing all facets of information from walking tests, specifically numerical data, provides medical practitioners with a trackable view of how patients' walking capabilities improve or decline over time. Although gait observations can be used to observe situations such as the ability to get out of a chair or climb stairs (Imms & Edholm, 1981), without a numerical measure, medical practitioners lack this level of detail to *accurately* track progress in patients. Therefore, a method of being able to combine these aspects derived from walking tests is necessary for pushing walking speed tests towards establishment as a standard vital sign measure, specifically in the form of an electronic device and interface.

3.1 Walking Speed Gait Devices

Generally, medical practitioners conduct walking speed tests using one of the two types of traditional approaches; walking a specified distance, which is then divided by the time recorded on a stopwatch, or walking on a treadmill (Green, Forster, & Young, 2002; Kiss, 2010; Peters, Fritz, & Krotish, 2013). However, both methods are highly subjective and are subject to inaccuracies in recording. To improve the validity of gait tests, researchers developed devices to measure gait in a more accurate and reliable manner.

In recent years, an iOS application called, "The 6th Vital Sign" was developed as a smartphone-based method of assessing population health (Morey et al., 2017), however it was found that this application underestimates the subject's walking speed in comparison with manual measurement and demonstrated poor validity overall (Martin et al., 2019). Another product used to measure walking speed is the Gait Mat II, which is a computer-interfaced system paired with a level pressure mat system (measuring 3.66 m or 11.91 ft long) that transfers data to an Excel sheet and has been shown to produce reliable results when used in nursing homes (Samantha Fien, Tim Henwood, Mike Climstein, Evelyne Rathbone, & Justin W. L. Keogh, 2019; Samantha Fien, Timothy Henwood, Mike Climstein, Evelyne Rathbone, & Justin William Leslie Keogh, 2019). Wearables such as wrist sensors can also be used to determine the speed of an individual while walking or running, but recent developments show that such technologies produce errors if wrist sensors are applied in a "one-size-fits-all" approach and not to specific individuals (Soltani, Dejnabadi, Savary, & Aminian, 2020). Other methods of measuring walking speed rely on gaitspeedometers (GSMs), where results show high validity when compared with actual time measurements, however it has not yet been validated for a clinical setting (Jung et al., 2019).

Although many of these devices have been shown to produce valid results, how easy is it for users in the medical field to operate these devices while maintaining validity in the results? To answer this question, it is necessary to investigate the usability of an interface that the user must rely on to perform the appropriate actions and retrieve necessary information. Specifically, analyzing the effectiveness, efficiency, and satisfaction based on users' interactions with an interface can be used to determine overall usability (Hornbæk, 2006). Furthermore, ensuring that an interface meets requirements based on usability heuristics increases the chances for success (Nielsen, 1992, 1994). Such issues with ignoring criteria for usability have been shown with electronic health record systems, where systems that lack prior usability tests score low in terms of usability, asserting the message that usability testing could improve the chances that a design is integrated with existing workflow processes in a clear and efficient manner (Corrao, Robinson, Swiernik, & Naeim, 2010). Moreover, previous research with comparing tasks of a drug-monitoring program showed that looking at measures of usability such as number of clicks and time to complete a task can help to understand how improve tasks within an interface overall (Poon et al., 2016). In addition, it is important to understand how learning curves could play a role in interface usability, where tasks get faster with practice and tasks that are very similar in difficulty task can be modeled with a similar learning curve (Ritter & Schooler, 2001). Based on these principles of usability, this paper aims to combine such evaluations of usability with our novel interface design to promote walking speed as a viable vital sign to be recorded in clinical settings.

4. RESEARCH OBJECTIVES

The main objective of the current paper was to understand how users interact with the LionSpeed interface through various analysis techniques, with the goal of creating a more viable interface. Specifically, the following research questions (RQs) were explored:

RQ1: What is the perceived and measured usability of the *current* LionSpeed interface when tested with target users?

Our hypothesis was that the time to complete both Task 1 and Task 2 in the newest iteration of the interface (created in Axure RP 9) would take less time in comparison to the previous interface (created in Adobe XD). This is important to analyze, as iterations of designs have been shown to improve the overall quality (source from ACM creativity paper?). Performing this analysis would show whether the current iteration actually improves the usability of the interface. For these reasons, we would also expect the satisfaction of users to increase, as we included more features that had been suggested before from Interface 1's study.

RQ2: How does the time on task relate to number of clicks in task?

The purpose of this RQ was to explore how using number of clicks and time for each task from keystroke data complements usability metrics in RQ1. Specifically, this RQ aims to show how keystroke data from Recording User Input (RUI) software can be used to explore users' interactions with an interface (Kukreja, Stevenson, & Ritter, 2006). Because a higher number of clicks tends to point to greater usability issues with an interface (Bevan, 1997), as shown in previous studies with interfaces for medical purposes (Poon et al., 2016), we wanted to explore the if any noticeable trends existed with respect to the time and the number of clicks from participant data.

RQ3: What are the factors that impact positive and negative satisfaction of users with the interface?

The purpose of this RQ was to explore what positive and negative qualitative measures impact users' satisfaction, based on the verbal protocol analysis (VPA). Comments from the VPA can be used to understand what aspects of the interface help users perform the tasks better, and what harms their abilities to perform the tasks.

5. METHODOLOGY

To answer the research questions presented above, empirical studies were conducted remotely due to COVID-19 restrictions. To answer these research questions, two datasets are utilized. The first dataset was taken from a previous study conducted for the Interaction Design (IE 548) course at The Pennsylvania State University, while the second dataset was conducted for the Human-Computer Interaction (IST 521) course at The Pennsylvania State University. More details about the participants, metrics, and procedure are highlighted in this section.

5.1 Participants

The first dataset was taken from a previous study (N=12) that aimed to compare the usability of the LionSpeed Interface created in AdobeXD. A total of three medical practitioners from family medicine, geriatric medicine, and physical therapy participated in this study, and nine nursing students in their junior or senior year enrolled at The Pennsylvania State University participated in this study. For the second dataset, data was analyzed based on a new usability study (N=10) conducted with medical professionals and nursing students as well. A total of two physician's assistants from the cardiology department of a hospital in southwestern Pennsylvania participated in this study, and eight nursing students in their junior or senior year enrolled at The Pennsylvania State University were recruited for this study. Each participant in both studies was given a unique identification number, and no personal identifiable information was used. Both studies were conducted following IRB guidelines, and participants were compensated with a $15 Amazon gift card.

5.2 Materials

To conduct this study, several tools were used. For verbalizations from the VPA task, the participants consented to being recorded via Zoom. Particularly, video and audio were recorded for each participant, and transcriptions were generated in otter.ai from audio files that had been cleaned up in Adobe Audition 2020. From there, one of the experimenters checked the transcriptions with the original audio file to correct any errors in transcription. The transcriptions were coded using the NVivo Software on Windows OS, which was run on a 13-inch 2018 Dell Inspiron laptop with 1.6 GHz Intel Core i5 processor and 8 GB of RAM, and Mac OS, which was run on a 13-inch 2015 MacBook Pro with a 3.1 GHz Intel Core i7 processor and 16 GB of RAM.

5.3 Interface Design

A LionSpeed interface was created and a usability study was conducted in a previous class using Adobe XD. Using the positive features of the interface and incorporating the improvements from the previous usability study, a similar but improved design of the interface was created in Axure RP 9. This interface can be found on the following link: https://706ssr.axshare.com/. With this, the participants from the second study had to complete the same two tasks as the first usability study but utilizing the LionSpeed 2.0 interface. The main screen of the interface is shown in Figure 1, and it provides the user with a list of the patient records that they currently have, with an option of searching for a specific patient. The user can click on the patient profile and click on "View" to view the record of the patient.

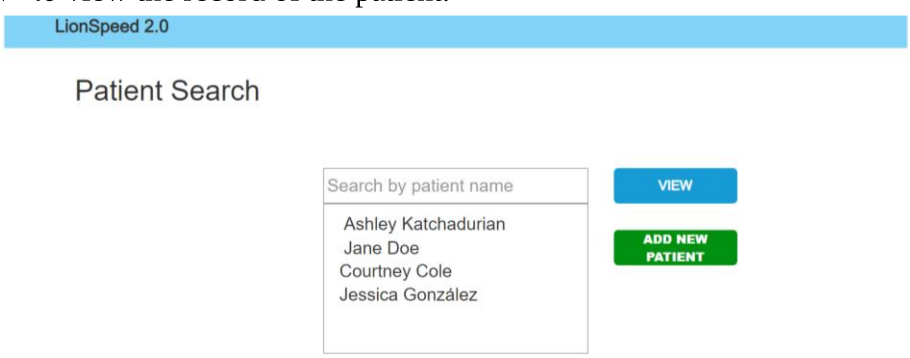


Figure 1: Main Screen

For each of the patients, the user can go directly to observe the patient's record, speed or trends, as shown in Figure 2.

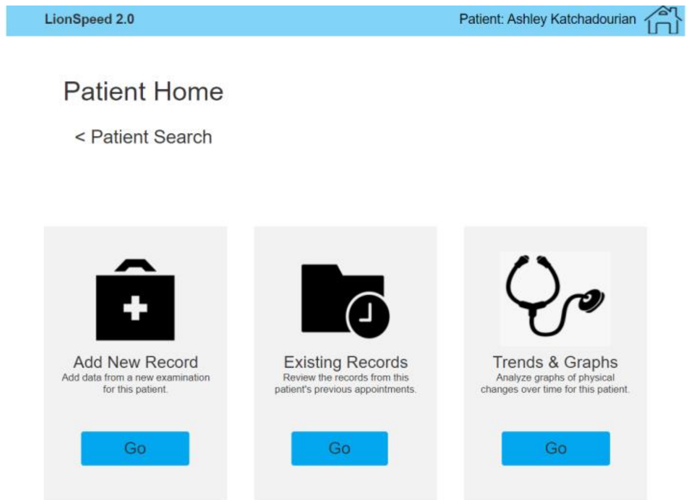


Figure 2: Menu that shows the available options for obtaining patient records

When the user clicks on the Existing Records option, they will be able to observe the different patient date records for that specific patient, as shown in Figure 3. By clicking the "View Record" button, they will observe the specific patient record for that date, as shown in Figure 4. If the user decides to click on "View" for the Speed graph, they will be directed to the menu in Figure 5. Moreover, if the user clicks on "View" for the Fatigue graph, they will be directed to the menu in Figure 6.

Figure 3: Menu with the Patient Existing Records

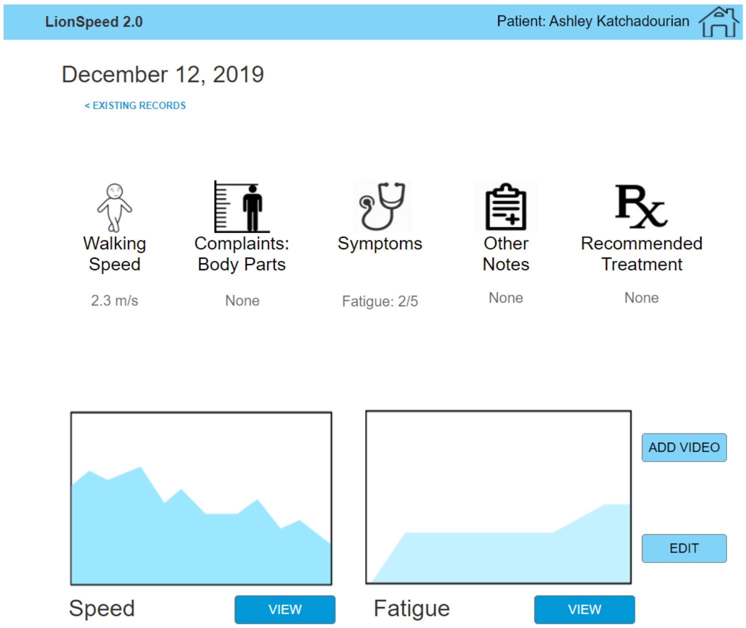


Figure 4: Patient Date Record

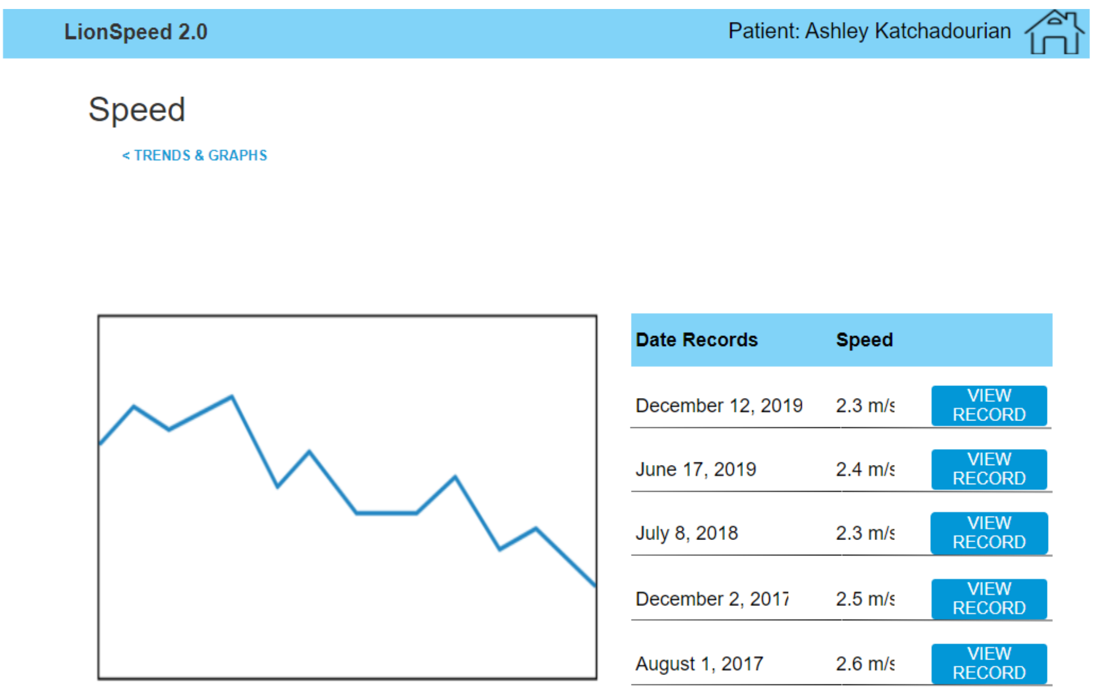


Figure 5: Screen showing the speed of the patient over time

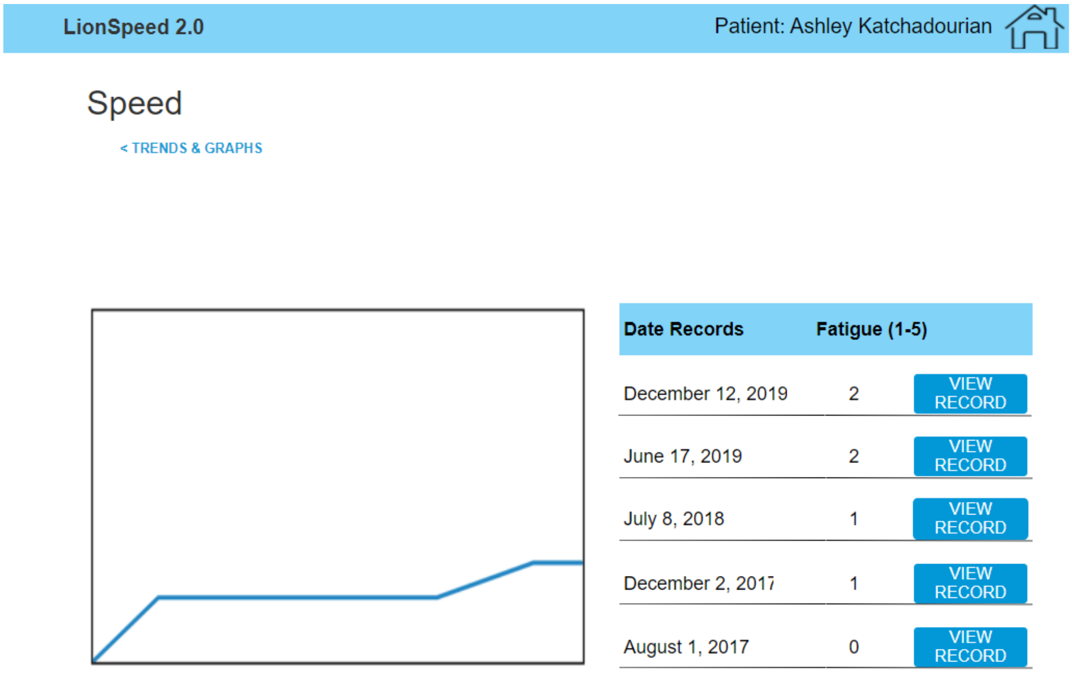


Figure 6:Screen showing the fatigue of the patient over time

If the user clicks on “Trends & Graphs” in Figure 2, they will be directed to the screen in Figure 7. This screen allows the user to observe the patient’s trends over time and determine any patterns with a visual inspection. Moreover, if the user clicks in “View” in the Speed and Fatigue graphs, they observe the menus in Figure 5 and Figure 6, respectively.

