Reverse Engineering of Content as a Task for Finding Usability Problems: An Evaluative Case Study using the Wikibreathe Tool for Online Creation of Asthma Action Plans

by

Flora Wan

A thesis submitted in conformity with the requirements for the degree of Master of Applied Science
Department of Mechanical and Industrial Engineering
University of Toronto

© Copyright by Flora Wan 2009
Reverse Engineering of Content as a Task for Finding Usability Problems: An Evaluative Case Study using the Wikibreathe tool for Online Creation of Asthma Action Plans

Flora Wan

Master of Applied Science

Department of Mechanical and Industrial Engineering
University of Toronto

2009

Abstract

After formulating the problem and reviewing relevant research literature, a study was performed that compared reverse engineering with traditional scenario-based techniques in usability evaluation. In this case study, an online tool for creating asthma action plans was created and evaluated through questionnaires and focus groups. The tool was then tested in a controlled study using both a traditional scenario-based approach and the reverse engineering method. A group of twelve users built asthma action plans using each method in a randomized order. Results concerning usability, efficiency and the types of usability problems found were reported, along with recommendations for further research in the use of reverse engineering as a method of usability evaluation.
Acknowledgments

I begin by thanking my Lord and Saviour Jesus Christ, to whom I owe my life.

Thanks to my husband James who is always encouraging me to go out and chase my dreams.

Much appreciation also goes to my supervisor and mentor Professor Mark Chignell, who believed in me and worked with me each step of the way. I would also like to thank the other members of my thesis committee, Sharon Straus and Deborah Fels.

This research would not have been possible without Dr. Samir Gupta and David Newton’s vision and collaboration on creating and implementing the Wikibreathe tool.

Last, but not least, thank you to my friends at the University of Toronto who brainstormed many interesting ideas with me and to all the participants in my study who gave me honest and valuable feedback.
# Table of Contents

Acknowledgments.................................................................................................................................................. iii  
Table of Contents................................................................................................................................................ iv  
List of Tables ....................................................................................................................................................... viii  
List of Figures ...................................................................................................................................................... ix  
Chapter 1............................................................................................................................................................ 1  
1 Introduction...................................................................................................................................................... 1  
1.1 Research Motivation..................................................................................................................................... 2  
1.2 Thesis Structure .......................................................................................................................................... 2  
Chapter 2............................................................................................................................................................ 3  
2 Literature Review............................................................................................................................................ 3  
2.1 What is Usability? ........................................................................................................................................ 3  
2.2 Usability Evaluation Methods ..................................................................................................................... 5  
2.2.1 Inquiry................................................................................................................................................... 6  
2.2.2 Inspection.............................................................................................................................................. 8  
2.2.3 Formal Usability Testing ....................................................................................................................... 9  
2.3 Reverse Engineering.................................................................................................................................. 11  
2.3.1 Comparison of Reverse Engineering to Other Usability Methods ................................................. 12  
2.4 Evaluation Instruments and Tools ............................................................................................................. 14  
2.4.1 System Usability Scale ....................................................................................................................... 14  
2.4.2 Heuristic Evaluation............................................................................................................................ 15  
Chapter 3............................................................................................................................................................ 17  
3 The Wikibreathe Project ................................................................................................................................. 17  
3.1 Asthma Action Plans ................................................................................................................................. 17  
3.2 The “Wiki” Concept .................................................................................................................................. 18
List of Tables

Table 2-1: Examples of Usability Evaluation Methods based on Categories (Battleson, 2001) ... 6

Table 5-1: Participants' Attitude toward Computers ................................................................. 32

Table 5-2: Inclusion of subtasks for each evaluation method .................................................. 35

Table 6-1: Average Ratings on the System Usability Scale ....................................................... 45

Table 6-2: Average Ratings on Attitude toward the Usability Testing Methods ...................... 45

Table 6-4: Usability Problems .................................................................................................. 48

Table 6-5: Number of Usability Problems Found for the First Time during each task .......... 49

Table 7-1: Converted SUS Scores .......................................................................................... 54

Table 7-2: Mapping of Usability Methods and Requirements ................................................. 58

Table D-1: Protocol A, Task 1 ................................................................................................. 84

Table D-2: Protocol B, Task 1 ................................................................................................. 85

Table D-3: Protocol A, Task 2 ................................................................................................. 85

Table D-4: Protocol B, Task 2 ................................................................................................. 86
List of Figures

Figure 2-1: A model of the attributes of system acceptability (Nielson, 1993)................................. 4

Figure 2-2: Typical Reverse Engineering Process (Storey et al., 2002)............................................. 12

Figure 3-1: Examples of Asthma Action Plans................................................................................... 19

Figure 3-2: Sections of an asthma action plan: 1) Header, 2) Zone Description, 3) Zone Instruction, 4) Footer .................................................................................................................... 21

Figure 3-3: Landscape (left) and Portrait (right) oriented plans....................................................... 21

Figure 4-1: Initial Prototype of the Wikibreathe Tool....................................................................... 24

Figure 4-2: Navigation Structure ...................................................................................................... 26

Figure 4-3: Choosing a tab............................................................................................................... 27

Figure 4-4: Choosing an option ....................................................................................................... 28

Figure 4-5: Options appears on action plan .................................................................................... 28

Figure 4-6: Deselecting an item in header (left) and in zones (right)............................................. 28

Figure 4-7: Adding comments ......................................................................................................... 29

Figure 4-8: Displaying terminology with hover ............................................................................... 29

Figure 4-9: Generating a PDF......................................................................................................... 29

Figure 4-10: Maximum number of options.................................................................................... 30

Figure 4-11: Space limitation alert message.................................................................................... 30

Figure 5-1: Asthma action plans generated by the Forward Flow (left) and Reverse Flow (right)........................................................................................................................................ 36
Figure 6-1: Comparison of Median Times for Forward and Reverse Flows for Protocol A (left) and Protocol B (right) ................................................................................................................... 41

Figure 6-2: Order Effect for Zone Description ........................................................................................................ 42

Figure 6-3: Order Effect for Footer ....................................................................................................................... 42

Figure 6-4: Comparison of average scores in participant attitude toward tool (SUS2) .................. 46

Figure 6-5: Comparison of average scores in participant attitude toward tool (SUS3) ............ 47

Figure 6-6: Selecting the layout of the asthma action plan................................................................. 50

Figure 7-1: Frequency Distribution of SUS Scores (Bangor, 2008) ......................................................... 53

Figure 7-2: Reordering asthma symptoms directly on asthma action plan.................................. 55

Figure 7-3: Peak Flow and Terminology ................................................................................................. 56

Figure B-1: Examples of Asthma Action Plan designed by Respirologists ................................... 71

Figure B-2: Examples of Asthma Action Plans generated by a Primary Care Physician or Asthma Educator ......................................................................................................................... 71

Figure B-3: Examples of Asthma Action Plans generated by Patients........................................... 72

Figure G-1: The dialogue/menu for Wikibreathe ...................................................................................... 94

Figure G-2: Tab and Sub-Tabs ....................................................................................................................... 95

Figure G-3: Arrows to indicate that a new submenu or drop-down has appeared .................... 95

Figure G-4: End of Section Message ......................................................................................................... 96

Figure G-5: Space limitation alert box ................................................................................................. 96
Chapter 1

1 Introduction

Usability is a fairly broad term that has been defined in a number of ways. Nielsen (1993) said that usability is about learnability, efficiency, memorability, errors, and satisfaction. In contrast, ISO 9241-11 defines usability as: “the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.”

The construct of usability also involves the study of the principles behind an object's perceived efficiency or elegance. Usability testing focuses on measuring a product's capacity to meet its intended purpose.

There are numerous existing usability evaluation methods, both quantitative (e.g. measuring the efficiency and number of errors as a user performs a task) and qualitative (e.g. using questionnaires and focus groups to gather feedback on user preferences). This study examines whether a new method based on the concept of “reverse engineering” can be a viable technique in the area of usability testing. In this method, users are given the final product such as a document or graphic that is created by a tool. With minimal training, they are asked to use the tool to “reverse engineer” or recreate the same product. The idea is that as the novice user attempts to rebuild the product, the effectiveness, efficiency and satisfaction of the tool (ISO 9241-11) can be measured through the amount of time required to complete a task, the number of errors encountered, and the difficulties in learning how to use the tool.

In order to test this concept, a series of experiments were conducted to evaluate an online tool that was developed to let users design a paper plan for asthma patients. In the experiments, users were asked to use the tool to perform the same task of creating an asthma action plan twice – once using the traditional scenario and task-based method and once using the reverse engineering technique. The usability problems found by using each method and the amount of time users took to complete the tasks were recorded and analyzed, and the results are documented in this thesis.
The results obtained using the two methods were then compared in a number of tests, recommendations were provided concerning guidelines on which method might be most beneficial in different situations. The limitations of this study are also noted and ideas for future work are suggested.

1.1 Research Motivation

The purpose of this research is to carry out a study to investigate the effectiveness of using a reverse engineering approach for usability testing.

The following research questions will be explored in this study:

1. How does reverse engineering compare to other usability evaluation techniques in terms of the number of usability problems it can uncover?

2. How do user attitudes and satisfaction level differ between methods?

3. What types of usability problems can be found using the reverse engineering method?

1.2 Thesis Structure

Chapters 1 and 2 focus on the introduction of the topic and relevant background literature. Chapters 3 and 4 introduce the Wikibreathe Project and the design and prototyping phases of the Wiki tool as well as results from the focus groups and questionnaires used to assess the usability of the tool. Chapters 5 and 6 discuss the overall methodology used in the usability testing phase and report the results. Finally, Chapters 7 and 8 summarize the conclusions from this study and discuss design recommendations and ideas for future work.
Chapter 2

2 Literature Review

This section will discuss the relevant background literature for this research study starting with definitions of usability and descriptions of some of the more common existing usability evaluation methods. It will then introduce the concept of reverse engineering as a new evaluation technique and conclude with some specific evaluation instruments used in this study.

2.1 What is Usability?

As noted earlier, the International Standards Organization standard, ISO 9241-11 defined usability as (ISO, 1995):

*The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.*

Effectiveness refers to the degree of accuracy and completeness in which users are able to achieve their goals. Efficiency relates to the amount of resources, whether time or effort, spent while achieving the goal. Satisfaction is determined by how comfortable the user feels when using the system (Helander, 2006). This standard defines usability in both quantitative and qualitative terms.

According to Nielson (1993), a system that is usable should have the following five attributes:

- **Learnability**: How easy is it for users to accomplish basic tasks the first time they encounter the design?
- **Efficiency**: Once users have learned the design, how quickly can they perform tasks?
- **Memorability**: When users return to the design after a period of not using it, how easily can they re-establish proficiency?
- **Errors**: How many errors do users make, how severe are these errors, and how easily can they recover from the errors?
- **Satisfaction**: How pleasant is it to use the design?
Figure 2-1: A model of the attributes of system acceptability (Nielson, 1993)

**Learnability**

Learnability is a particularly important attribute for novice users to a system. Humans appear to be able to learn new concepts without needing to be programmed explicitly (Valiant, 1984). In software testing, learnability is the capability of a software product to enable the user to learn how to use it (ISO 9126).

The concept that a product has a single measure of learnability may be problematic, however, as the learnability of a product may depend on how well the learning methods it encourages match the learning styles of users. There are more than 50 different learning-style theories (Gerding, 2007), many of which are based on sensory preferences with broad categories that include: Visual, Auditory, and Kinesthetic (VAK). *Visual Learners* need to see the teacher's body language and facial expression to fully understand the content of a lesson. They may think in pictures and learn best from visual displays including: diagrams, illustrated text books, overhead transparencies, videos, flipcharts and hand-outs. *Auditory Learners* learn best through verbal lectures, discussions, talking things through and listening to what others have to say. *Tactile/Kinesthetic Learners* learn best through a hands-on approach, actively exploring the physical world around them. They may find it hard to sit still for long periods and may become distracted by their need for activity and exploration.
**Efficiency and Error**

“Efficiency” is defined as “the amount of effort required to accomplish a goal” (Jordan, 1998). For most tasks in a system, there is a path that requires the least effort. When users deviate from this ideal path, there would be a cost in terms of additional time and effort expended. For example, if a system is designed such that the user has to continually look up commands in a manual, then the design is not efficient. In this way, efficiency can be measured by the amount of time required by a user to accomplish a specific task.

On the other hand, “Error” is a deviation from the path of accomplishing a goal (Jordon, 1998). The number of errors committed by a user can also be used as a measurement of how usable the system is designed.

**Satisfaction**

Finally, “Satisfaction” is a qualitative measurement of the enjoyment that users get from using system. Some questions to consider are: Do users enjoy using the system? Are users frustrated or confused? Do users prefer a particular design? The answers to these questions, though subjective in nature, are vital in measuring the overall usability of the system.

Learnability, Efficiency, Error and Satisfaction are the characteristics that will be used to measure usability in this study. Memorability is not included since the focus of this research is on novice users of a tool.

### 2.2 Usability Evaluation Methods

This section describes the usability methods that will be used in this study.

### Table 2-1: Examples of Usability Evaluation Methods based on Categories (Battleson, 2001)

<table>
<thead>
<tr>
<th>Inquiry</th>
<th>Inspection</th>
<th>Formal Usability Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Contextual Inquiry</td>
<td>• Heuristic Evaluation</td>
<td>• Scenario-based testing</td>
</tr>
<tr>
<td>• Ethnographic Study</td>
<td>• Cognitive Walkthrough</td>
<td>• Thinking Aloud</td>
</tr>
<tr>
<td>• Interviews / Focus Groups</td>
<td>• Formal Usability Inspection</td>
<td>• Co-discovery method</td>
</tr>
<tr>
<td>• Surveys / Questionnaires</td>
<td>• Guideline checklists</td>
<td>• Performance measurement</td>
</tr>
</tbody>
</table>

#### 2.2.1 Inquiry

Methods of inquiry include focus groups, interviews, questionnaires, and surveys. Interviews and focus groups are structured methods of inquiry which are used to gather information about users' experiences and preferences. While surveys and questionnaires may also be used to gather such information, interviews and focus groups allow for more interaction with the users and for immediate answers to questions raised during the interview or focus group. In addition, focus groups consist of multiple users whose interaction may raise additional issues of interest.

**Focus Groups**

A focus group is a form of qualitative research in which a group of people are asked about their attitude towards a product, service, concept, advertisement, idea, or packaging. Questions are asked in an interactive group setting where participants are free to talk with other group members. In most cases, a group of 6-10 carefully selected individuals is led by a moderator in a meticulously organized discussion lasting approximately two hours (Greenbaum, 2009). The goal is to explore the general attitudes of the participants to the topics selected for inclusion in the session.
Krueger et al. (2009) writes that the purpose of conducting a focus group is to listen and gather information, and opinions. Participants are selected because they have certain characteristics in common that relate to the topic of the focus group. The researcher creates a permissive environment that encourages participants to vote or reach consensus. Drawbacks to focus groups include (Krueger, 2009):

- Focus group participants tend to intellectualize
- Focus groups don’t tap into emotions
- Focus group participants may make up answers
- Focus groups produce trivial results
- Dominant individuals can influence results

Although focus groups can be a powerful tool in system development, they should not be used as the only source of usability data (Nielson, 1997).

**Interviews**

The interview is a research method defined by Cannell and Kahn (1968) as a “two-person conversation, initiated by the interviewer for the specific purpose of obtaining research-relevant information, and focused by him on contents specified by research objectives of systematic description, prediction, or explanation.”

Interviews can be categorized as unstructured, semi-structure, or structured (Lindgaard, 1994):

- An *unstructured* interview is exploratory and the research goal is discovering relevant issues to be explored through future research. No restrictions are placed on the issues that are discussed and questions are flexible and open-ended.

- A *semi-structured* interview is more focused than the unstructured interview. Generally, a list of question would be prepared as a guide but the order of the questions are not rigidly enforced to encourage exploration.
A structured interview is one where the questions that are asked are fixed and inflexible. The questions are often formatted to yield “codeable” answers, for example, there may be several categories of which the user chooses one. The interviewer should behave in the exactly same manner towards all participants to avoid bias and error.

**Questionnaires and Surveys**

A questionnaire consists of a series of questions for the purpose of gathering information from respondents. Questionnaires have advantages over some other types of surveys in that they are cheap, do not require as much effort from the questioner as verbal or telephone surveys, and often have standardized answers that make it simple to compile data.

Questionnaires and surveys can either be an open or closed format (Leung, 2001). A questionnaire designed in an open format allows exploration of a range of themes whereas a closed format enforces specific choices and the results are generally easy to code, record and analyze quantitatively.