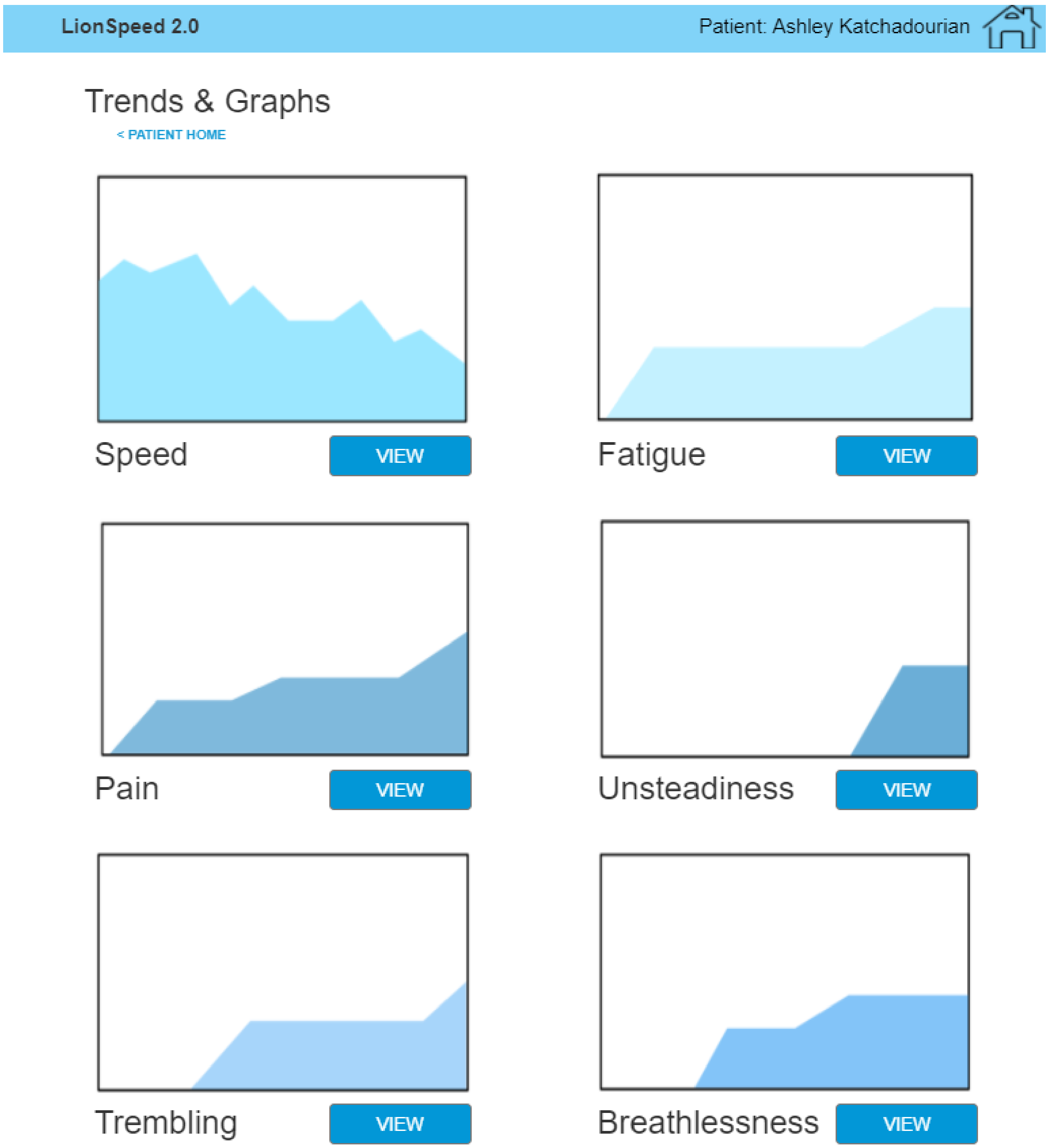


Figure 7: Screen with the patient's trends over time

If the user decides to conduct a new walking speed test for a specific patient, they can create a new record by clicking "Add New Record," as shown in Figure 2. This action will redirect the user to the menu in Figure 8. Then, the user can see the screen in Figure 8 and choose between two options: (1) From Sensor, which allows the user to initiate a walking speed test using the LionSpeed device (via Bluetooth), or (2) From Records, which allows the user to manually enter the speed of a new patient.

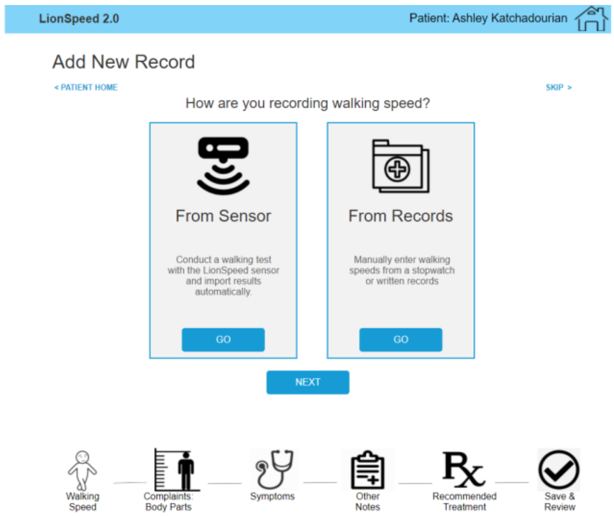


Figure 8: Screen that shows the method to record speed

After the speed is obtained using the LionSpeed device, the user will be redirected to the screen in Figure 9 and the user will be able to record additional information pertaining to other aspects of the patient's gait.

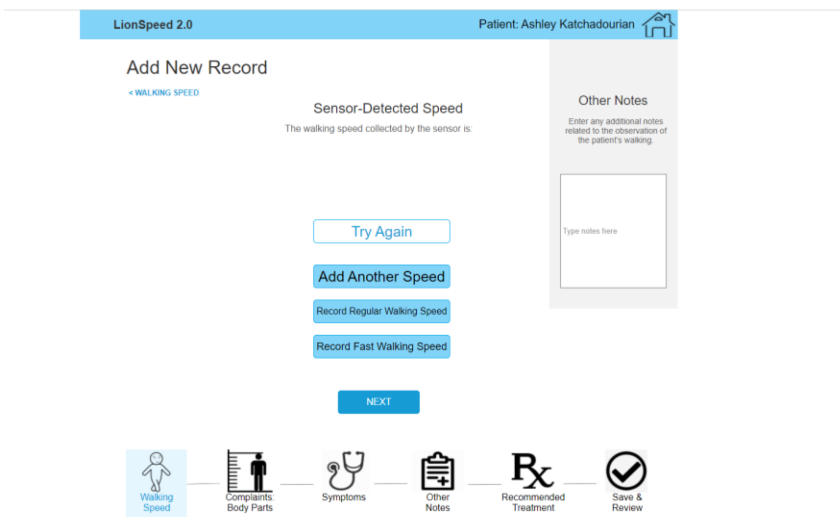


Figure 9: Menu that shows the recorded speed using LionSpeed

First, the user has the option of selecting the body parts that the patient has reported pain in while walking, as shown in Figure 10. Users also have the option of skipping this screen if the information is not relevant to them.

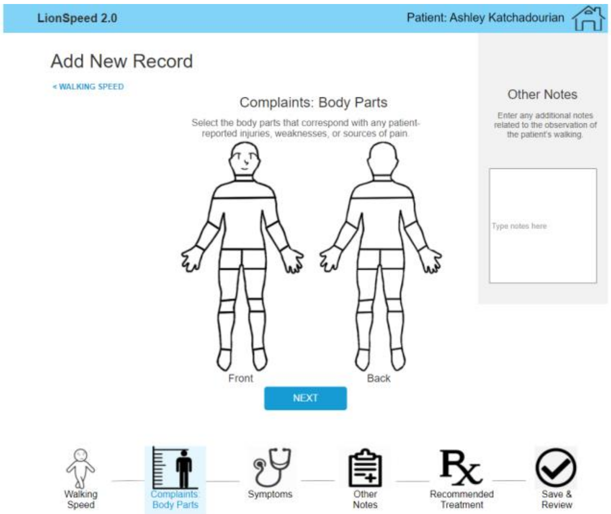


Figure 10: Screen that shows the patient's body parts to indicate if they are hurt or injured

Following this, users will have the option to rate the patient's gait on Likert-like scales based on a number of metrics pre-selected by the user (Figure 11). For example, if a patient reports extreme pain while walking, the user might give a rating of 5 on the pain scale.

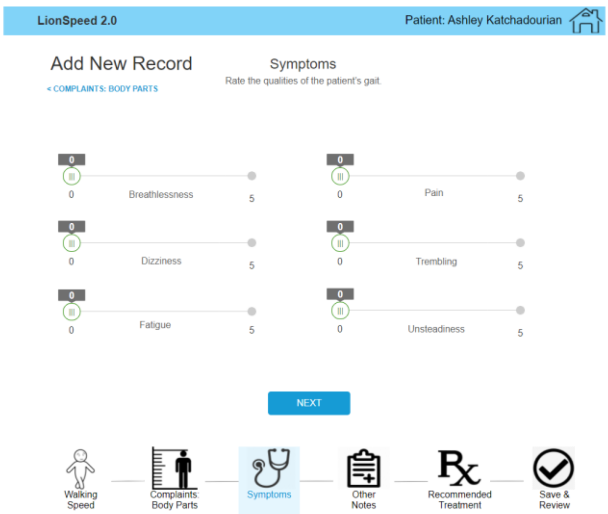


Figure 11: Menu that shows the patient's characteristics and their ratings

Then, as shown in Figure 12 and Figure 13, the user will be able to write any additional notes and the recommended treatment for the patient.

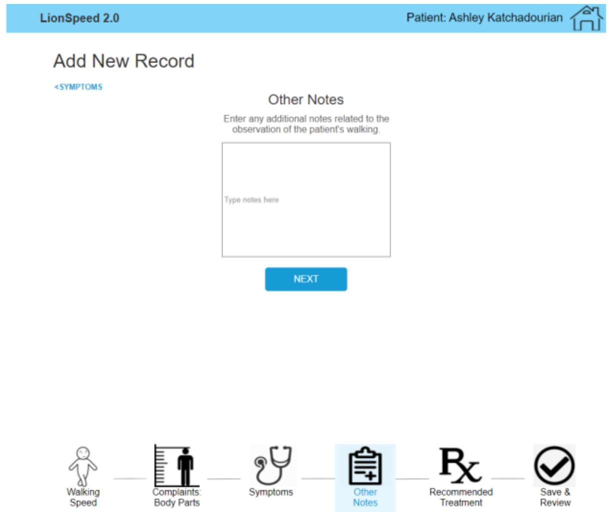


Figure 12: Menu that allows the user to write additional notes

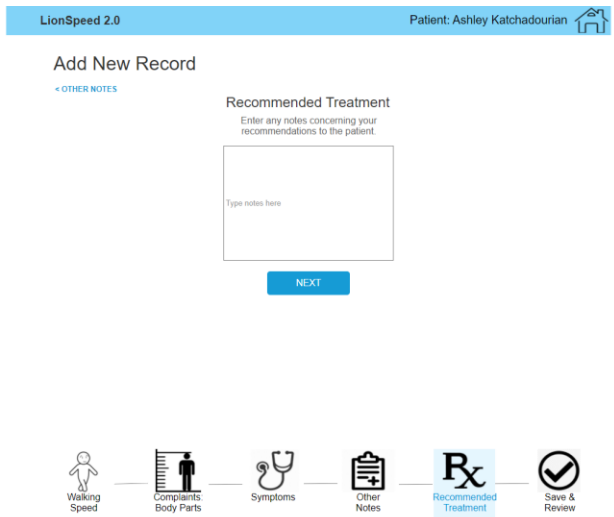


Figure 13: Menu that allows the user to indicate the recommended treatment

Finally, the user will be able to review and save the record by clicking the "Save" button, as shown in Figure 14.

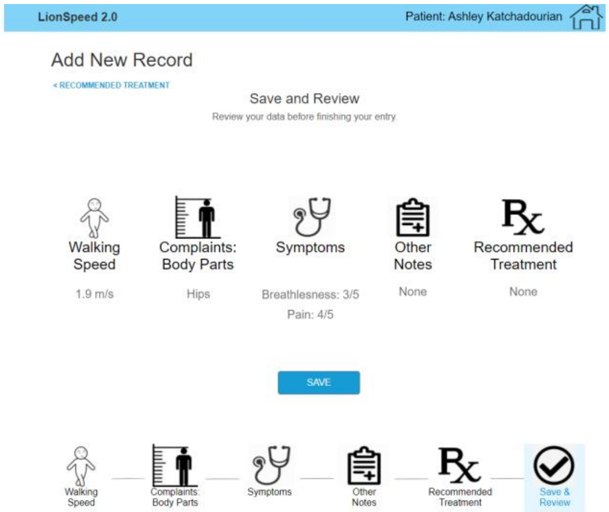


Figure 14: Screen that shows a summary of the data and allows the user to save the data

5.4 Quantitative Metrics

The following quantitative metrics were utilized for both studies conducted:

Time on Task: Prior research has shown that time on task is a valid metric to measure the effectiveness in a usability study (Hornbæk & Law, 2007). To measure how effective were the interfaces for a participant to navigate through them, in both studies we measured the time taken to complete each task.

Number of Errors: Hornbaek & Law (2007) showed how the occurrence of errors correlates with usability metrics like satisfaction, completion rate, and task completion time. Therefore, we measure the number of errors participants had when interacting with both interfaces. We classified an error as being a mistake that is inherent to the interface, rather than a slip-up. As an example, we would count as an error, when the participant repeatedly clicked on a text that was not a button or would produce an action. However, slip-ups like typos and incorrect data entries, that are not inherent to the interface will not be counted as errors.

Number of Clicks: To measure how the number of clicks relate to the time that participants took to complete each task, data from a Recording User Input (RUI) was used.

5.5 Qualitative Metrics

The following qualitative metrics were utilized for both studies conducted:

Verbal Protocol Analysis: Prior research has shown how usability can be measured by asking participants to verbally articulate their thought process when conducting a task (Benbunan-Fich, 2001). Therefore, in these studies we asked participants to "verbally articulate their thought process" when they completed the two different tasks in the interface.

5.6 Surveys

This study also counted with participants completing a pre-survey prior to the study and a post-survey provided at the end of the study.