For questionnaires in a closed format, the most widely used scaling technique is the Likert scale, named after the psychologist Rensis Likert (Polit et al., 1994). A Likert scale typically consists of 5 or 7 choices that ask the respondents to evaluate their level of agreement to a statement. For example, if the statement is “I enjoyed performing Task A with this website”, the choices on the Likert scale might then be:

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Disagree

**2.2.2 Inspection**

With inspection methods like heuristic evaluation and cognitive walkthrough, a site's designers and information specialists serve as testers and subjects, often putting themselves in the place of the user to perform various tasks using the site (Hom, 2003). Unlike inquiry and formal usability testing, these forms of assessment do not enlist the participation of actual users. While evaluations based on inspection are relatively inexpensive to conduct, they are less useful in identifying usability errors than tests with actual users.
Heuristic Evaluation

In 1990, Nielsen and Molich introduced a new method for evaluating user interfaces called heuristic evaluation (Nielsen et al, 1990). The method involves having a small group of usability experts evaluate a user interface using a set of guidelines and noting the severity of each usability problem and where it exists. They found that the aggregated results of five to ten evaluators of four interfaces identified 55 to 90 percent of the known usability problems for those interfaces. They concluded that heuristic evaluation was a cheap and intuitive method for evaluating the user interface early in the design process (Hollingsed et al, 2007).

However, though simple to do, this method is open-ended and easily leads to unreliable results. There is little agreement in the problems reported by different evaluators of the same interface and a high number of false alarms and problems that are missed (Chattratichart, 2008).

Nevertheless, it is one of the most popular methods of usability evaluation due to its low cost of implementation (Nielsen, 1994).

Cognitive Walkthrough

The cognitive walkthrough method is used to evaluate the usability of a user interface by focusing on the ease of learning (Stanton, 2005). It works by having the analyst “walking” through each user action involved in a step of a task. The analyst then considers the effect the interface has on the user’s interactions with the device. The method is rooted in the notion that users typically prefer to learn a system by using it to accomplish tasks, rather than, for example, studying a manual.

2.2.3 Formal Usability Testing

In formal usability testing, users are observed using a site, or prototype, to perform given tasks or achieve a set of defined goals. This method involves employing experiments to gather specific information about a design.

Dumas and Redish (1999) discuss five facets of formal usability testing: (1) the goal is to improve the usability of the interface; (2) testers represent real users; (3) testers perform real
tasks; (4) user behavior and commentary are observed and recorded; and (5) data are analyzed to recognize problems and suggest solutions.

A formal usability test should consist of the following materials (Nemeth, 2004):

1. **Orientation Script:** A brief description of what will happen during the session and is used to prepare the participant. The script describes the test facility, purpose, expectations and any unusual requirements.

2. **Task Scenarios:** The scenario describes the state of the product or system, why the work needs to be accomplished, the information that the user will experience and results the participant will pursue during the test. This may consist of one or a few instruction sheets.

3. **Training Materials:** These are materials used to train participants before a usability session begins.

**Think Aloud**

The “Thinking Aloud” protocol is a popular technique used during usability testing. During the course of a test, where the participant is performing a task as part of a user scenario, they are asked to vocalize their thoughts, feelings and opinions while interacting with the product (Aykin, 2007).

In the most commonly used approach, users’ verbalization takes place simultaneously with their task performance. The verbalization can also take place after users have completed the tasks, in which case this method is variously called retrospective think aloud (Ramey et al., 2006). By collecting user’s verbal reports on their task performance, usability practitioners hope to gain insight into how users interact with the product and identify the barriers that hinder users’ interaction.

**Selection of Usability Evaluation Methods**

Overall, there is no conclusive evidence as to the “best” usability evaluation method. Olson and Moran (1998) wrote that usability evaluation methods should not be used in isolation; instead,
the goal of the research should be to use multiple methods in which important findings can be replicated and validated.

The next section introduces a new usability evaluation method proposed in this research. It is not meant to be a replacement for existing methods but instead should be treated as an alternate technique to be used in conjunction with other relevant methods.

2.3 Reverse Engineering

Reverse engineering is typically associated with software development but it is of interest to the field of Human Factors because its concepts can be adapted to usability testing. Reverse engineering involves examining and analyzing software systems to determine questions such as (Wills, 1996):

- What are the components of the software and how do they interact?
- What is their functionality?
- What design decisions were made in the construction of the software?

Examples of problem areas in software engineering where reverse engineering has been successfully applied include redocumenting programs and relational databases, identifying reusable assets, recovering architectures, recovering design patterns, building traceability between code and documentation, identifying clones, code smells and aspects, computing change impacts, reverse engineering binary code, renewing user interfaces, translating a program from one language to another, migrating or wrapping legacy code (Canfora & Penta, 2007).

However, in this study, reverse engineering is used in a more general sense, abstracted to a higher level. We begin by looking at the definition of reverse engineering. The term reverse engineering as it applies to software engineering is defined by Chikofsky et al. (1990) as:

*The process of analyzing a subject system to (i) identify the system’s components and their inter-relationships and (ii) create representations of the system in another form or at a higher level of abstraction.*
Typical reverse engineering projects follow a sequence of steps such as those illustrated in Figure 2-2 (Storey et al., 2002). Source code from the subject system is parsed to extract information that is then analyzed to produce documentation or diagrams. These two steps are expected to be iterative and may be followed by a reengineering phase with some implementation work to apply specific changes to the source code.

![Figure 2-2: Typical Reverse Engineering Process (Storey et al., 2002)](image)

From a usability testing perspective, these four phases can be adapted as follows:

1. **Parse**: User is given the final product of a system (e.g. a document or graphic) and looks at the elements that compose this product.

2. **Analyse**: User analyzes the parts of the product in order to understand how they are related to the features of the tool.

3. **Document/Visualise**: User looks at each element in more detail and maps them to the functionality of the tool.

4. **Reengineer**: User recreates the product using the tool while the observer notes any difficulties encountered during the reengineering process due to the design flaws in the system.

This method can be seen as an extension to the Formal Usability Testing category of usability evaluation methods since it is based on observations of participants in an experiment.

### 2.3.1 Comparison of Reverse Engineering to Other Usability Methods

The following section compares the pros and cons of using the Reverse Engineering method for usability testing against three other types of testing: Standard User Testing, Focus Groups and Heuristic Evaluation.
Reverse Engineering vs Standard User Testing

The closest comparison for the reverse engineering method would be standard user testing. Standard user testing is a scenario and task-based usability evaluation method where subjects follow either written or verbal instructions to complete a series of tasks using the tool. Standard user testing is often seen as expensive and time consuming (Desurvire et al., 1992) as it requires the tester to recruit a group of subjects, set up the equipment and create a series of tasks (Nemeth, 2004). Reverse engineering also requires a group of subjects and the set up of equipment but the test preparations are much easier – it simply involves creating an end product that can be recreated by the subject.

Secondly, the reverse engineering method encourages more exploration of the tool instead of having users follow a preconceived scenario. By not restricting the user to a specific task, a greater number of unexpected usability issues might be discovered.

Finally, reverse engineering may appeal more to users who are Visual Learners. Visual learners think in pictures and learn best from visual displays (Gerding, 2007). In contrast, instructions for standard user testing are often written or verbal, thus appealing more to those who are more abstract thinkers.

Reverse Engineering vs Focus Groups

Krueger et al. (2009) writes that the purpose of conducting a focus group is to listen and gather information, and opinions. This contrasts with the purpose of using reverse engineering as a usability method which is to focus on the interaction between the user and the system. However, both methods strive to gather a list of usability issues that can be addressed in the system.

The advantage of reverse engineering as a method is that it is done in a one-on-one setting, thus eliminating the effect of group interaction where a subject’s feedback may be influenced by the opinions of other participants. The qualitative data gathered from a focus group may also be more difficult and time-consuming to analyze and report. On the other hand, reverse engineering can be used to generate comparable quantitative data such as the time taken to complete a task or the number of errors that occurred.
Reverse Engineering vs Heuristic Evaluation

Heuristic Evaluation involves having a small group of usability experts evaluate a user interface using a set of guidelines and noting the severity of each usability problem and where it exists (Nielson et al, 1990). The most notable difference between heuristic evaluation and reverse engineering is that heuristic evaluation does not involve end users in the process. By replacing end users with usability experts, some problems that novice users would encounter might be missed by experts who are following a set of pre-existing guidelines. It may also be difficult to find a group of usability evaluation experts and even then, these experts may not agree on the problems reported for the same interface (Chattratichart, 2008). On the other hand, reverse engineering records problems directly encountered by the end users of a system and is therefore a more realistic approach.

2.4 Evaluation Instruments and Tools

This section describes other evaluation instruments specifically used in this study.

2.4.1 System Usability Scale

The System Usability Scale (SUS) is a simple, ten-item scale giving a global view of subjective assessments of usability (Brooke, 1996). It uses a Likert scale ranging from 1 to 5 where 1 = Strongly Disagree and 5 = Strongly Agree.

The ten items on the scale are:

1. I think that I would like to use this system frequently
2. I found the system unnecessarily complex
3. I thought the system was easy to use
4. I think that I would need the support of a technical person to be able to use this system
5. I found the various functions in this system were well integrated
6. I thought there was too much inconsistency in this system
7. I would imagine that most people would learn to use this system very quickly
8. I found the system very cumbersome to use

9. I felt very confident using the system

10. I needed to learn a lot of things before I could get going with this system

2.4.2 Heuristic Evaluation

Heuristic Evaluation has been defined in ISO standard 9241-110 as a set of heuristics to evaluate the usability of a system. The principles, and their definition in the standard, are as follows:

1. Is the dialogue suitable for the user's task and skill level? (Suitability for the task)
   "A dialogue is suitable for a task when it supports the user in the effective and efficient completion of the task. In a dialogue which is suitable for the task, the user is enabled to focus on the task itself rather than the technology chosen to perform that task."

2. Does the dialogue make it clear what the user should do next? (Self-descriptiveness)
   "A dialogue is self-descriptive to the extent that at any time it is obvious to the users which dialogue they are in, where they are within the dialogue, which actions can be taken and how they can be performed."

3. Is the dialogue consistent? (Conformity with user expectations)
   "A dialogue conforms with user expectations if it corresponds to predictable contextual needs of the user and to commonly accepted conventions."

4. Does the dialogue support learning? (Suitability for learning)
   "A dialogue is suitable for learning when it supports and guides the user in learning to use the system."

5. Can the user control the pace and sequence of the interaction? (Controllability)
   "A dialogue is controllable when the user is able to initiate and control the direction and pace of the interaction until the point at which the goal has been met."
6. Is the dialogue forgiving? (Error tolerance)

“A dialogue is error-tolerant if, despite evident errors in input, the intended result may be achieved with either no or minimal corrective action by the user. Error tolerance is achieved by means of damage control, error correction, or error management to cope with errors that occur.”

7. Can the dialogue be customised to suit the user? (Suitability for individualisation)

“A dialogue is capable of individualization when users can modify interaction and presentation of information to suit their individual capabilities and needs.”
Chapter 3

3 The Wikibreathe Project

The Wikibreathe Project, also titled OCTAPUS (“Online Collaboration Tool for Asthma Plans with USability”), is developed by clinical researchers at the Li Ka Shing Knowledge Institute. The purpose of this project is to create a tool that will help ascertain the preferences of various stakeholders when designing an asthma action plan.

My role in this study consisted of collaborating with a clinical researcher and a software developer at the LKSKI to design, implement and evaluate the Wikibreathe tool. In particular, my focus was on developing the usability evaluation process through questionnaires and focus groups.

3.1 Asthma Action Plans

Asthma is a condition that causes occasional tightening of the air passages, which makes it difficult breathing air in and out of the lungs (Levy et al., 2006). Asthma affects approximately 10-15% of children and 5-10% of adults and the prevalence of asthma is increasing world-wide (Sloane et al., 2008). In order to control their asthma, patients need to understand the disease, its triggers, how to recognize worsening status, and what to do in response.

An Asthma Action Plan consists of written management guidelines that provide clear instructions on how to treat asthma when the condition is stable as well as when the symptoms are active (Plottel & Feldman, 2008). Asthma action plans usually use the “zone” system, patterned after the red-yellow-green lights of a traffic signal, to describe the severity of asthma symptoms and to determine the appropriate response (Sloane et al., 2008).

Studies have shown that the use of asthma action plans for self-management of asthma symptoms greatly improves the health of adult patients (Gibson et al., 2003) and significantly reduces the number of acute care visits per child (Zemek et al., 2008). However, despite enthusiastic support from professional medical societies and managed care insurance companies, written asthma plans are not universally popular in the asthma community. The most common reason a person with asthma does not follow an asthma action plan is simply that the physician never provided one (Plottel & Feldman, 2008).
Furthermore, a survey of existing asthma action plans found that there is no consistent version of a plan used by health care professionals. Although the goals of the action plans are essentially the same, many hospitals and states have developed their own versions of the plans with differing format and content.

3.2 The “Wiki” Concept

The purpose of this project is to determine the preferences of three groups of stakeholders (respirologists, patients, and asthma educators/primary care physicians) when designing an asthma action plan. Since the goal is to allow collaboration between participants, the concept of a Wiki was introduced. A “Wiki” is as a collection of Web pages that anyone can edit (Woods & Thoeny, 2007). An example of a Wiki-based application is Wikipedia, a user-contributed online encyclopedia. The same concept can be extended to the tool to be developed which will allow users to collaboratively contribute to the design of a single asthma action plan, thus resulting in a standardized design.

One of the hurdles that needs to be addressed is the issue of allowing users to select format preferences. Websites such as Wikipedia focus entirely on content and therefore the structure of the website is similar to that of a document editing tool. However, since the final product of this tool is the design of an asthma action plan, it needs to incorporate options to support modifying both content (e.g. asthma-related descriptions and instructions) and format (e.g. whether the plan should be oriented in landscape or portrait format and how many zones it should contain).

3.3 Requirements Gathering

Before the design of the Wikibreathe tool (“Wiki” due to its collaborative features+ “Breathe” as it relates to asthma) can begin, it was necessary to gather an exhaustive list of options to be included in the tool. A total of 50 asthma action plans from Canada and the U.S. were analyzed for both content and format differences. Below are several examples of asthma action plans:
Although the purpose of each asthma action plan is essentially the same – to educate asthma patients on actions to take based on the severity of their asthma symptoms – there were many notable differences between plans.

### 3.3.1 Format differences

Format differences ranged from the orientation of the plan and the number of zones it contains to the font style and size. Some notable differences include:

1. Number of Pages used in the plan: 1 page (27), 2 pages (23)
2. Orientation: Landscape (18), Portrait (32)
3. Were images used in the plan? Yes (21), No (29)
4. Overall format: Table (45), Decision Flow (5)
5. For the plans using a table format, the description and instruction sections were arranged: Horizontally (31), Vertically (14)

### 3.3.2 Content differences

There were also many content differences between plans. For example, the simple task of describing a patient who is breathing well produced six different variations:

- Able to do normal activities and/or sports without being short of breath
• Breathing is easy
• Breathing is good
• No difficulty breathing
• No shortness of breath
• Normal breathing

3.3.3 Basic Requirements

After analyzing the differences between plans and gathering an exhaustive list of options, it was decided that the Wiki tool must support the following basic requirements:

1. The tool should have the flexibility to allow users to modify both content and format options as it relates to an asthma action plan.

2. The tool must be able to generate an asthma action plan with the following sections (see figure below):

   a. **Header:** Contains the basic administrative information like a line to write the date and patient name, as well as other pertinent information such as the severity of a patient's condition.

   b. **Zone Description:** This section contains the description of asthma symptoms separated by severity. For example, a description of “breathing” in the Green zone might be “Breathing is easy” whereas in the Red zone the description would be “Very short of breath.”

   c. **Zone Instruction:** This section describes the actions an asthma patient should take when their symptoms match the zone description in the corresponding zone colour. An example of an instruction would be “Take reliever inhaler ____ puffs every 4 to 6 hours.”
d. **Footer:** This section is at the bottom of the plan and may include a space for patient or physician signatures or important statements about asthma.

Figure 3-2: Sections of an asthma action plan: 1) Header, 2) Zone Description, 3) Zone Instruction, 4) Footer

3. The tool should allow the option of changing orientations: Landscape or Portrait.

Figure 3-3: Landscape (left) and Portrait (right) oriented plans
4. The tool should allow the option of having either a 3-zone or 4-zone plan. A 4-zone plan will include an orange zone in addition to the usual colours of red, yellow and green.

5. The list of options should be exhaustive and based on existing asthma action plans that are in use in clinical settings.

6. The plan produced should be readable in black and white since many clinics do not have colour printers.
4 Prototyping

This chapter describes the process of designing and developing the Wikibreathe tool and reports the results from focus groups conducted to evaluate the usability of the tool.

4.1 Initial prototype

An initial prototype was developed based on the format and content options in the requirements gathering phase. The overall concept is that the user will choose from a list of options on the left side of the screen to build up the asthma action plan which will be displayed on the right side of the screen. When an option is selected, the change is immediately reflected on the plan.