Pre-Survey: Participants were asked to complete a pre-survey to gather information about their demographics, occupation, and specialty (when they were a medical practitioner). The purpose of this pre-survey was to differentiate results according to the different group of participants.

Post-Survey: Participants were asked to complete a post-survey with the purpose of understanding their satisfaction with the interface. The post-survey is based on the Nielsen's 10 Usability Heuristics that allow us to understand the usability of the interface with valid metrics (Nielsen, 1994). Participants had to indicate their level of satisfaction regarding the duration of time they took to complete the tasks, learnability of interface, and legibility of content. Participants that were medical experts like practitioners or physician assistant, they had to indicate the applicability of the interface in their job, if the interface met their needs, and their likelihood of using it in their job.

5.7 Prototype Testing Procedure

Both studies followed a similar methodology and prototype testing procedure shown in Figure 15. Each usability study was conducted for approximately 25 minutes. Initially, the participants were asked to provide consent to participate in the study, as required by the IRB guidelines and were asked to choose a date available to conduct the usability study. The usability studies were conducted virtually using the Zoom platform. Prior to the study, some background information was provided and consent to record the session was obtained. Participants were initially provided with a link through the Zoom Chat for them to access and complete the Pre-Survey. For the first study, participants were provided with a link through the chat that allowed them to access the interface and they were asked to share their screen during the study. Participants had the task descriptions in their chat for them to reference it as many times as they preferred. However, for the second study, participants took control of the proctor's screen to interact with the interface and had the task descriptions in an email tab for them to reference it as many times as they preferred. Before interacting with the interface, they had to start the recording for a "Recording User Input (RUI)" that provides data about the location of the mouse and clicks pressed when interacting with the interface. A different RUI recording was recorded for each of the two individual tasks.

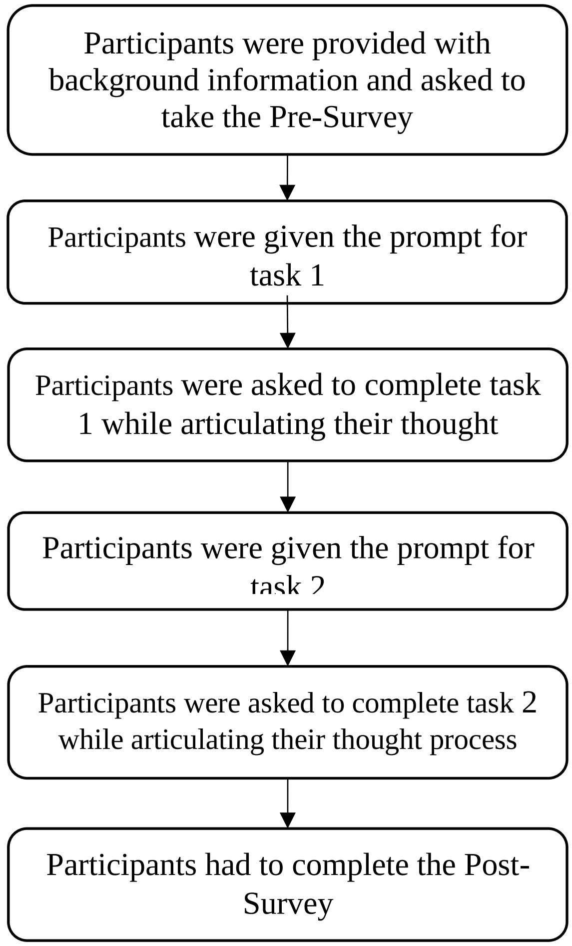


Figure 15: Methodology for Usability Study

The first task had the purpose of allowing the participant to interact with the interface and observe a specific patient, Ashley's existing records and report her last recorded speed and fatigue level. The prompt of the first task is shown below. Participants in were asked to verbally articulate their thought process while completing the task. If the participant went quiet, we prompted the participant to continue articulating their thought process. After participants completed the first task, they were provided with the second task description.

> *"Ashley is an 82-year-old patient who has been complaining about pain in her lower hip and feels like she has been walking slower. She comes into your clinic and wants to get your opinion. Check Ashley's speed and trends for the past year and report her speed and fatigue level recorded during her last visit."*

The second task aimed for the participant to interact with the interface and create a new record for a specific patient, Ashley. Similarly, to the first task participants were asked to verbally articulate their thought process while they complete the task. The prompt of the second task is shown below.

> *"You then conduct a walking speed test with Ashley using the LionSpeed device. Ashley specifies that she has pain in her hip, and is experiencing moderate breathlessness and severe pain while walking. When asked to rate her pain and*

breathlessness on a scale from 1 to 5, she reports a 3 from breathlessness and a 4 for pain. She does not report issues with anything else. Create a new entry for Ashley."

After participants completed both tasks, they were provided with a link in the Zoom chat to access the Post-Survey. Then, the participant was thanked for their time.

6. RESULTS AND DISCUSSION

To answer our research questions, we collected data from nine undergraduate (junior or senior) nursing students, as well as three medical and healthcare professionals for the first iteration (created in Adobe XD) of the interface. For the second iteration of this interface (created in Axure), we collected data from another set of eight undergraduate nursing students, as well as two medical professionals. In the first interface, it took an average of 1.4409 minutes (SD=0.432 minutes) to complete Task 1 and an average of 2.248 minutes (SD= 0.97214 minutes) to complete Task 2. In the second interface, it took an average of 3.91 minutes (SD= 1.915 minutes) to complete Task 1 and an average of 2.858 minutes (SD= 0.571 minutes) to complete Task 2. The statistical data were analyzed via the SPSS v.26. A value of $p < .05$ was used to define statistical significance (Fisher, 1925).

6.1 RQ1: What is the perceived and measured usability of the *current* LionSpeed interface when tested with target users?

The objective of our first research question was to determine the perceived and measured usability of the current LionSpeed interface. Moreover, this research question provides the measured usability of the current LionSpeed interface and compares it to the measured usability of the first version of the LionSpeed interface. To measure the usability, we used three quantifiable metrics; number of errors to measure the effectiveness, time taken to complete each task to measure the efficiency, and the rate of satisfaction collected at the post-survey to measure the satisfaction.

Effectiveness:

To collect the number of errors, the zoom recordings were observed and quantified. When comparing Task 1, we obtained that the first interface had a mean number of errors of 1.91 with standard deviation of 2.879 and for the second interface we had a mean number of errors was 2.20 with a standard deviation of 1.229. When comparing the number of errors for the Task 2, for a first interface we obtained a mean number of errors of 5 with a standard deviation of 4.219 and for the second interface we obtained a mean number of errors of 3.60 with a standard deviation 1.35. A boxplot of these results can be seen in Figure 16 and Figure 17, respectively.

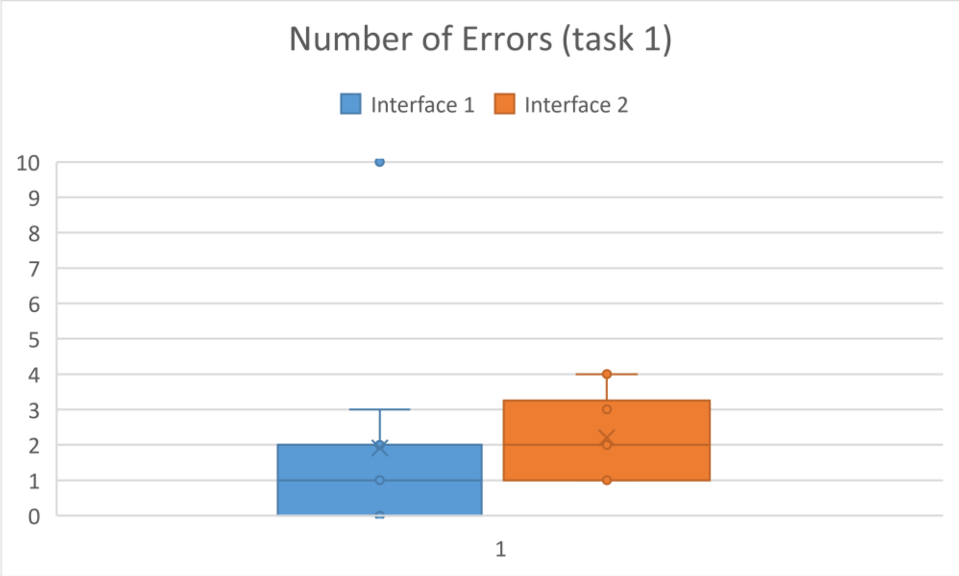


Figure 16: Number of Errors for Task 1

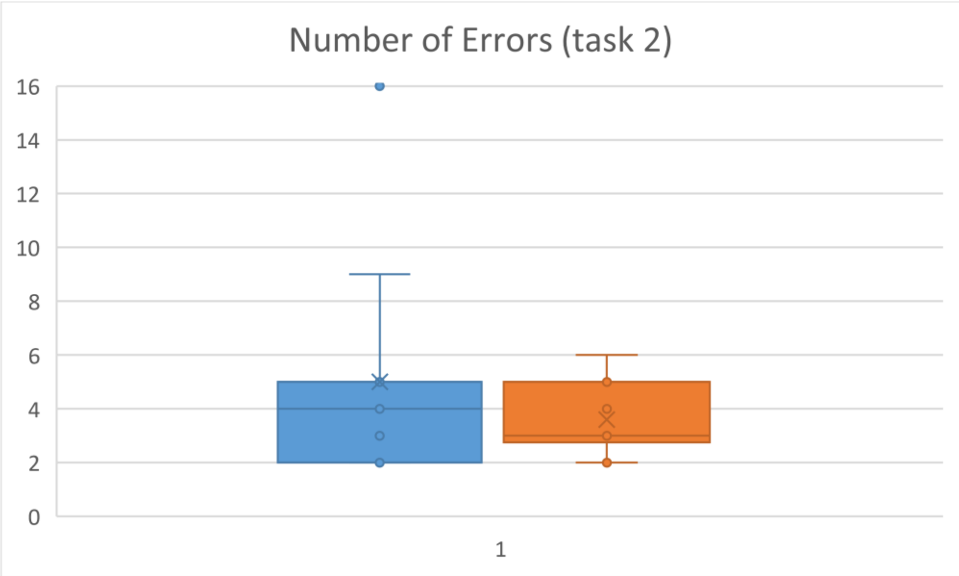


Figure 17: Number of Errors for Task 2

One of the errors that was common and frequent between the two interfaces, was the speed recording screen in Figure 8 during the completion of Task 2. Participants seemed to confuse themselves where they had to record their screen, rather the LionSpeed sensor or manually entry the data. Therefore, participants that made this error click the "From Records" option rather than the "From Sensor" option as the task asked them to do. Another frequent error observed in Interface 2 was in Figure 9, where participants seemed to get confused by the multiple options available to record the walking speed. When asked as to why this screen caused them confusion, they mentioned that they expected the speed to be automatically displayed rather than them clicking or choosing one of the options available.

Another frequent error in the second interface was in Figure 1 since participants seemed to click the name of the participant rather than clicking the "View" option available. It seems that participants were expecting that by clicking the name of the patient in the search box, they would be able to access the patient's record.

A paired t-test was conducted between both groups of participants to compare if there were significant differences in the number of errors depending on the type of interface. Results from this paired t-test showed that there was not a statistical difference in number of errors between both interfaces when completing either Task 1 or Task 2. Participants in the Interface 2 study made more errors in Task 1 (M = 2.20, SD = 1.229), as opposed to the number of errors in the Interface 1 study (M = 1.80, SD = 3.011); although this was not statistically significant. Participants in the interface 1 made a greater number of errors in Task 2 as well (M = 5.30, SD = 4.322) as opposed to the Interface 2 (M = 3.60, SD = 0.427), failing to establish statistical significance.

Efficiency:

To collect the time taken to complete each of the two tasks, the recordings were observed and the time where each participant started and acknowledge that they finished with the task was quantified. When comparing Task 1, we obtained that the first interface had a mean time to complete of 1.4409 minutes with standard deviation of 0.432 minutes and for the second interface we had a mean time to complete was 3.91 minutes with a standard deviation of 1.915 minutes. When comparing the time to complete for the Task 2, for a first interface we obtained a mean time to complete of 2.248 minutes with a standard deviation of 0.97214 minutes and for the second interface we obtained a mean time to complete of 2.858 minutes with a standard deviation 0.571 minutes. A boxplot of these results can be seen in Figure 18 and Figure 19, respectively.

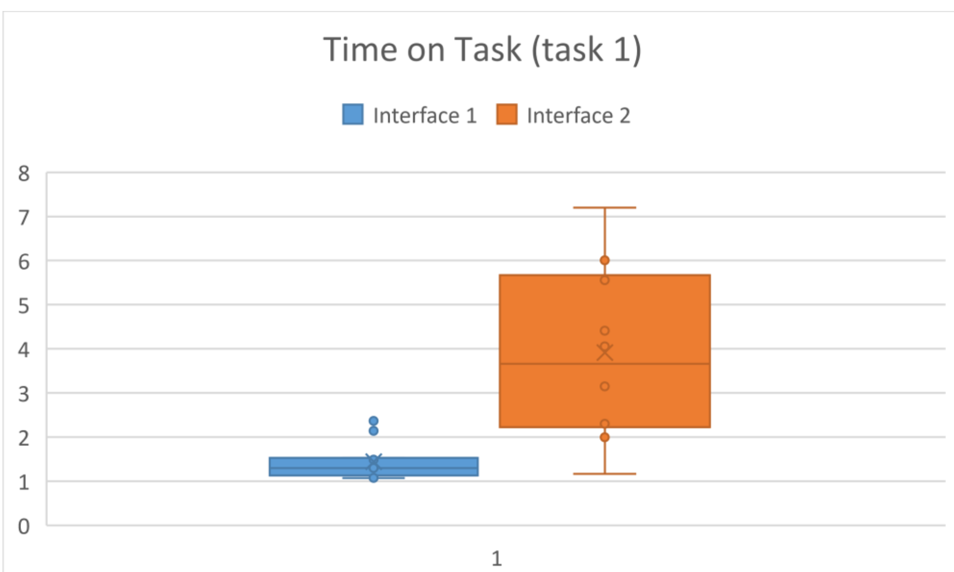


Figure 18: Time on Task 1

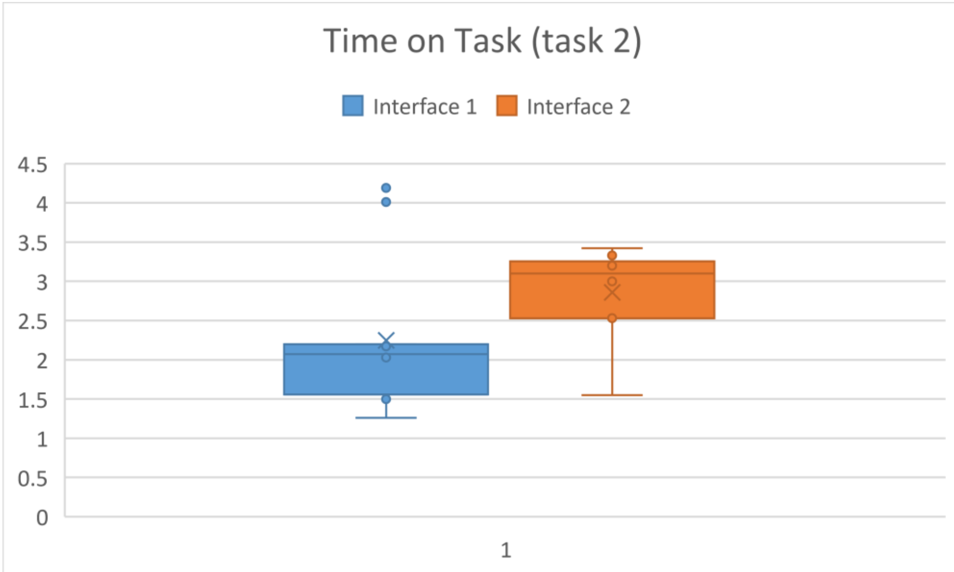


Figure 19: Time on Task 2

A paired T-Test was conducted between both groups of participants to compare if there were significant differences in the time taken to complete each task, depending on the type of interface. Results from this paired T-Test showed that there was a statistically significant difference in time taken to complete Task 1 between both interfaces. Participants in the Interface 2 study spent more time performing Task 1 (M = 3.9120, SD = 1.9142) as opposed to the Interface 1 study participants (M = 1.4320, SD = 0.4546), a statistically significant mean increase of 2.48, 95% CI [1.1825, 3.77746], $t(9) = 4.324$, $p < .05$. However, results showed that there was not a statistically significant difference in time taken to complete Task 2 between both interfaces. Participants in the Interface 2 study took more time to complete Task 2 (M = 2.8580, SD = 0.1804) as opposed to the Interface 1 participants (M = 2.3470, SD = 0.3051); failing to establish statistical significance.