The large number of content options will be grouped into high level categories roughly corresponding to the sections of the plan. The categories are:

- Asthma Action Plan Layout
- Basic Setup Choices
- Header
- Asthma Zones (includes Green/Yellow/Orange/Red zones)
- Footer

It was also decided that a complete list of options should be provided in this initial phase, which will then be reduced in subsequent versions of the prototype based on user feedback.
4.1.1 Issues with initial prototype

An informal focus group was conducted with a group of health care researchers to gather feedback on the initial design of the Wikibreathe tool. A major difficulty found with the system was the lack of organization while navigating through options. Since the structure of the three or four zones (e.g. green, yellow or red zones) are almost identical and repeats the same titles of Zone Description, Zone Instructions, Basic Choices and Content Options, users (and even the creators of the system!) would often inadvertently expand menus and select options from the wrong zone section. An example of the navigation structure is shown below:

- Green, Yellow, Orange or Red Zone
  - Zone Description
    - Basic Choices (Defines overall asthma-related terminology)
    - Content Options (Options to describe asthma symptoms)
  - Zone Instructions
    - Basic Choices (Specifies the format of the instructions area)
    - Content Options (Describes responses to asthma symptoms)
Aside from the navigation structure, an additional source of confusion stemmed from having the entire navigation panel in the same font size and style.

Another issue with the initial prototype was that there was no way for users to determine how many options can be added in each zone. For example, there are over ten different categories of options when defining the description for each zone, but the plan only has enough space for 4-5 options to be included. If users selected more options than the allotted space, the additional options would simply “disappear” off the screen.

4.2 Revised Prototype

Many of the issues raised in regards to the initial prototype were addressed in the second version. In particular, the navigation structure was improved by using tabs to represent the high level sections of the asthma action plan with the appropriate menus embedded in each tab.

4.2.1 Appearance

![Figure 4-1: Revised Prototype of the Wikibreathe tool](image)

The Wikibreathe tool consisted of three main sections:

1. Selecting the number of zones and the page orientation: This section allows the user to select the number of asthma zones (3 or 4) and whether the plan should be oriented vertically (portrait) or horizontally (landscape).

2. Selecting Content and Format preferences: This is the main section of the tool which contains
both content (i.e. asthma specific) and format (i.e. appearance specific) choices.

3. Asthma action plan display: This section displays the current state of the asthma action plan and updates automatically when an option is selected.

4.2.2 Format and Content

Under the Zones tab, there are three or four sub-tabs depending on the number of zones chosen for the plan. The sub-tabs are colour-coded to match the zone section on the plan. The content of each tab is further separated into Zone Description and Zone Instruction.

Figure 4-2: Navigation Structure

Basic Choices and Content Options

Basic Choices refers to general options such as setting the overall terminology for asthma symptoms (“Description”) and choosing whether the zone instructions should be displayed in a table format or as a guided prescription (“Instruction”).

Content Options refers to asthma-specific content choices such as whether to display a symptom like breathlessness (“Description”) or how frequently a patient should use a reliever (“Instruction”).
“Don’t Show” and “Show, but phrasing doesn’t matter”

Many of the content options are set to the default option “Don’t Show.” This means that the particular content element will be not be included in the action plan unless you change this option. You can change this option by choosing any of the phrasing options offered in the drop-down list. The drop-down list also contains the option “Show, but phrasing doesn’t matter.” Choose this option to indicate that you would like the particular content element to be included in the asthma action plan, but have no particular preference for how it is phrased (in this case it will be included in the asthma action plan and phrased according to a default setting).

4.2.3 Navigation structure

In addition to Zones, the tabs also include general options (Setup, Other) and options for a specific section in the action plan (Header).

![Figure 4-3: Choosing a tab](image)

To select an option, the user would first click on a tab, then expand a menu and select an option from the drop-down box. The selected option will be automatically reflected in the action plan on the right side of the display.

For example, to add a space at the top of the asthma action plan to write in the patient name, the user would click on Header, and then expand the menu Patient Information.
Figure 4-4: Choosing an option

Once an option is selected, the item appears immediate on the displayed asthma action plan.

Figure 4-5: Options appears on action plan

Items can be easily deselected by choosing the “Don’t create space…” option in the Header section or the “Don’t show” option in the zones section.

Figure 4-6: Deselecting an item in header (left) and in zones (right)
4.2.4 Special Features

Other features implemented in the Wikibreathe tool include the ability for users to comment on option choices, the feature to display descriptions of medical terminology when hovering over certain words, and the ability to review and print the asthma action plan as a PDF document.

![Breathlessness Options](image1)

**Breathlessness Options**

[Don't show]

**hide comments**

Type a comment here

Post Comment

**Figure 4-7: Adding comments**

![Wheezing Options](image2)

**Wheezing Options**

[Don't show]

**show/make**

Wheezing is a high-pitched sound that patients with asthma often make when breathing out, especially when they are having an asthma attack.

**Figure 4-8: Displaying terminology with hover**

![PDF Save/Preview](image3)

[Generate a new PDF preview]

View previously generated PDFs

**Figure 4-9: Generating a PDF**

4.2.5 Space Limitations

Since there are many content options, the issue of space limitation on the action plan was an important consideration.

**Shifting Zones**

One feature implemented to address this issue was that the size of the zones would shift depending on the number of items in the zone. For example, if there are five options in the Green zone and another one is added, the Green zone would automatically become larger to accommodate the additional item.
User Guide to Number of Options

Another feature that was implemented to address the space limitation was a pop-up box that would appear if the user hovers over a plan orientation icon. In the example below, the maximum total number of options allowed by the space on this particular plan orientation is 21 to 24 items. Note that this is an approximate number based on the orientation and number of zones chosen.

![Figure 4-10: Maximum number of options](image)

Error message

Finally, if the user exceeds the maximum amount of space allowed on the plan, an error message will appear. The user must then remove one or more items from the plan in order to create more empty space before adding the next option.

![Figure 4-11: Space limitation alert message](image)
4.3 Summary

The development of the Wikibreathe tool followed an iterative process over several months. The resulting tool contained an exhaustive list of choices for both content and format based on options found during the requirements gathering phase.

The results from all three stakeholder groups (Respirologists, Primary Care Physicians/Asthma Educators, and Patients) in the focus groups can be found in Appendix B.
Chapter 5

5 Methodology

This chapter discusses the process used to evaluate the Wikibreathe tool using two different usability evaluation methods: A traditional task and scenario-based approach (“Forward Flow”) and the reverse engineering method of recreating a completed plan (“Reverse Flow”).

5.1 Participants

Twelve subjects (4 Male, 8 Female) between the ages of 20-40 participated in this study. Their occupations ranged from high school teacher and music therapist to programmer and engineer. Each participant was asked to fill out a survey in regards to their technical abilities and attitude toward computers.

From the survey, it was noted that the participants spent an average of 9 hours each week on the internet at work and 10.4 hours each week on the internet at home.

On average, participants had a generally positive attitude toward computers and almost all felt confident in their ability to master new skills on the computer (based on Likert scale of 1 to 5 where 1=Strongly Disagree and 5=Strongly Agree):

<table>
<thead>
<tr>
<th>Statement</th>
<th>Average Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>I use computers only because they are necessary for work.</td>
<td>2.3</td>
</tr>
<tr>
<td>Computers have a positive impact on my quality of life.</td>
<td>4.6</td>
</tr>
<tr>
<td>I find dealing with computers to be frustrating.</td>
<td>1.9</td>
</tr>
<tr>
<td>I am confident in my ability to master new skills with computers.</td>
<td>4.5</td>
</tr>
</tbody>
</table>

Table 5-1: Participants' Attitude toward Computers

Sampling Strategy

Participants were selected based on their experience and attitude toward technology. Since the tasks in this experiment did not require a medical background to complete, it was not necessary to recruit physicians and/or patients in this process. Aside from the medical difference, the participants in this study represented the general population of novice users who would use the Wikibreathe tool.
5.2 Process

Participants were first introduced to the goal of the project and then given a brief training on the Wikibreathe tool. Each participant received the same amount of training in the same way to avoid skewing the results.

They were then asked to build an asthma action plan using the Wikibreathe tool based on given specifications. Each user first performed the task using either the Forward or Reverse method and then repeated the task again using the alternate method.

The order of tasks was randomized such that six participants did the Forward Flow first and six did the Reverse Flow first. The experiment instructions can be found in Appendix F.

5.2.1 Forward Flow

The Forward Flow task used the traditional task and scenario-based approach. Participants were told to imagine that they were respirologists who were designing an asthma action plan. They were then given a set of written instructions representing the specifications of the plan that this “Respirologist” would prefer to have and told to create the plan that most closely match these specifications.

5.2.2 Reverse Flow

In the Reverse Flow task, the participant was simply given a completed asthma action plan (that was previously created using the Wikibreathe tool) and was told to recreate it using the tool as closely as possible.

5.3 Selection of Representative Tasks

A list of subtasks was generated based on the most important features of the tool. A complete asthma action plan can be created through these subtasks, from selecting the initial layout to filling in zone-specific content. Additional subtasks related to special features of the tool were also added, such as the ability to set specific terminology to be used throughout the tool.

Since participants were to use the same tool in each evaluation method, it was decided that two different asthma action plans should be used to reduce the learning effect of having to create the same plan twice. A great deal of consideration was given to make sure that these two plans are
different but have similar components so that the results are comparable. The subtasks listed in Table 5-2 were chosen to represent significant features in the Wikibreathe tool.

Some subtasks were not included in both the Forward and Reverse flows. For example, in the Forward flow, the subtask to add an additional statement was not included since it only required that the user make a simple selection. Similarly, the subtask to modify the zone title in the Reverse flow was not included because the task was straightforward.

The Zone Instruction subtasks also differ in that they used the two different formats for instructions: Table Format and Guided Prescription.
<table>
<thead>
<tr>
<th>Description of Subtask</th>
<th>Forward Flow</th>
<th>Reverse Flow</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selecting the orientation of the plan</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Changing the overall terminology in the plan</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Adding an item in the Header (exact match such as patient name)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Adding an item in the Header (not an exact match such as asthma severity)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Adding an additional statement</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Adding the peak flow parameters</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Modifying zone titles</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Changing overall terminology for asthma specific text</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Adding descriptions in each zone</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Adding instructions in each zone (table format)</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Adding instructions in each zone (guided prescription format)</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Adding an additional statement from the “Other” tab</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

**Total number of Subtasks**

<table>
<thead>
<tr>
<th></th>
<th>Forward Flow</th>
<th>Reverse Flow</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

*Table 5-2: Inclusion of subtasks for each evaluation method*
The two plans generated by the Forward Flow and Reverse Flow methods are shown below. A respirologist was consulted and determined that these were reasonable representations of complete asthma action plans.

Figure 5-1: Asthma action plans generated by the Forward Flow (left) and Reverse Flow (right)

5.4 Quantitative Measurements

The following aspects of the experiment were recorded for each task:

Time

The total amount of time each subject took to complete the task was recorded using a stopwatch. Each section (Setup, Header, Zone Description, Zone Instruction, and Footer) was also timed individually for more accurate results.
Errors

The percentage of errors made in each task was calculated. The decision was made to use the percentage instead of the number of errors because the Forward Flow task required 30 steps to complete whereas the Reverse Flow task required 36 steps; therefore, the percentage of errors proportional to the number of steps would be a more accurate measurement than the number of errors.

For example, a complete asthma action plan for the Reverse Flow required that the user correctly complete 36 subtasks. If two of these subtasks were incorrectly performed or missed, the error percentage would be calculated as $\frac{2}{36} \times 100 = 5.55\%$. In other words, each “miss” of a subtask would count as an error.

5.5 Qualitative Measurements

The following qualitative measurements were recorded:

Observations and Think Aloud

Participants were encouraged to verbalize their thoughts while performing the tasks. The moderator took note of any comments as well as any perceived difficulties the participant had while using the Wikibreathe tool.

Questionnaire

A short questionnaire was administered to the participant after each task. The questionnaire was based on the System Usability Scale and gathered feedback from the participant on the usability of the Wikibreathe Tool. A few additional questions were also added to determine the participant’s attitude towards the specific task (for example, if they felt confident or frustrated while using the tool). The questions on this survey were developed to analyze how much users enjoyed performing a task using the specific method (Forward or Reverse).

Semi-Structured Interview

After all the tasks were completed, a semi-structured interview was conducted with each participant. The interview was audio-recorded and lasted approximately 10 minutes. The
interview began with a review of any difficulties the participant had with the tasks. This included issues raised by the participant as well as issues noted by the observer.

All participants were then asked the following series of questions:

1. You did this task twice, once to follow a set of written instructions and once to recreate a completed plan – which one did you enjoy more and why?
2. Which method did you feel was better in helping you learn how the tool works? Why?
3. If you were running a usability test, which method would you choose to use or think would be more effective in finding usability problems with the tool?
4. What did you like most about the tool?
5. What did you dislike most or find most confusing about the tool?
6. Which part was hardest to use / most difficult?
7. In what ways would you improve the design of the tool?
Chapter 6

6 Usability Test Results

This chapter will describe the findings from the experiment described in Chapter 5. It begins by reporting the amount of time participants spent in completing each task as well as the percentage of errors committed. It then describes the results from the questionnaires participants filled out after each task. The last section of this chapter discusses the differences in the types and number of usability problems encountered.

6.1 Time and Errors

In each task, the amount of time spent in completing a specific section and the total percentage of errors were recorded.

The statistical significance of the differences in times and errors between the two usability testing methods was assessed using repeated measures analysis of variance. Prior to the analysis, response times were subjected to natural log transformation in order to record for positively skewed response times. Similarly, the error proportions were subjected to arcsin transformation as recommended by Snedecor and Cochrane (2000). There was a significant multivariate effect across all the dependent measures (the response times for the six different subtasks, plus the proportion of errors (as assessed by Wilk’s Lambda) for the Task main effect (F[7,4]=11.27, p<.05). However, although not significant, the multivariate effect for the task by order interaction had a large effective size of 0.8 suggesting that it too would have been significant in a larger study with more participants.

For the task main effect, the zone description time and the total response time were significantly different (p<.05),

Since the task by order effect interaction had a large effect size, the univariate effects were also examined. The following measures were significantly affected by the task x order interaction (p<.05): Header response time; zone description response time; footer response time; total response time; error proportion.
Comparison of Median Times based on Order of Tasks

The graphs on the next page plot the median times for each section separately based on the order of tasks. Protocol “A” refers to participants who were asked to first perform the forward task and then to perform the reverse task whereas Protocol “B” refers to participants who were asked to first perform the reverse task and then perform the forward task.
Overall, users spent the least time working on the Setup and Header sections and the most time on the Zone Description and Zone Instructions sections. It is also interesting to note that the total time spent is much greater for the Reverse task compared to the Forward task, particularly when the Reverse task is done first.

Figure 6-1: Comparison of Median Times for Forward and Reverse Flows for Protocol A (left) and Protocol B (right)
6.1.1 Order Effects

The following figures show the task by order interaction effect on the Zone Description and Footer response times respectively. In both cases, participants completed each section faster as the second task compared to the time they required for the first task or method they were exposed to.

**Figure 6-2: Order Effect for Zone Description**

**Figure 6-3: Order Effect for Footer**
The benefit of going second was likely greatest for the subtasks that were most similar across the two methods. For example, in the Zone Description section, the subtasks consisted of adding asthma symptoms and peak flow parameters. In the Footer section, the subtasks consisted of adding a section for triggers and a signature for the patient or physician. Since the tasks were similar (even though the content added was different), the second performance of the task was speeded up.

On the other hand, other sections such as the Zone Instruction differed slightly for the Forward and Reverse tasks. For instance, in the Forward task, the user was asked to create the zone instruction in a table format whereas in the Reverse task, the user was shown a paper plan with a zone instruction in a guided prescription format. In this case, carry-over from the first performance of the task did not apply as readily.

6.1.2 Greater Amount of Time and Errors for Reverse Task

In terms of the total amount of time, the reverse engineering task took longer to complete (around 13 minutes versus around 10 minutes for the forward task) and there was also a non-significant (p=.115) tendency for participants to commit a greater percentage of errors while doing the reverse task (9% vs. 6%).

The higher error rate and longer performance time in the reverse task was likely due to increased task demands of that task. In the Reverse task, participants had to recreate the plan to look exactly like the one given, whereas in the Forward task, participants had a small degree of freedom to make certain decision, for example, whether the plan should be in portrait or landscape format.

The more rigid requirements caused participants to go through the tool in more depth (more time) and through more paths (more errors). One participant commented in regards to encountering more errors during the Reverse task:

"[The Reverse task] was more frustrating in that it was harder and so there you can see more of where "red flags" come out."
Nine out of the eleven usability problems uncovered by this experiment (described in more detail in section 6.3) were found due to user error. The remaining two usability problems are overall suggestions for the tool. Therefore, in this case, a higher percentage of errors translated into a greater number of usability problems found.