Satisfaction:

The satisfaction was measured by the ratings that participants provided in the Post-Survey. Participants had to rate their overall satisfaction on a 7-point Likert-scale that ranged from "Extremely Dissatisfied" to "Extremely Satisfied". For the first interface, 8 participants indicated that they were Extremely Satisfied with the interface, while 4 participants indicated that they were Moderately Satisfied with the interface. Participants were asked to justify their ratings, and results from this are the following:

> *"Very user friendly, I think a nurse of medical assistant could obtain and enter most data, with the physician reviewing and making recommendations....an additional "vital sign" or more formal way to track components of the geriatric assessment." – Doctor at the Hershey Medical Center*
>
> *"I thought the design on the interface was very pleasing and simple, which made it easy to navigate through. Even when I wasn't 100% sure what to do, I still was able to back track and figure it out, without direction from the study leaders. Overall, I really think this interface could be of use to many people." – Nursing student at the Pennsylvania State University*

For the second interface, 2 participants indicated that they were Extremely Satisfied with the interface, while 4 participants said they were Moderately Satisfied, 2 participants said they were Slightly Satisfied, 1 participant said they were Slightly Dissatisfied, and 1 participant said that they were Moderately Dissatisfied. When participants were asked to justify their ratings, they mentioned the following:

> *"The interface was very easy to navigate. Things were clearly marked and even when I made a mistake it was easy to correct. Everything was relatively easy to find and the tasks were very easy to complete. It was very clear and obvious what to click next and to figure out how to work the interface."- Nursing student at The Pennsylvania State University*

> *"I would say that I found it difficult without any training prior to using the interface to navigate it. The place where you add data maybe could have been better as a list laid out instead of a sequence of entries you have to move through. It is easier to see all the options in front of you that you can move through on one page versus seeing the area of pain then hitting next to the ratings. I did like the actual look of the interface. I do think once trained on the interface it would be easy to use but overall there could be improvements to make it more efficient when filling out new records."- Nursing student at The Pennsylvania State University*

> *"Easy to make adjustments, view entry, and navigate different tasks."-Physician's Assistant at University of Pittsburgh Medical Center*

A graph showing a comparison of the overall level of satisfaction between the two interfaces is shown in Figure 20.

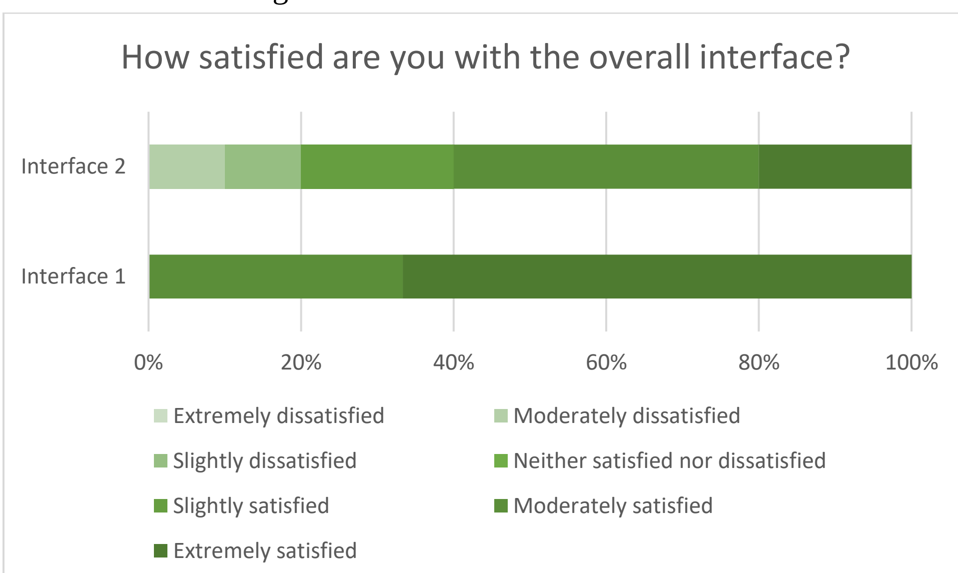


Figure 20: Overall satisfaction with both Interfaces

To verify if there were significant differences in the level of satisfaction depending on the type of interface, a Mann-Whitney U-Test was conducted between both groups of participants. A Mann-Whitney U test was run to determine if there were differences in overall satisfaction ratings score between interface 1 and interface 2. Distributions of the engagement scores for both interfaces were similar, as assessed by visual inspection. Overall satisfaction ratings score was statistically significantly different between interface

1 (Mdn = 7) and interface 2 (Mdn = 6), U = 24, z = -2.561, $p < .05$, using an exact sampling distribution for U (Dineen & Blakesley, 1973). These result show that overall satisfaction with the first interface was higher than the second interface.

The post-survey provided us with more information about the level of satisfaction of each participant according to the Nielsen's usability heuristics (Nielsen, 1994). The purpose of these satisfaction statements was to identify which aspects of the interface the users were more satisfied with. Results of the satisfaction with each statement for both interfaces are in Figure 21. Moreover, the participants were also asked to provide their ratings when asked to indicate the level of difficulty when interacting with the interface considering the ability of correcting mistakes, learnability of the interface, and legibility of text on the screen. Results of the participants ratings for both interfaces is shown in Figure 22.

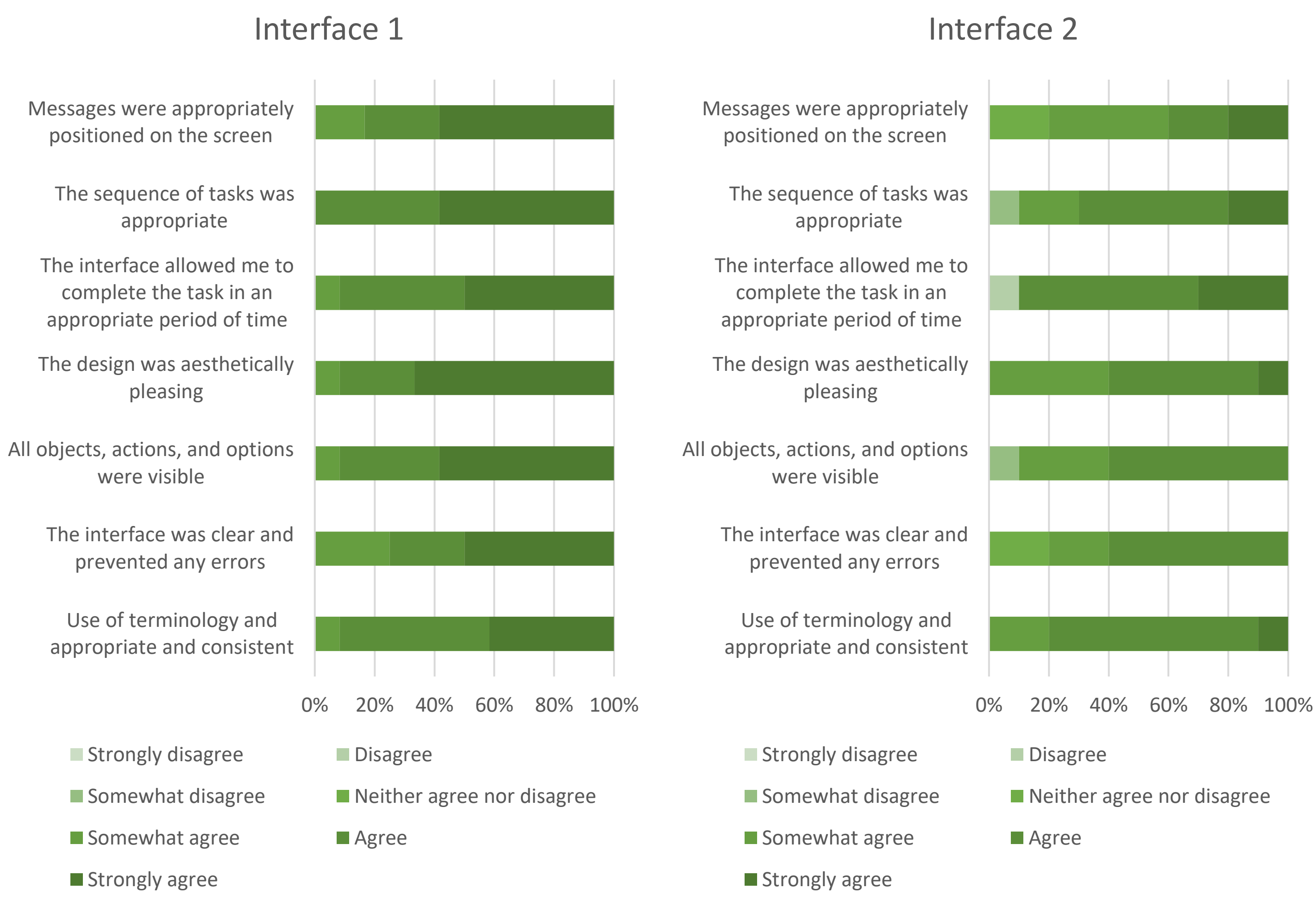


Figure 21: Participant indicated Extent of Agreement with Usability Statements

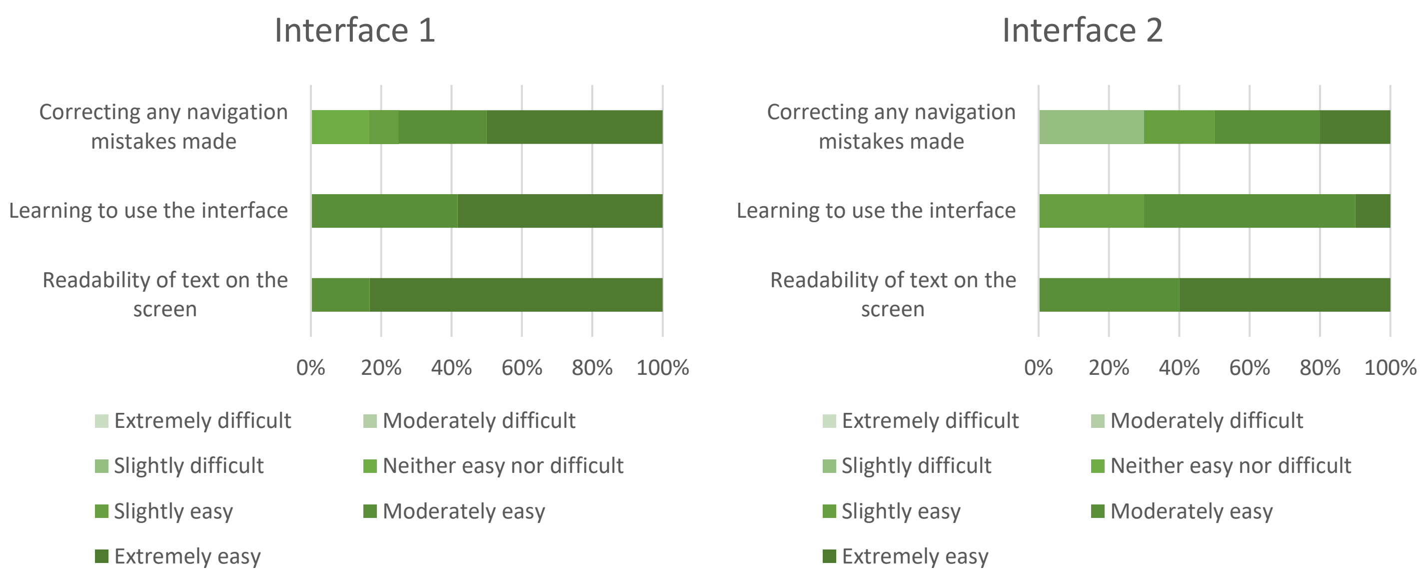


Figure 22: Participants' Level of Difficulty in Interacting with the Interface

6.2 RQ2: How does the time on task relate to number of clicks in task?

The second research question looked at how the number of clicks of each participant in each of the two tasks relate to the time in each task. Taking the data from the Recording User Input (RUI), we used the total number of clicks in each of the two tasks and graphed them together, as shown in Figure 23. The average number of clicks on Task 1 resulted in an average of 70.5 clicks with standard deviation of 42.11 clicks, with a total sum of 705 clicks. For Task 2, the average number of clicks was 45.7 clicks with a standard deviation of 5.56 clicks, and a total sum of 457 clicks. This shows that the participants had a greater number of clicks of Task 1 when compared to Task 2. This may be due to the participants having never seen nor interacted with the interface prior to the first task, therefore they had a greater number of clicks as a result of following a learning curve.

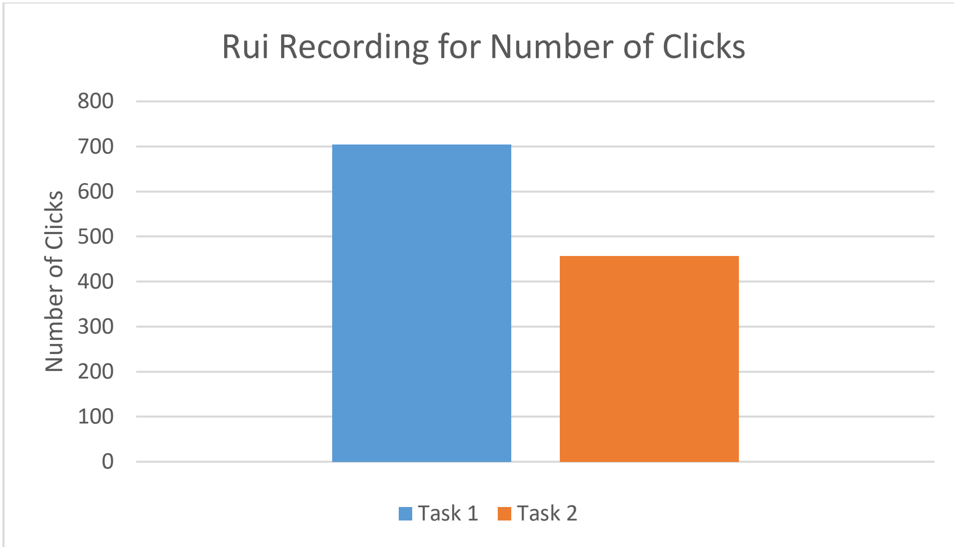


Figure 23: Graph of total number of clicks per each task

The RUI provides data about regarding when the participant pressed a key, moved, and the number of clicks to conduct an action. The total number of counts per key, clicks, and movements are graphed in Figure 24. The participants had a higher number of movements when compared to key and clicks. This can be due to the participant moving around through the interface since to find the different keys they should press to complete each of the tasks. The difference between movement and clicks and keys could be due to the learning curve participants experienced when interacting with the interface (Ritter & Schooler, 2001).