6.2 Usability Questionnaire

Immediately after each task, participants were asked to complete a questionnaire about the usability of the tool and their attitude toward that particular usability testing method.

6.2.1 Questionnaire Results

System Usability Scale (SUS)

The first part of the questionnaire is based on the System Usability Scale. This questionnaire uses a 5-point Likert scale where 1 is Strong Disagree and 5 is Strongly Agree. Repeated measures ANOVA was carried out on the SUS data. Using an alpha level of .05, and adopting an exploratory approach (with no correction for family-wise alpha), significant differences between the tasks were found on SUS2 only. However, the first three SUS items were significantly affected ($p<.05$) by the task by order interaction effect: SUS1; SUS2; SUS3.

The average scores are listed in the table below:

<table>
<thead>
<tr>
<th>Usability Statement</th>
<th>Average Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUS1 I think that I would like to use this system frequently.</td>
<td>3.3</td>
</tr>
<tr>
<td>SUS2 I found the system unnecessarily complex.</td>
<td>2.8</td>
</tr>
<tr>
<td>SUS3 I thought the system was easy to use.</td>
<td>3.6</td>
</tr>
<tr>
<td>SUS4 I think that I would need the support of a technical person to be</td>
<td>1.8</td>
</tr>
<tr>
<td>able to use this system.</td>
<td></td>
</tr>
<tr>
<td>SUS5 I found the various functions in this system were well integrated.</td>
<td>3.9</td>
</tr>
<tr>
<td>SUS6 I thought there was too much inconsistency in this system.</td>
<td>1.7</td>
</tr>
<tr>
<td>SUS7 I would imagine that most people would learn to use this system</td>
<td>3.5</td>
</tr>
<tr>
<td>very quickly.</td>
<td></td>
</tr>
</tbody>
</table>
SUS8 | I found the system very cumbersome to use. | 2.4  
SUS9 | I felt very confident using the system. | 3.5  
SUS10 | I needed to learn a lot of things before I could get going with this system. | 2.2

**Table 6-1: Average Ratings on the System Usability Scale**

Overall, users had a mildly positive response to the Wikibreathe tool where ratings generally fall in the neutral range of 2.0 – 4.0. Scores outside the neutral range showed that in general, users did not feel that they need the support of a technical person to be able to use this tool and also that there was not too much inconsistency in the system.

**Attitude toward Usability Method**

The next set of statements on the questionnaire dealt with the participants’ attitude (positive or negative) toward the particular usability evaluation method used to accomplish the task. Again repeated measures ANOVA was carried out. Neither the main effect of task nor the task by order interaction effect was significant on any of the five attitude measures.

<table>
<thead>
<tr>
<th>Statement</th>
<th>Forward Task</th>
<th>Reverse Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>I feel confident that I could this tool in the future to create an effective asthma action plan.</td>
<td>4.4</td>
<td>4.4</td>
</tr>
<tr>
<td>I enjoyed completing the task assigned using this tool.</td>
<td>4.3</td>
<td>4.1</td>
</tr>
<tr>
<td>I felt lost and frustrated while using the tool to complete this task.</td>
<td>1.8</td>
<td>2.1</td>
</tr>
<tr>
<td>I felt that the instructions to complete the task were clear and straightforward.</td>
<td>4.2</td>
<td>4.3</td>
</tr>
<tr>
<td>After completing the tasks I was asked to do, I have a better idea of how the tool works.</td>
<td>4.3</td>
<td>4.4</td>
</tr>
</tbody>
</table>

**Table 6-2: Average Ratings on Attitude toward the Usability Testing Methods**
Thus, the Tool Attitudes don't appear to differ between the Forward and Reverse tasks even though results showed that participants spent more time on the Reverse task and committed more errors. Overall, participants had a positive attitude toward both tasks.

There is a slight difference in the rating for the statement “I felt lost and frustrated while using the tool to complete this task” in that those who had just completed the task using the Forward method had an average rating of 1.8 which is slightly less than the average of 2.1 for those who had just completed the task using the Reverse method. The difference, however, is not statistically significant and therefore should be treated as simply an overall observation.

### 6.2.2 Differences in Usability Ratings due to Order of Tasks

Participants who performed the task using the Reverse method first felt that the system was less complex and easier to use (SUS2 and SUS3) when they performed the second task using the Forward method whereas those who did the task using the Forward method first did not show a difference in their attitude toward the tool in the second task.

![Bar chart showing comparison of average scores in participant attitude toward tool (SUS2)](image)

**Figure 6-4: Comparison of average scores in participant attitude toward tool (SUS2)**
I thought the system was easy to use.

![Comparison of average scores in participant attitude toward tool (SUS3)](image)

**Figure 6-5: Comparison of average scores in participant attitude toward tool (SUS3)**

The results showed that users who used the reverse engineering method to recreate an asthma action plan perceived the tool as more complex and difficult to use.

### 6.3 Usability Problems

A total of 11 usability problems were compiled using observation notes during the experiments, verbal feedback from participants and the asthma action plans produced through the two tasks. These usability problems are summarized in the table below:

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Severity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Peak Flow</td>
<td>The option to add Peak Flow parameters to the Zone Descriptions is located in the Setup tab whereas users expected it to be in the Zone tab and therefore had trouble finding it.</td>
<td>High</td>
</tr>
<tr>
<td>2 Default Options</td>
<td>Two of the Header options are pre-selected in the plan (Patient Name and Asthma Plan Title). Users did not notice that the options were already selected and therefore deselected them by accident.</td>
<td>Low</td>
</tr>
<tr>
<td>3 Guided Prescription</td>
<td>In order to add specific options in the Zone Instructions section, the Guided Prescription</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Terminology</td>
<td>In order to match the appropriate terminology for Asthma Symptoms in the plan, users must go to Zone &gt; Green &gt; Basic Choices and change the default wording. They were not able to find this selection.</td>
</tr>
<tr>
<td>---</td>
<td>-------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>Layout</td>
<td>The wrong layout (portrait vs landscape) was selected. Users did not recognize the proper layout based on the small layout icons in the zone orientation section.</td>
</tr>
<tr>
<td></td>
<td>Submenus</td>
<td>In order to add puffer colours or checkboxes to the Zone Instructions, users must select the option in a submenu in the upper level Basic Choices. They were not able to find this option.</td>
</tr>
<tr>
<td></td>
<td>Search</td>
<td>Users tried to use Ctrl-F to perform a search on the menu options in the application. This feature is not supported.</td>
</tr>
<tr>
<td></td>
<td>Basic Choices vs Content Options</td>
<td>Both Zone Description and Zone Instruction sections are divided into the Basic Choices and Content Options menus. Users were confused as to the purpose and distinction between the content in the two menus.</td>
</tr>
<tr>
<td></td>
<td>Hidden Triggers</td>
<td>In order to select specific asthma triggers, users must first choose the location of the asthma triggers box as well as whether it should be displayed as text, figures, or text and figures. They did not realize that the asthma triggers were by default hidden under the previous options were selected.</td>
</tr>
<tr>
<td></td>
<td>Font Size</td>
<td>Users mentioned that the Font Size used in the application is too small.</td>
</tr>
<tr>
<td></td>
<td>Additional Statements</td>
<td>Due to the large number of options in the Other &gt; Additional Statements menu, users were not able to find the option needed to match the plan.</td>
</tr>
</tbody>
</table>

Table 6-4: Usability Problems
The table below summarizes the number of usability problems found during each task. Only problems found the first time were included. For example, if problem #1 was encountered in both tasks, it would be counted only as a problem in Task 1. The complete list of usability problems can be found in Appendix D.

<table>
<thead>
<tr>
<th></th>
<th>Protocol A (Forward &gt; Reverse)</th>
<th>Protocol B (Reverse &gt; Forward)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1</td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>Task 2</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total number of usability problems found</strong></td>
<td><strong>9</strong></td>
<td><strong>11</strong></td>
</tr>
</tbody>
</table>

Table 6-5: Number of Usability Problems Found for the First Time during each task

When users performed the task using the Forward method first, there were less total usability problems found (9). On the other hand, users who performed the task with the Reverse method first found more usability problems in total (11), and in fact, all the problems were found during the first task!

**The “Suppression” Effect of Task-based Testing**

In comparing the two usability evaluation methods, it was found that users who did the Reverse Engineering task first found two additional problems that were not found by the other group of users who did the task-based method first. The two problems that were missed are:

1. **Layout (#5):** This usability problem relates to users not being able to distinguish the layout of the asthma action plan using the small icons on the top of the screen (Figure 8-6). It is not entirely clear why this usability problem was not apparent to the group of users who did the Forward task first. A possible reason is that they had learned how to use the tool through following the written instructions in the first task and when they got
to the second task, they’ve simply adapted to how the tool works.