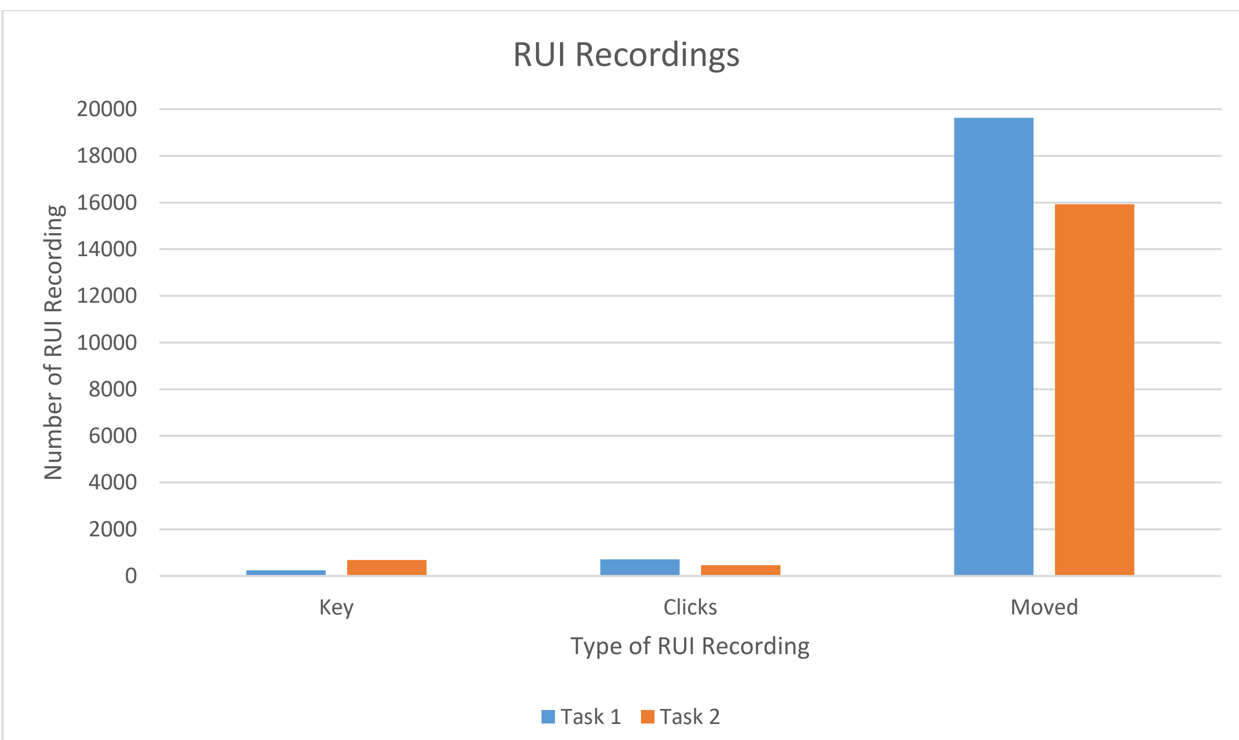


Figure 24: Number of clicks, key, and movements

The number of clicks that the participant took to complete each task was plotted related to the time that the participants took to complete each task as shown in Figure 25. We can see that for Task 1, the participants had a more variable number of clicks across participants and participants tended to take more time to complete the task. When compared to Task 2, due to the learning curve that participants had experienced (Ritter & Schooler, 2001), they had a more consistent number of clicks across participants and tended to take less time to complete the task. These findings support our findings from RQ1.

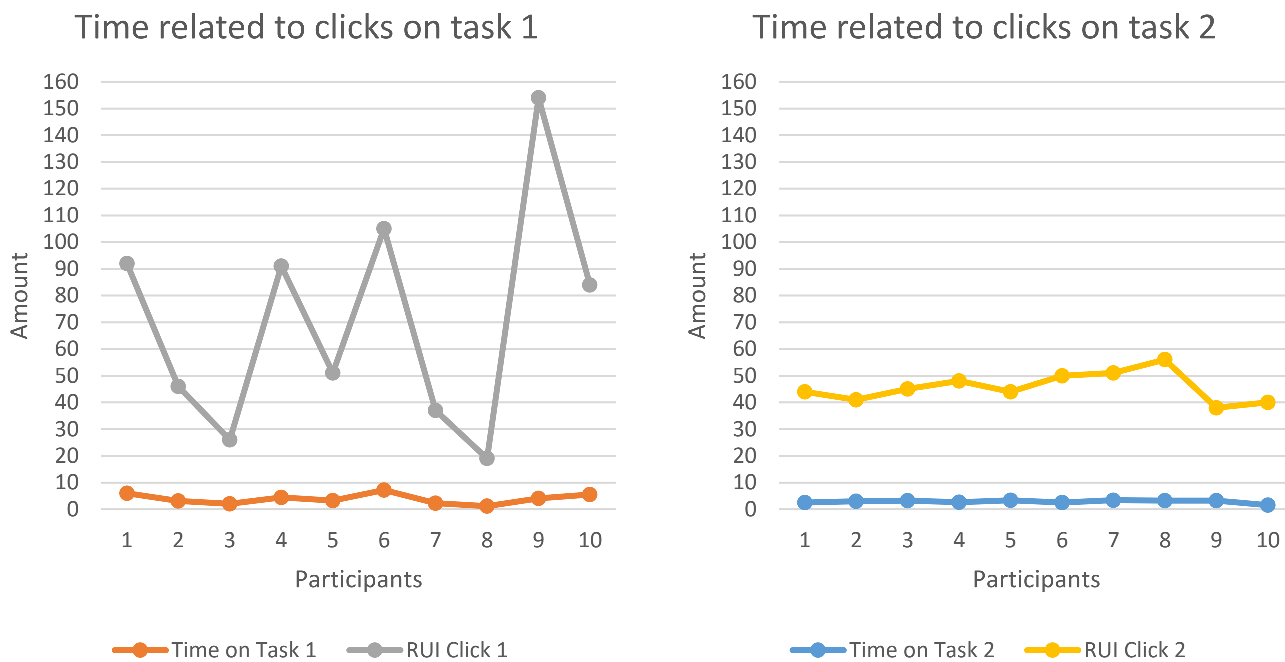


Figure 25: Time related to number of clicks

6.3 RQ3: What are the factors that impact positive and negative interactions with the interface?

The objective of our third research question was to identify positive and negative factors regarding the Axure interface. To answer this question, we conducted a qualitative content analysis with principles of deductive content analysis (Hsieh & Shannon, 2005) with a verbal protocol analysis, having participants follow a think aloud procedure (Ericsson & Simon, 1980). Specifically, we analyzed the verbal protocol analysis to identify general factors that might be positive or negative based on satisfaction metrics such as *ease of use* and *attitudes towards content* (Hornbæk, 2006). Initially a codebook was created with a total of 3 main high-level themes and a total of 5 sub-level themes, and in each of these sub-levels a positive and negative category was created. The first and second level themes were obtained by Hornbæk (2006) as validated metrics to measure satisfaction. To view the full codebook, see Appendix D. To ensure the interrater reliability, two raters, who are both PhD Industrial Engineering students coded using the same codebook at an overlap of 20% (MacPhail, Khoza, Abler, & Ranganathan, 2016) of the verbal protocol analysis data (VPA) using Nvivo Pro. An interrater reliability (Cohen's Kappa) of 0.79 was reached, which was considered acceptable (Landis & Koch, 1977).

The discussion topic that was most frequently discussed by the participants was Users' Attitudes and Perceptions (f=122), Satisfaction with the interface (f=96), and Others (f=11). The number of coding for each of these topics per each participant is shown in Figure 26, Figure 27, and Figure 28, respectively. Therefore, we can conclude that based on frequency, the interface needs improvements in the Users' Attitudes and Perceptions and Satisfaction with the Interface categories. Thus, looking into what made those categories have higher frequency, we evaluated the positive and negative sentiment of each coding category.

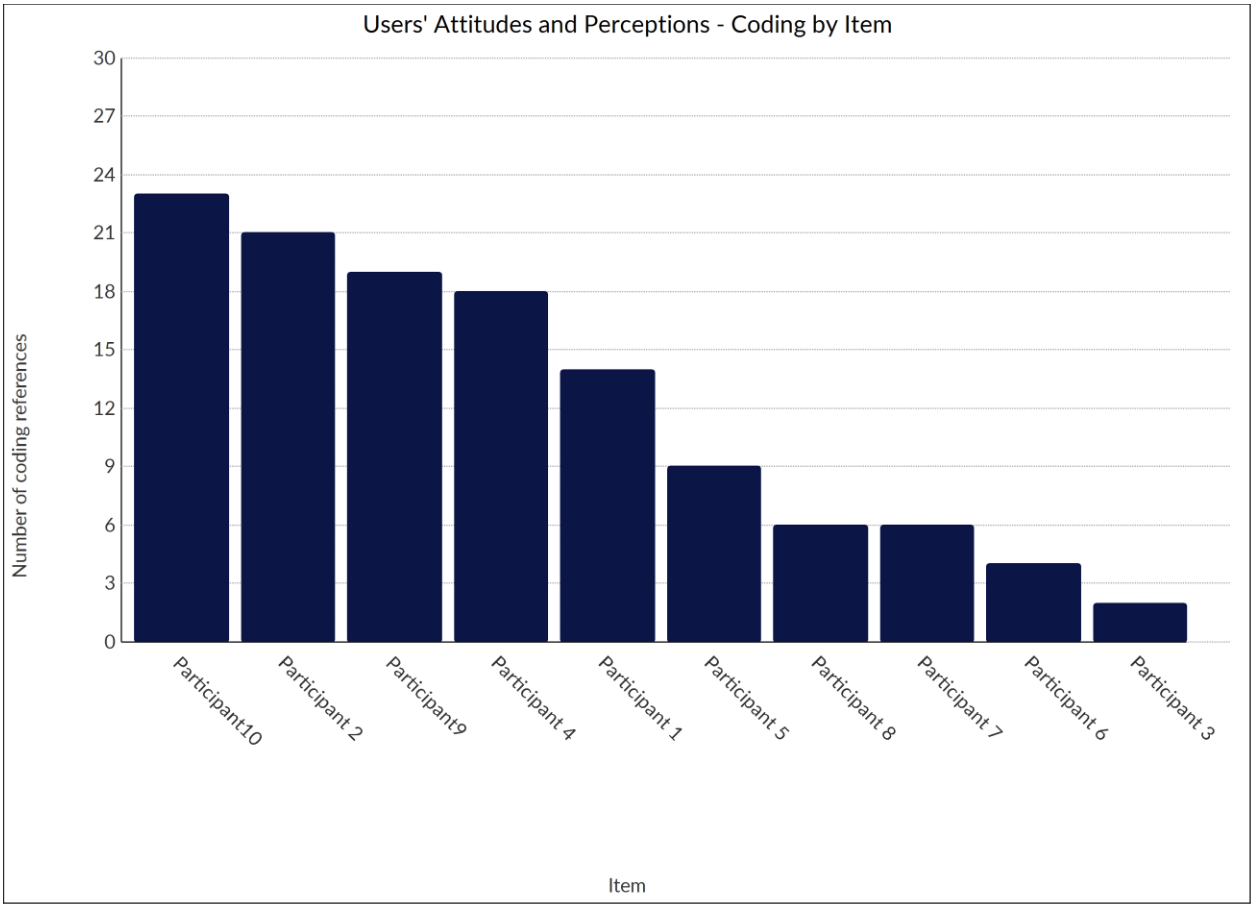


Figure 26: Frequency for Users' Attitudes and Perceptions

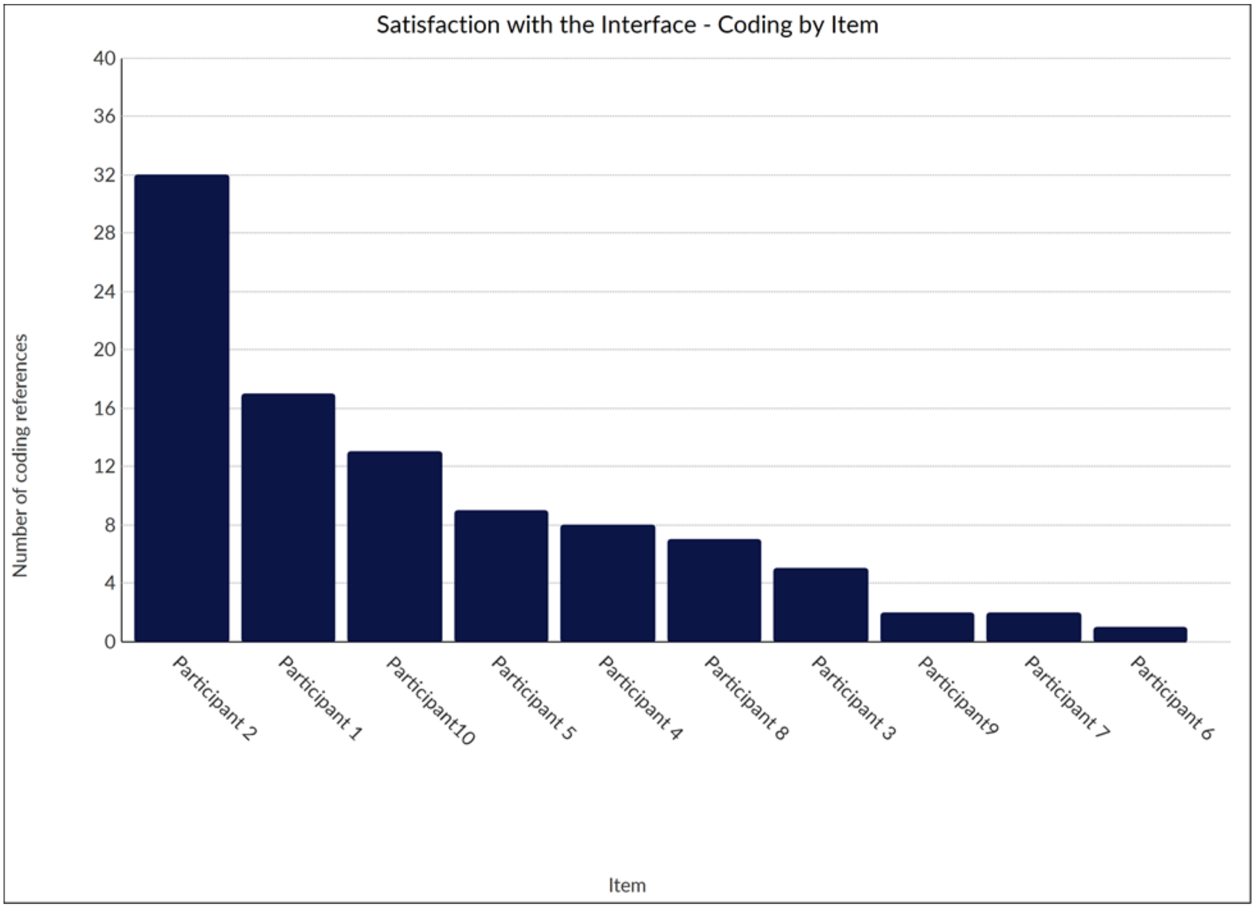


Figure 27: Frequency for Satisfaction with the Interface

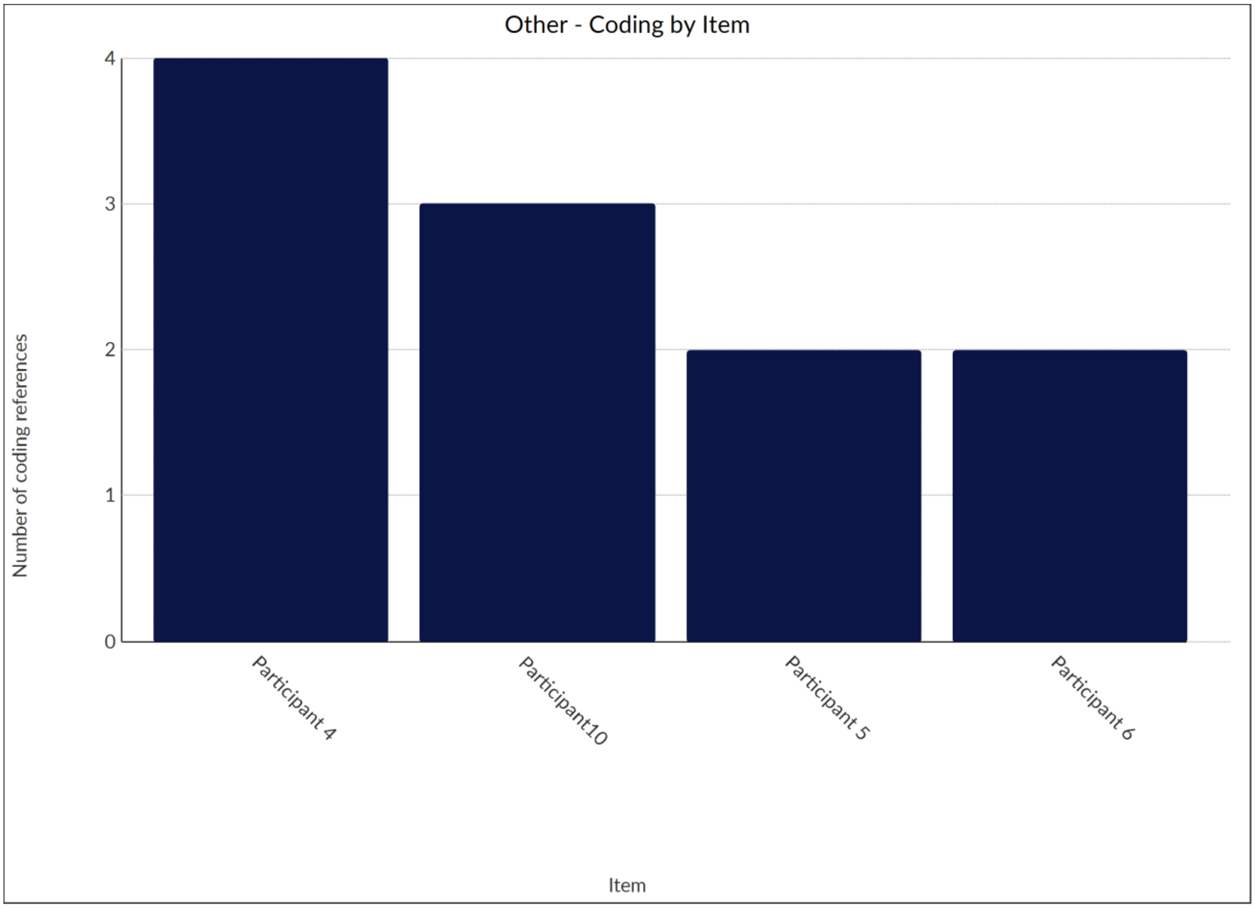


Figure 28: Frequency for Others

Knowing that there was a high frequency for the Users' Attitudes and Perceptions and Satisfaction with the Interface, we wanted to compare them in terms of positive and negative sentiment to evaluate how was the sentiment related to each category. Therefore, to compare the positive and negative sentiment in each of the coding categories, a graph was created in NVivo as seen in Figure 29. Specifically, we also coded the number of references into each of the sentiment for each of the codes can be seen in Figure 30: Users' Attitudes and Perceptions Sentiment Coding, **Error! Reference source not found.**, and Figure 32 to see the frequency for each level of coding.

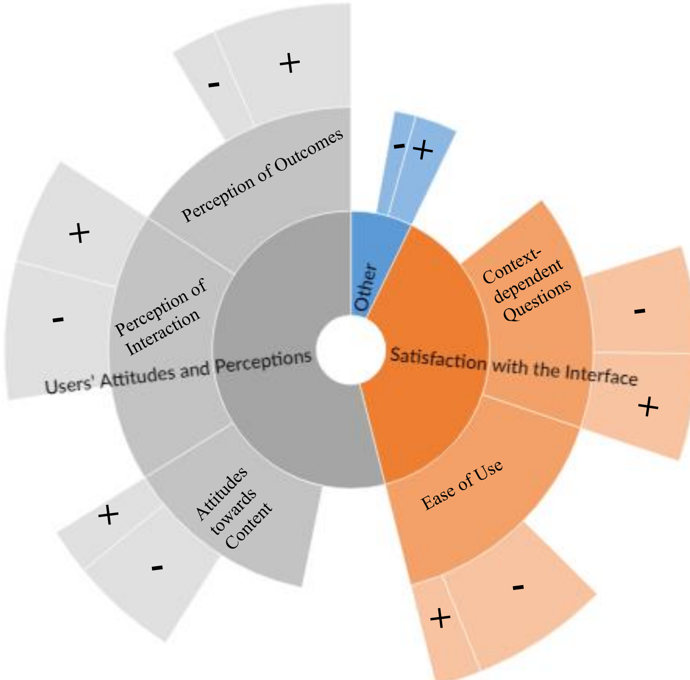


Figure 29: Sentiment Coding

As shown in Figure 30: Users' Attitudes and Perceptions Sentiment Coding, there were some negative and positive codes for "Users' Attitudes and Perceptions." Specifically, in relation to perception of the interface, there were more negative sentiment comments than positive sentiment comments. For "Perception of Outcomes" and "Attitudes Towards Content," there were approximately the same quantity for positive and negative sentiment comments.

To improve our interface, we evaluated the negative sentiment which is related to users clicking the hips in the Figure 10, confusion with who the speed is being recorded in Figure 8 and Figure 9, and looking for the patient name in the search box for Figure 1. These can be seen in some of the following expressions:

> *"So I'm clicking this because the pain is in her lower hip, but I don't know why I can't click any other side?"*
>
> *"How do I go to like conduct the tests?"*
>
> *"I'm trying to like kind of click her name."*

When observing the features that caused negative sentiment in the perception of outcomes, we found that users had confusion with the hips in Figure 10 and recording the speed in Figure 9. Some of the participants' expressions specifically mentioned the following:

> *"Okay, I can only click one area it seems like."*
>
> *"Oh, so I can skip the walking speed, did she get a walking speed?"*

Finally, when observing the negative sentiment for attitudes towards content, some participants expressed negative sentiment towards the physical characteristics graphs in Figure 7, the recording the speed with the sensor in Figure 8 and Figure 9, and hips in Figure 10.

"The graphs for, say, if I select it, I'm selecting just the view of the speed graph, and it looks like the data points or graphs just aren't labeled, which it would be something I think that would be beneficial."

"It does not say which hip so I guess I would put."

"Um, am I recording from the sensor?"

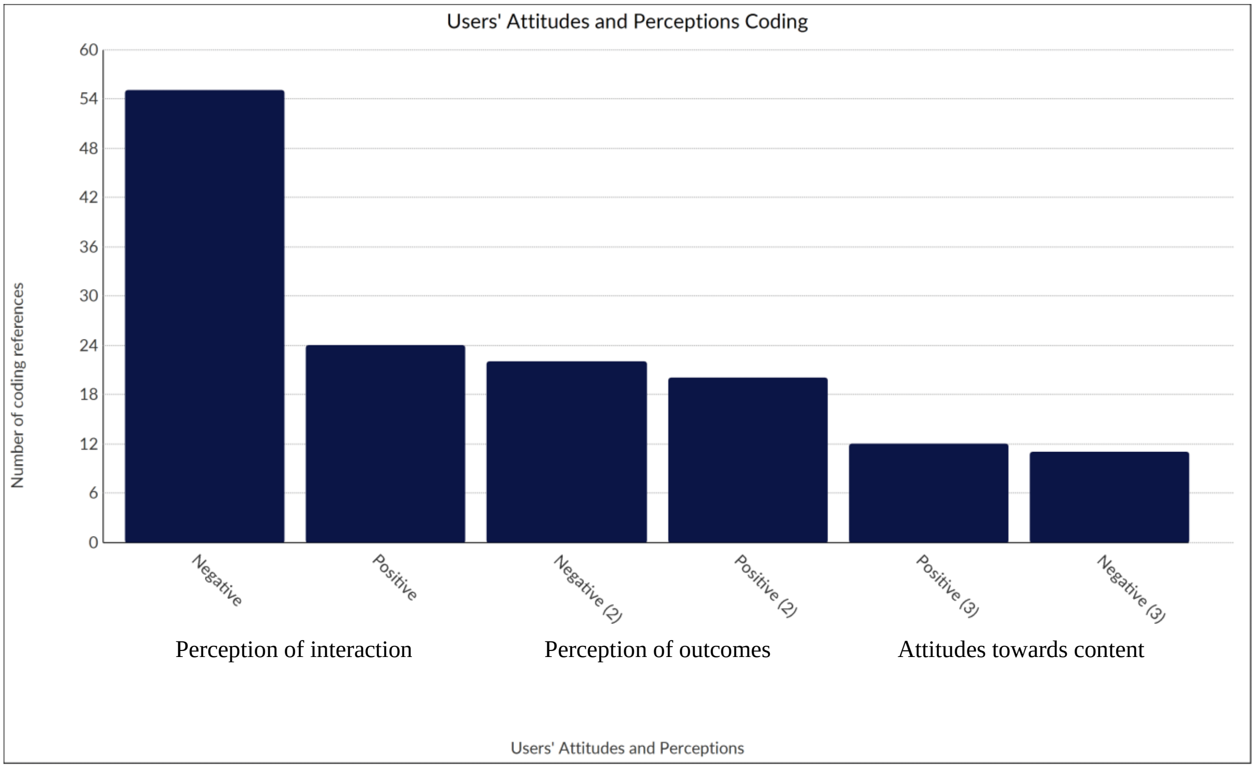


Figure 30: Users' Attitudes and Perceptions Sentiment Coding

Participants expressed negative and positive sentiments for "Ease of Use" and "Context-dependent Questions" in when related to satisfaction with the interface. When identifying the negative sentiment for "Ease of Use," participants mentioned that they were confused with the search box in Figure 1, identifying which patient record they were looking at, clicking the hips in the Figure 10, confusion in the process of conducting a walking speed test, comparing actual medical record with reviewing past records, and feelings of not knowing which is the next step in the interface.

"I'm trying to like kind of click her name."

"How do I go to like conduct the tests?"

"I don't know what I did before."

"So you don't really know how to get to the record to compare it from her trend from last year."

"I feel like I'm kind of stuck at this point."

For the "Context-dependent Questions," participants expressed negative sentiment related to the recommended treatments the patient should have, type of hip in which the patient is expressing pain, and recording the speed.

"Any, should I put any recommended treatments in or is that just like an option?"

"It doesn't...pain in her lower hip, it doesn't say which."

"I don't know if I want it back here to do like a fast-walking speed too, I don't know if it worked."

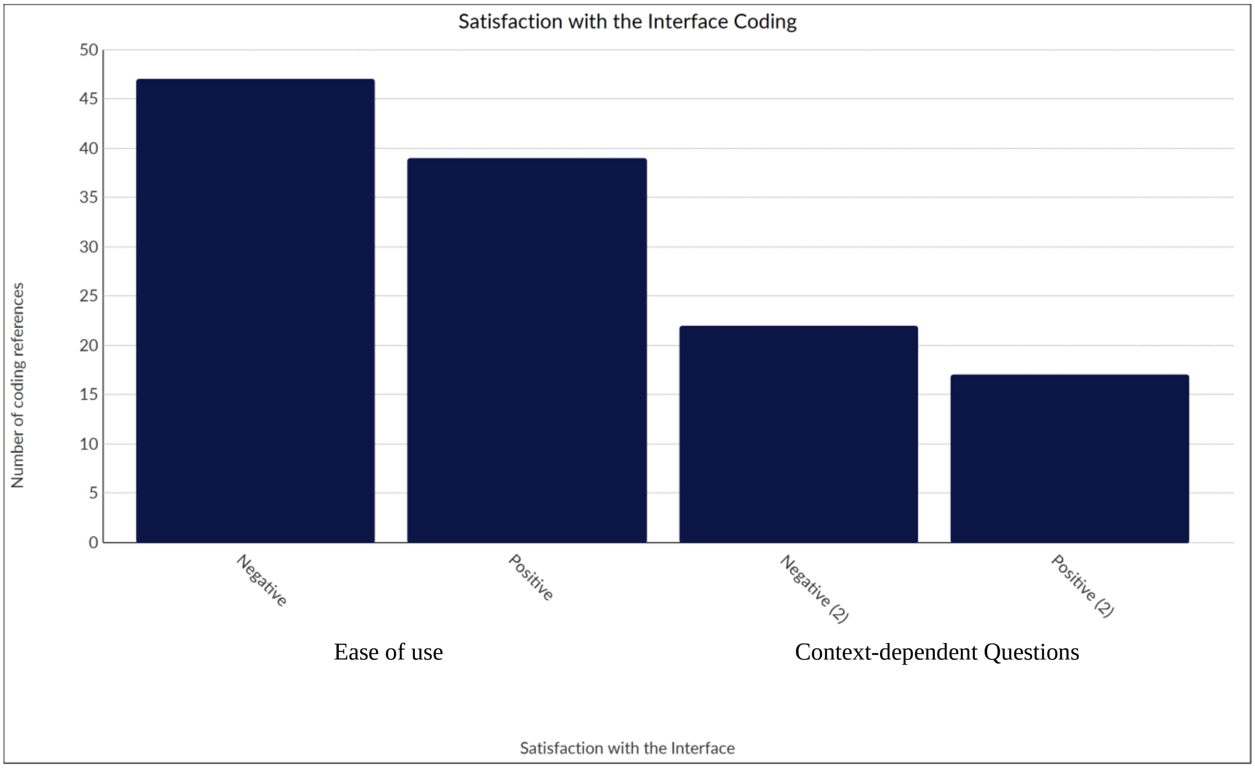


Figure 31: Satisfaction with the Interface Sentiment Coding

In the last category, Other, participants expressed positive and negative sentiment with the usability study. The negative sentiments where related to confusion about what they should do and the Zoom interface.

"Okay, I don't know, oh, yeah, we'll do that."

"I just keep moving our zoom faces around so I can see the screen."

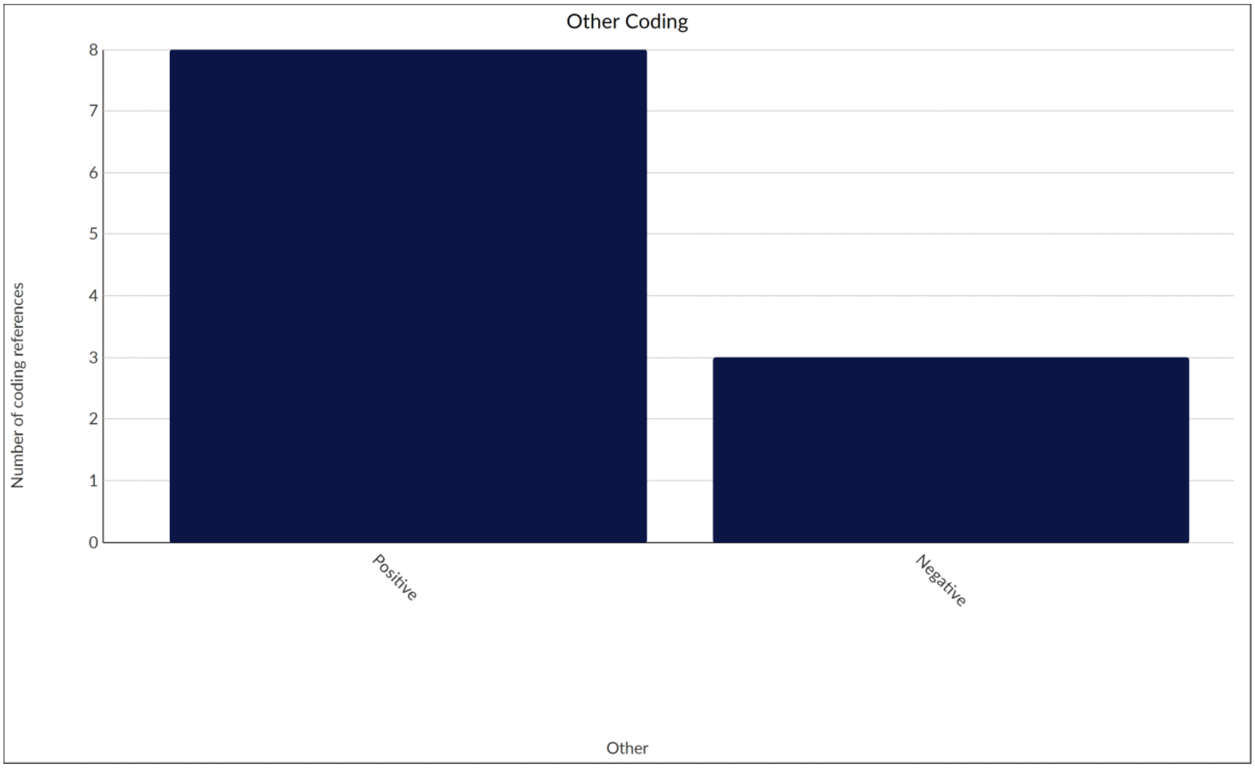


Figure 32: Other Sentiment Coding

With the data collected from the negative sentiments and frequency, we can conclude that the interface has areas of improvement to better accommodate the users' needs. Major negative sentiment coding was related to Perception of the interactions and Ease of use. Therefore, improvements in these two areas should be made to improve the interface. The results from this research question provides insight into what are those areas of improvements in the interface. For future work, it is recommended to improve the interface based on those negative sentiments to provide a better interface for users to interact with.

7. STUDY LIMITATIONS, CONCLUSION, AND FUTURE WORK

This study aimed to analyze the effectiveness of an interface created to measure and track walking speed tests using usability metrics from previous literature (Hornbæk, 2006). Specifically, we analyzed two iterations of an interface and compared them with various usability metrics. The key findings from our study are as follows:

- More time was spent performing Task 1 in Interface 2 in comparison to Interface 1, but there was not a significant difference between both interfaces in Task 2.
- There were no differences in the number of errors made in both tasks between both interfaces.
- Overall satisfaction was higher for Interface 1 in comparison to Interface 2.
- RUI data shows greater variance in the number of clicks among participants for Task 1, whereas Task 2 had a more equal balance across participants.
- Sentiment coding shows that many negative comments were related to the perception of interactions, as well as the ease of use.

Based on the findings mentioned above, several implications can be drawn. First, while no significant difference was found between the numbers of errors between both interfaces for both tasks, more time was spent completing Task 1 in Interface 2 in comparison to Interface 1, but no significant difference in time to complete the task was shown in Task 2. This could be attributed to a learning curve (Ritter & Schooler, 2001), where the participants were able to become faster with practice from performing Task 1, familiarizing them with the interface. This could also be seen as a limitation to our testing method, where we should have either alternated which task each participant started with first, or all participants should have been trained on the interface in all studies to decrease the impact of this confounding factor.

Another implication drawn from this study is that while we included features that were requested from participants during the study for Interface 1, such as a search function for looking up patients, as well as including multiple options to record walking speed (through a "regular walking speed" option and a "fast walking speed option") that was requested by one of the doctors we were working with, these tended to confuse the participants even more. Specifically, participants either didn't understand why the search function wasn't automatically filling in the name of the patient, and didn't know which walking speed button to click. While the issue of not knowing which walking speed button to click could be a limitation of participants' knowledge of walking speed, which requires both a regular pace recording and fast pace recording (Middleton et al., 2016), the confusion with the search function could just be a limitation of Axure itself and getting options to automatically fill in as soon as a user interacts with the prototype of the interface. While it is possible to get options to automatically fill in as shown in some tutorials on the Internet, the researchers in this study were not able to get this to work, which ended up causing some confusion for participants in the end. For these reasons, satisfaction may have dropped in the second iteration of the interface due to such issues causing confusion for participants. A lesson learned from this is to run a smaller pilot study with unpaid participants willing to help with figuring out what needs debugged before running a full usability study with paid participants.

In addition to the decrease in satisfaction that could be due to issues with newly added features, running RUI during usability studies can be a good way to check whether findings from usability metrics make sense. For example, when we ran RUI during Task 1 for Interface 2, we found that both the number of clicks and time to complete the task were much more variable and typically higher than Task 2, which makes sense when compared with our findings that Task 1 took a significantly longer amount of time to complete in comparison to Task 1 from Interface 1, but differences were not apparent in Task 2 between both interfaces. The very high number of clicks illustrates that participants tended to spend more time clicking things in Task 1, whereas this dropped and stabilized across all participants for Task 2, also pointing to the learning curve that helped participants become more efficient (Ritter & Schooler, 2001), particularly in going from Task 1 to Task 2.

Along with RUI data, sentiment coding showed that participants seemed to be more frustrated with the usability of the interface in the form of the perception of interactions and the ease of use. These findings complement our empirical results from the satisfaction survey, which showed a decrease in overall satisfaction. These results convey that making the interface easier to understand and providing something that would guide users through the interface based on what they want to do by making the effect of the functions more explicit and easier to interpret their meanings would be of great importance in the next iteration.

While incorporating the feedback from the latest iteration of this interface in the form of the results, future work would involve having another team with experience in electrical engineering and computer programming program a full version of this interface to be paired with the existing LionSpeed walking speed recording device. From there, studies can be run to see how users in the medical field, such as nurses, interact with both the software and the device itself to see if there are any usability issues. Decisions could be made while following the risk-driven spiral model to ensure that effort is inputted when necessary to ensure the greatest chances for success (Pew & Mavor, 2007). This would help bring this project one step closer to becoming a product on the market for use in the medical field, or it could promote other researchers to come up with even more innovative methods of gathering patient information.

8. DELEGATION OF TASKS

In the creation of this report, Courtney was tasked with writing the introduction and literature review sections, as well as the limitations, conclusions, and future work section. Jessica wrote the methodology and results sections and generated the graphs and all statistical analyses. Both Jessica and Courtney were present during the usability studies, where Jessica read the tasks to the participants and Courtney took notes. Both Jessica and Courtney edited other sections of the document that the other was responsible for to ensure accuracy. This study was supported under IRB guidelines, where consent was given by all participants and they were paid for their time.

References

Barthuly, A. M., Bohannon, R. W., & Gorack, W. (2012). Gait speed is a responsive measure of physical performance for patients undergoing short-term rehabilitation. *Gait & Posture, 36*(1), 61-64. doi:https://doi.org/10.1016/j.gaitpost.2012.01.002

Benbunan-Fich, R. (2001). Using protocol analysis to evaluate the usability of a commercial web site. *Information & management, 39*(2), 151-163.

Bevan, N. (1997). *Usability Issues in Web Site Design.* Paper presented at the HCI.

Bijleveld-Uitman, M., van de Port, I., & Kwakkel, G. (2013). Is gait speed or walking distance a better predictor for community walking after stroke? *Journal of rehabilitation medicine, 45*(6), 535-540. doi:https://doi.org/10.2340/16501977-1147

Bohannon, R. W., & Crouch, R. (2017). Minimal clinically important difference for change in 6-minute walk test distance of adults with pathology: a systematic review. *Journal of Evaluation in Clinical Practice, 23*(2), 377-381. doi:https://doi.org/10.1111/jep.12629

Bowden, M. G., Balasubramanian, C. K., Behrman, A. L., & Kautz, S. A. (2008). Validation of a Speed-Based Classification System Using Quantitative Measures of Walking Performance Poststroke. *Neurorehabilitation and Neural Repair, 22*(6), 672-675. doi:10.1177/1545968308318837

Brach, J. S., & VanSwearingen, J. M. (2002). Physical Impairment and Disability: Relationship to Performance of Activities of Daily Living in Community-Dwelling Older Men. *Physical Therapy, 82*(8), 752-761. doi:10.1093/ptj/82.8.752

Chester, J. G., & Rudolph, J. L. (2011). Vital Signs in Older Patients: Age-Related Changes. *Journal of the American Medical Directors Association, 12*(5), 337-343. doi:https://doi.org/10.1016/j.jamda.2010.04.009

Cooper, R., Kuh, D., Cooper, C., Gale, C. R., Lawlor, D. A., Matthews, F., . . . Teams, H. S. (2010). Objective measures of physical capability and subsequent health: a systematic review. *Age and Ageing, 40*(1), 14-23. doi:10.1093/ageing/afq117

Corrao, N. J., Robinson, A. G., Swiernik, M. A., & Naeim, A. (2010). Importance of Testing for Usability When Selecting and Implementing an Electronic Health or Medical Record System. *Journal of Oncology Practice, 6*(3), 120-124. doi:10.1200/jop.200017

Dobkin, B. H. (2006). Short-distance walking speed and timed walking distance: Redundant measures for clinical trials? *Neurology, 66*(4), 584-586. doi:10.1212/01.wnl.0000198502.88147.dd

Elbaz, A., Sabia, S., Brunner, E., Shipley, M., Marmot, M., Kivimaki, M., & Singh-Manoux, A. (2013). Association of walking speed in late midlife with mortality: results from the Whitehall II cohort study. *AGE, 35*(3), 943-952. doi:10.1007/s11357-012-9387-9

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*(3), 215-251. doi:10.1037/0033-295X.87.3.215

Faggiano, P., D'Aloia, A., Gualeni, A., Brentana, L., & Cas, L. D. (2004). The 6 minute walking test in chronic heart failure: indications, interpretation and limitations from a review of the literature. *European Journal of Heart Failure, 6*(6), 687-691. doi:https://doi.org/10.1016/j.ejheart.2003.11.024

Ferrucci, L., Bandinelli, S., Benvenuti, E., Di Iorio, A., Macchi, C., Harris, T. B., & Guralnik, J. M. (2000). Subsystems Contributing to the Decline in Ability to Walk: Bridging the Gap Between Epidemiology and Geriatric Practice in the InCHIANTI Study. *Journal of the American Geriatrics Society, 48*(12), 1618-1625. doi:https://doi.org/10.1111/j.1532-5415.2000.tb03873.x

Fien, S., Henwood, T., Climstein, M., Rathbone, E., & Keogh, J. W. L. (2019). Exploring the feasibility, sustainability and the benefits of the GrACE + GAIT exercise programme in the residential aged care setting. *PeerJ, 7*, e6973. doi:10.7717/peerj.6973

Fien, S., Henwood, T., Climstein, M., Rathbone, E., & Keogh, J. W. L. (2019). Gait Speed Characteristics and Their Spatiotemporal Determinants in Nursing Home Residents: A Cross-Sectional Study. *Journal of Geriatric Physical Therapy, 42*(3), E148-E154. doi:10.1519/jpt.0000000000000160

Fisher, R. A. (1925). Theory of Statistical Estimation. *Mathematical Proceedings of the Cambridge Philosophical Society, 22*(5), 700-725. doi:10.1017/S0305004100009580

Fritz, S., & Lusardi, M. (2009). White Paper: “Walking Speed: the Sixth Vital Sign”. *Journal of Geriatric Physical Therapy, 32*(2), 2-5.

Godha, S., & Lachapelle, G. (2008). Foot mounted inertial system for pedestrian navigation. *Measurement Science and Technology, 19*(7), 075202. doi:10.1088/0957-0233/19/7/075202

Graham, J. E., Ostir, G. V., Fisher, S. R., & Ottenbacher, K. J. (2008). Assessing walking speed in clinical research: a systematic review. *Journal of Evaluation in Clinical Practice, 14*(4), 552-562. doi:https://doi.org/10.1111/j.1365-2753.2007.00917.x

Green, J., Forster, A., & Young, J. (2002). Reliability of gait speed measured by a timed walking test in patients one year after stroke. *Clinical Rehabilitation, 16*(3), 306-314. doi:10.1191/0269215502cr495oa

Guimaraes, R. M., & Isaacs, B. (1980). Characteristics of the gait in old people who fall. *International Rehabilitation Medicine, 2*(4), 177-180. doi:10.3109/09638288009163984

Hornbæk, K. (2006). Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human-Computer Studies, 64*(2), 79-102. doi:https://doi.org/10.1016/j.ijhcs.2005.06.002

Hornbæk, K., & Law, E. L.-C. (2007). *Meta-analysis of correlations among usability measures*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, San Jose, California, USA. https://doi.org/10.1145/1240624.1240722

Hsieh, H.-F., & Shannon, S. E. (2005). Three Approaches to Qualitative Content Analysis. *Qualitative Health Research, 15*(9), 1277-1288. doi:10.1177/1049732305276687

Imms, F. J., & Edholm, O. G. (1981). Studies of Gait and Mobility in the Elderly. *Age and Ageing, 10*(3), 147-156. doi:10.1093/ageing/10.3.147

Ingemarsson, A. H., Frändin, K., Mellström, D., & Möller, M. (2003). Walking ability and activity level after hip fracture in the elderly--a follow-up. *Journal of rehabilitation medicine, 35*(2), 76-83. doi:10.1080/16501970306113

Jung, H. W., Roh, H. C., Kim, S. W., Kim, S., Kim, M., & Won, C. W. (2019). Cross-Comparisons of Gait Speeds by Automatic Sensors and a Stopwatch to Provide Converting Formula Between Measuring Modalities. *Ann Geriatr Med Res, 23*(2), 71-76. doi:10.4235/agmr.19.0016

Kiss, R. M. (2010). Effect of walking speed and severity of hip osteoarthritis on gait variability. *Journal of Electromyography and Kinesiology, 20*(6), 1044-1051. doi:https://doi.org/10.1016/j.jelekin.2010.08.005

Kong, W., Lin, J., Waaning, L., Sessa, S., Cosentino, S., Magistro, D., . . . Takanishi, A. (2016, 3-7 Dec. 2016). *Comparison of gait event detection from shanks and feet in single-task and multi-task walking of healthy older adults.* Paper presented at the 2016 IEEE International Conference on Robotics and Biomimetics (ROBIO).

Kukreja, U., Stevenson, W. E., & Ritter, F. E. (2006). RUI: Recording user input from interfaces under Windows and Mac OS X. *Behavior Research Methods, 38*(4), 656-659. doi:10.3758/BF03193898

Kullo, I. J., & Rooke, T. W. (2016). Peripheral Artery Disease. *New England Journal of Medicine, 374*(9), 861-871. doi:10.1056/NEJMcp1507631

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics, 33*(1), 159-174. doi:10.2307/2529310

Lusardi, M. M. (2012). Is Walking Speed a Vital Sign? Absolutely! *Topics in Geriatric Rehabilitation, 28*(2), 67-76. doi:10.1097/TGR.0b013e31824385a4

MacPhail, C., Khoza, N., Abler, L., & Ranganathan, M. (2016). Process guidelines for establishing Intercoder Reliability in qualitative studies. *Qualitative Research, 16*(2), 198-212. doi:10.1177/1468794115577012

Maki, B. E. (1997). Gait Changes in Older Adults: Predictors of Falls or Indicators of Fear? *Journal of the American Geriatrics Society, 45*(3), 313-320. doi:https://doi.org/10.1111/j.1532-5415.1997.tb00946.x

Martin, E., Kim, S., Unfried, A., Delcambre, S., Sanders, N., Bischoff, B., & Saavedra, R. (2019). 6th vital sign app: Testing validity and reliability for measuring gait speed. *Gait & Posture, 68*, 264-268. doi:https://doi.org/10.1016/j.gaitpost.2018.12.005

McDermott, M. M., Liu, K., Ferrucci, L., Tian, L., Guralnik, J. M., Liao, Y., & Criqui, M. H. (2011). Greater Sedentary Hours and Slower Walking Speed Outside the Home Predict Faster Declines in Functioning and Adverse Calf Muscle Changes in Peripheral Arterial Disease. *Journal of the American College of Cardiology, 57*(23), 2356-2364. doi:doi:10.1016/j.jacc.2010.12.038

Middleton, A., Fritz, S. L., & Lusardi, M. (2015). Walking Speed: The Functional Vital Sign. *23*(2), 314. doi:10.1123/japa.2013-0236 10.1123/japa.2013-0236 10.1123/japa.2013-0236 10.1123/japa.2013-0236

Middleton, A., Fulk, G. D., Beets, M. W., Herter, T. M., & Fritz, S. L. (2016). Self-Selected Walking Speed Is Predictive of Daily Ambulatory Activity in Older Adults. *24*(2), 214. doi:10.1123/japa.2015-0104 10.1123/japa.2015-0104

Morey, M. C., Ryan, S., Kelly, J., Liu, C., Hawkins, K., Schwartz, K., & Prvu Bettger, J. (2017). THE 6TH VITAL SIGN: A MOBILE APP FOR POPULATION HEALTH SURVEILLANCE OF WALKING SPEED. *Innovation in Aging, 1*(Suppl 1), 669-669. doi:10.1093/geroni/igx004.2377

Müller, R., Ertelt, T., & Blickhan, R. (2015). Low back pain affects trunk as well as lower limb movements during walking and running. *Journal of Biomechanics, 48*(6), 1009-1014. doi:https://doi.org/10.1016/j.jbiomech.2015.01.042

Nielsen, J. (1992). *Finding usability problems through heuristic evaluation*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Monterey, California, USA. https://doi.org/10.1145/142750.142834

Nielsen, J. (1994). *Usability engineering*: Morgan Kaufmann.

Odonkor, C. A., Schonberger, R. B., Dai, F., Shelley, K. H., Silverman, D. G., & Barash, P. G. (2013). New utility for an old tool: can a simple gait speed test predict ambulatory surgical discharge outcomes? *American journal of physical medicine & rehabilitation, 92*(10), 849-863. doi:10.1097/PHM.0b013e3182a51ac5

Peters, D. M., Fritz, S. L., & Krotish, D. E. (2013). Assessing the Reliability and Validity of a Shorter Walk Test Compared With the 10-Meter Walk Test for Measurements of Gait Speed in Healthy, Older Adults. *Journal of Geriatric Physical Therapy, 36*(1), 24-30. doi:10.1519/JPT.0b013e318248e20d

Pew, R. W., & Mavor, A. S. (2007). *Human-System Integration in the System Development Process: A New Look*. Washington, DC: The National Academies Press.

Poon, S. J., Greenwood-Ericksen, M. B., Gish, R. E., Neri, P. M., Takhar, S. S., Weiner, S. G., . . . Landman, A. B. (2016). Usability of the Massachusetts Prescription Drug Monitoring Program in the Emergency Department: A Mixed-methods Study. *Academic Emergency Medicine, 23*(4), 406-414. doi:https://doi.org/10.1111/acem.12905

Ritter, F. E., & Schooler, L. J. (2001). The learning curve. In W. Kintch, N. Smelser, & P. Baltes (Eds.), *International Encyclopedia of the Social and Behavioral Sciences*. Oxford, UK: Pergamon.

Soltani, A., Dejnabadi, H., Savary, M., & Aminian, K. (2020). Real-World Gait Speed Estimation Using Wrist Sensor: A Personalized Approach. *IEEE Journal of Biomedical and Health Informatics, 24*(3), 658-668. doi:10.1109/JBHI.2019.2914940

van Hedel, H. J. A. (2009). Gait Speed in Relation to Categories of Functional Ambulation After Spinal Cord Injury. *Neurorehabilitation and Neural Repair, 23*(4), 343-350. doi:10.1177/1545968308324224

Wang, J., Lin, C., Yang, Y. C., & Ho, Y. (2012). Walking Pattern Classification and Walking Distance Estimation Algorithms Using Gait Phase Information. *IEEE Transactions on Biomedical Engineering, 59*(10), 2884-2892. doi:10.1109/TBME.2012.2212245

Witte, U. S., & Carlsson, J. Y. (1997). Self-selected walking speed in patients with hemiparesis after stroke. *Scandinavian journal of rehabilitation medicine, 29*(3), 161-165.

Woo, J., Ho, S. C., & Yu, A. L. M. (1999). Walking Speed and Stride Length Predicts 36 Months Dependency, Mortality, and Institutionalization in Chinese Aged 70 And Older. *Journal of the American Geriatrics Society, 47*(10), 1257-1260. doi:10.1111/j.1532-5415.1999.tb05209.x

9. APPENDIX

9.1 Appendix A: Usability Study Script

Introduction:

Hi ______,

First, we want to thank you for taking the time to meet with us today. As part of the IRB guidelines and if you agree, you will be compensated with a $15 Amazon Gift Card that will be provided some time soon. If you wish to have receive this compensation, let us know so we can make the necessary arrangements to send you the compensation.

As background information we are part of the LionSpeed Team. There currently exists a prototype of a sensor, called the LionSpeed sensor, that can automatically detect walking speeds of patients. Our aim is to create an interface that would allow clinicians, nurses, and physicians to effectively enter, track, and share data from walking speed tests using the sensor. For this reason, we improved on a previous developed and interface using Axure to provide the users with a medium-fidelity prototype to conduct our usability studies. The goal of the usability studies is to understand the interactions and the ease of use that you as a user have with the interface.

As you might know, this usability study will be used of our data analysis, and we require your consent to record this meeting. We want to make sure we do not miss any of the key information that you tell us today. For this reason, ______, do you consent to record this meeting?

Perfect, thank you!

Pre-Survey:

We are going to begin by requesting you to complete the Pre-Survey on the link we just sent you on the Zoom chat. This pre-survey will allow us to keep track of your information. After you complete the pre-survey, let us know so we can continue to the usability study.

https://pennstate.qualtrics.com/jfe/form/SV_1N7NiUqNqG7st4V

Testing:

Team: Click on interface **https://706ssr.axshare.com**

Poner Log File Location

Team: Open RUI interface

We will be screen sharing the screen with the interface,

As you can see this is the interface you will be interacting and the other screen is a Recording User Input will be recording the clicks during your interaction with the interface. During this study you will be prompted to verbally articulate your thought process during the two different tasks. After you finish the two tasks, you will be given a link to a Post-Survey where you can provide us with information about your interaction with the interface. Do you have any questions?

The first task is:

"Ashley is an 82-year-old patient who has been complaining about pain in her lower hip and feels like she has been walking slower. She comes into your clinic and wants to get your opinion. Check Ashley's speed and trends for the past year, and report her speed and fatigue level recorded during her last visit."

Team: Paste description on chat (Courtney).

We have the description of the task in the other tab for you to reference it and can go to the other tab as many times as you want to make sure you understand the task.

So now I will give you control over my screen (**Team:** Give control and make them click and move mouse). Now you can click on the RUI interface "Start Recording" and begin clicking along the interface and verbally articulating your thought process.

After finishing: You can stop the RUI recording. Thank you! Now we will take control again really quickly to make a new RUI record.

Team: Create a new RUI recording, Poner Log File Location, Poner

Now we will give to control again of our screen.

The second task is:

"You then conduct a walking speed test with Ashley using the LionSpeed device. Ashley specifies that she has pain in her hip, and is experiencing moderate breathlessness and severe pain while walking. When asked to rate her pain and breathlessness on a scale from 1 to 5, she reports a 3 from breathlessness and a 4 for pain. She does not report issues with anything else. Create a new entry for Ashley."

We have pasted the task description in the Zoom Chat for reference. Now you can click on the RUI interface "Start Recording" and begin clicking along the interface and verbally articulating your thought process.

After finishing: You can stop the RUI recording. Thank you! Now we will get control again of our screen.

Post-Survey:

Now, we will provide you the link for the post-survey through the Zoom Chat so you can provide us with comments about the interface and how was your interaction with the interface.

https://pennstate.qualtrics.com/jfe/form/SV_0H9IvPbfQ9o4GTX

https://pennstate.qualtrics.com/jfe/form/SV_1KMQrM63Kk4j5YN (med)

Thank you for your time!

9.2 Appendix B: Pre-Survey

PennState

First Name:

Last Name:

Please select your occupation:

- Doctor
- Physical Therapist
- Nursing Student

If you are a practicing clinician, what is your area of specialty? (Please leave the question blank and submit if you are not)

→

9.3 Appendix C: Post-Survey

PennState

Please indicate your extent of agreement with the following statements:

	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
Use of terminology and appropriate and consistent	○	○	○	○	○	○	○
The interface was clear and prevented any errors	○	○	○	○	○	○	○
All objects, actions, and options were visible	○	○	○	○	○	○	○
The design was aesthetically pleasing	○	○	○	○	○	○	○
The interface had appropriate features that were suited for my needs	○	○	○	○	○	○	○
The interface allowed me to complete the task in an appropriate period of time	○	○	○	○	○	○	○
The sequence of tasks was appropriate	○	○	○	○	○	○	○
Messages were appropriately positioned on the screen	○	○	○	○	○	○	○

→

PennState

Please indicate the level of difficulty with respect to the following criteria:

	Extremely easy	Moderately easy	Slightly easy	Neither easy nor difficult	Slightly difficult	Moderately difficult	Extremely difficult
Readability of text on the screen	○	○	○	○	○	○	○
Learning to use the interface	○	○	○	○	○	○	○
Correcting any mistakes made	○	○	○	○	○	○	○

→

PennState

Do you feel that this system would be a good addition to the services at Hershey Medical Center?

- ○ Yes, perfect the way it currently is
- ○ Probably yes, with some minor changes
- ○ Maybe needs some major changes
- ○ Probably not, needs to be redesigned completely
- ○ Definitely not, I don't see a use

→

PennState

How likely are you to use this system in the future?

Extremely unlikely	Moderately unlikely	Slightly unlikely	Neither likely nor unlikely	Slightly likely	Moderately likely	Extremely likely
○	○	○	○	○	○	○

Please explain why:

→

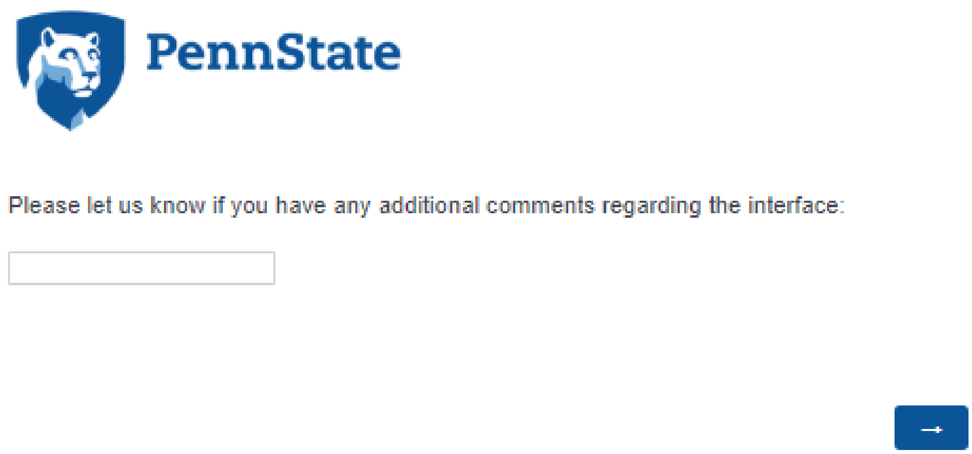
PennState

How satisfied are you with the overall interface?

Extremely dissatisfied	Moderately dissatisfied	Slightly dissatisfied	Neither satisfied nor dissatisfied	Slightly satisfied	Moderately satisfied	Extremely satisfied
○	○	○	○	○	○	○

Please explain why:

→



9.4 Appendix D: Codebook for Research Question 3

Table 1: Codebook for Verbal Protocol Analysis

Level 3	Level 2	Definition	Level 1
Satisfaction with the Interface	Context-dependent Questions	"Users' satisfaction with specific features or circumstances in the specific context of use"	Positive
			Negative
	Ease of Use	"Broad measures of users' overall satisfaction or attitudes towards the interface or user experience"	Positive
			Negative
Users' Attitudes and Perceptions	Attitudes towards content	"Attitudes towards the content of the interface when content can be distinguished from the interface"	Positive
			Negative
	Perception of interaction	"Measures users' perception of the interaction"	Positive
			Negative
	Perception of outcome	"Users' perception of the outcome of the interaction"	Positive
			Negative
Other		"Other measures of satisfaction"	Positive
			Negative

9.5 Appendix E: Snippets of some transcripts for Research Question 3

Unknown Speaker 0:00
I'm just trying to select Ashley's name.

Unknown Speaker 0:04
I guess maybe I'll have to type it. No?

Unknown Speaker 0:15
Okay, that's the right person.

Unknown Speaker 0:20
I'll probably want to check the existing records.

Unknown Speaker 0:28
I can see what she's been doing the last year.

Unknown Speaker 0:34
So it'd be good to know as background information for actually going to talk to her.

Unknown Speaker 0:48
Okay, so looks like nothing has been hurting her at least on December 12. There's graphs.

Unknown Speaker 1:10
Okay, so looks like she's been becoming more fatigued as it goes on. There's more graphs.

Unknown Speaker 1:26
Okay, on her speed has been decreasing.

Unknown Speaker 1:42
Looks like just from this the pain has been increasing. We'll probably want to see what the actual levels have been.

Unknown Speaker 1:51
Shows me...

0:04
Okay, so I'm assuming that you click...no? So I want to add, I'm assuming, that's kind of confusing. I don't know how to get to her chart. Maybe I'll sear

0:55
Maybe I'm doing this wrong. Okay, I'm gonna try to go to the existing records. Oh, I guess I'm already in her. Oh, I see. Okay, this zoom thing was blocki

1:56
So I'm gonna view her records from her last visit, which would've been December 12. She has fatigue at a five. Her walking speed was 2.3 meters per secon

2:32
So, her average for the last visit was 2.3 meters per second. So um, it has been decreasing since her, looks like her first visit and August. And her fati

3:01
Okay. Awesome. Thank you. So now if you can click on the we think that's the bottom to stop this recording. This?

3:12
Yeah, that's about recording. Awesome. Yep. And I'm going to control again. So before we do Test, test two, let me just start another recording.

3:38
Okay.

3:45
There we go.

3:53
So the second task is you then conduct a walking speed test with Ashley using the line speed device. Ashley specifies that she has pain in her hip, and is

4:50
Click that again. Three for breathlessness, and four for pain. Create an entry. So, I'm assuming I would hit this because it says "add data from a new exa

6:03
Maybe it's the trend, this is kind of confusing to like figure out what to put it under.

0:00
What was her name? Ashley?

0:08
Can I click her?

0:40
I'm just going through and reading, manage everything I can get so I can understand what's been going on with her. Oh, I think I can only click on that o

0:58
Okay, so she's been getting kind of slower since they first saw her. Um, fatigue, let's look at that.

1:10
So she's been feeling I guess more fatigue every time we've been checking on her and pain, it looks like that's also been kind of increasing. Unsteadines

1:21
Umm, trembling, and increased breathlessness, okay, which would make sense if she's having trouble walking. So these are all things that I would expect w

1:56
So how are you recording?

2:00
Um, am I recording from the sensor?

2:05
So this, basically, you can reference the first task again, but it's just like, going through checking the speeds, and checking the trends.

2:17
Oh, just checking the trends. Check your speed, [inaudible], record your speed, if it's [inaudible], according to during her last visit.

2:29
So if I should add when we say report, we mean to verbally state to us, so let us know what the findings are.