**Figure 6-6: Selecting the layout of the asthma action plan**

2. **Font Size (#10):** It was observed on several occasions that participants who were following the written instructions (“Forward Flow”) focused more on the paper than the computer screen. A typical approach would be to read the instructions on the page, glance up quickly to select the option on the screen, and then return the attention to the paper. Participants in this group spent much less time exploring the tool and therefore did not report a problem with the size of the fonts used in the tool.

**Severity of Usability Issues**

Although the Reverse method was able to uncover two more usability problems, it is important to note that these two problems were rated as “Low” and “Medium” severity. Therefore, the group that performed the task using the Forward method first was able to find the same number of High Severity problems as the other group.

**Reverse Engineering Method finds more Usability Problems the First Time**

During the post-experiment interview, several participants commented that it was more difficult to create the plan using the reverse engineering method but it also encouraged them to explore the tool in more depth. By exploring the tool, these users encountered many of the errors that the other group did not. This would correspond with the results of the questionnaire which indicated that users rated the task as more confusing and frustrating when they had used the reverse engineering method. One of the subjects mentioned that "[The Reverse task] was more frustrating in that it was harder and so there you can see more of where "red flags" come out." In this case, the “red flags” referred to the design flaws in the Wikibreathe tool.
On the other hand, users who followed the written instructions first had commented that it helped them “learn” how to use the tool. Unfortunately, this meant that when they switched to the reverse engineering task, they had already been accustomed to the design of the tool and therefore did not explore the tool in as much depth.

6.4 Summary

In summary, the results of this experiment showed that the reverse engineering method is able to uncover as many, if not more, usability problems compared to the traditional task-based testing approach. Overall, users had a positive attitude toward the new usability method.

Order effects were also noted through statistical analyses, indicating that users were able to learn how to use the tool after performing the task the first time. However, the exact effect of learning cannot be measured due to the design of the experiment.
Chapter 7

7 Discussion

This chapter proposes some changes to the Wikibreathe tool, discusses the implications of the experiment results and provides a mapping of the advantages and disadvantages of each usability evaluation method used in this study.

7.1 Repeated Measures ANOVA vs T-Tests

It was noted that since this study had two groups of participants, a series of t-tests could have been done in place of repeated measures analysis of variance. However, the decision was made to analyze the data using repeated measures ANOVA instead of t-tests since this is an exploratory study and therefore ANOVA provides the following benefits not inherent to t-tests:

1. **Reduction of Type-I Errors**: As multiple t-tests are done, the likelihood of finding a significant result by chance (Type-I Error) increases. ANOVA reduces this probability by comparing all the means with each other at once, thus retaining the same alpha.

2. **Controlling Factors**: ANOVA can be used to test each factor while controlling for all others and therefore is more statistically powerful than the simple t-test.

3. **Interaction Effects**: ANOVA can be used to detect interaction effects between variables and provide a more complete analysis of the data.

Therefore, although similar conclusions can be achieved through multiple paired t-tests, the use of repeated measures ANOVA provided a broader platform for exploration of the data that otherwise could not be done.

7.2 Comparison of SUS scores

In order to analyze the validity of the conclusions based on the SUS scores recorded in this study, it is important to understand how these scores compare with those in other studies. Figure 7-1 shows the frequency distribution of SUS scores from 50 studies (Bangor, 2008). The score is
calculated by extrapolating the SUS score from a Likert scale of 1-5 to a score from 0-100. For example, a rating of 4/5 would be scored as 80.

![Frequency Distribution of SUS Scores for 129 Conditions from 50 Studies](image)

**Figure 7-1: Frequency Distribution of SUS Scores (Bangor, 2008)**

The most frequent average SUS score was 71-80, or 3.55 to 4.00 on a Likert scale from 1-5. A check of the SUS scores recorded in this study shows that the range is from 64 to 86 with an average of 73.8 (Table 7-1). This falls into the average range of SUS scores and therefore, the results, although positive, cannot be considered highly significant.

<table>
<thead>
<tr>
<th>Usability Statement</th>
<th>Average Rating</th>
<th>Converted SUS Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUS1 I think that I would like to use this system frequently.</td>
<td>3.3</td>
<td>66</td>
</tr>
<tr>
<td>SUS2 I found the system unnecessarily complex.</td>
<td>2.8</td>
<td>64</td>
</tr>
<tr>
<td>SUS3 I thought the system was easy to use.</td>
<td>3.6</td>
<td>72</td>
</tr>
<tr>
<td>SUS4 I think that I would need the support of a technical person to be able to use this system.</td>
<td>1.8</td>
<td>84</td>
</tr>
<tr>
<td>SUS5 I found the various functions in this system were well integrated.</td>
<td>3.9</td>
<td>78</td>
</tr>
</tbody>
</table>
Table 7-1: Converted SUS Scores

<table>
<thead>
<tr>
<th>SUS6</th>
<th>I thought there was too much inconsistency in this system.</th>
<th>1.7</th>
<th>86</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUS7</td>
<td>I would imagine that most people would learn to use this system very quickly.</td>
<td>3.5</td>
<td>70</td>
</tr>
<tr>
<td>SUS8</td>
<td>I found the system very cumbersome to use.</td>
<td>2.4</td>
<td>72</td>
</tr>
<tr>
<td>SUS9</td>
<td>I felt very confident using the system.</td>
<td>3.5</td>
<td>70</td>
</tr>
<tr>
<td>SUS10</td>
<td>I needed to learn a lot of things before I could get going with this system.</td>
<td>2.2</td>
<td>76</td>
</tr>
</tbody>
</table>

This is considered a limitation of this study and poses difficulties when trying to generalize results as positive or negative.

7.3 Recommended Changes to the Wikibreathe Tool

This method of reverse engineering was developed at a latter stage of the Wikibreathe development process. Due to time constraints, not all the recommended changes can be applied to this iteration of development. In the future, these types of usability test methods can be integrated earlier into the development cycle when the tool can be more easily modified.

However, the results of this study did generate many design recommendations for changes to the Wikibreathe tool based on the results from the focus groups, questionnaires, heuristics evaluation, and usability testing. Below are four that were particularly interesting ideas:

1. **Reorder Asthma Symptoms (Focus Group):** To be able to drag items to the preferred spot directly on the asthma action plan display. Previously, the order of how the asthma symptoms appear on the plan cannot be changed and some participants in the focus groups noted that it is an important feature to be able to reorder the symptoms. Being able to manipulate the items directly on the asthma action plan display also removes the
hassle of finding and choosing the option in the navigation menus.

You Feel Good

**Description**  
You have ALL of the following:

- Breathing is easy
- No cough
- No chest tightness
- Can do daily activities without difficulty

Figure 7-2: Reordering asthma symptoms directly on asthma action plan

2. **Search Capabilities (Usability Testing):** Due to the large number of options, it is at times difficult to find one particular option within the Wikibreathe tool. Several users have suggested that a search function can be implemented either as an extension to the search from the browser (ex. Ctrl-F), as an open text box or as a voice command (“I use [the Ctrl-F function] a lot when I’m on a website and I can’t find something.”). This search function would search through the entire database of options and present a separate screen to the user with a list of options matching the search criteria.

3. **Change Location of Peak Flow and Terminology (Usability Testing):** During the User Testing phase, participants were constantly confused as to where the Peak Flow parameters can be selected. Since Peak Flow is an indication of asthma symptoms, users naturally assumed that it would be under the Zones tab. However, it is actually located in the Setup tab. One user mentioned in the post-experiment interview, “I expected [the peak flow option] to be in the Zones tab, not Setup.” Another user said, “By the time I got to the Zones section, I didn’t think of going back to find it under Setup.” Therefore, the recommended change is to move the Peak Flow parameters into the Zones tab so that it is more consistent with the other zone-related options.

The process of changing the overall terminology of the tool was also a source of confusion since some of the terminology can be set in the Setup tab and others in the Green Zone. The recommended change is to move the Terminology options to the Setup
tab so that there is more consistency across the navigation structure.

![Figure 7-3: Peak Flow and Terminology]

4. **Management of Space Limitations (Focus Group, Heuristic Evaluation):** The space available on one page of an asthma action plan is a limiting factor to how many options a user can select. The problem occurs when a user had selected too many options in a previously zone and runs out of space at the bottom of the plan. The program shows an error message box but leaves it up to the user to figure out how to resolve the problem. The recommended change would be to display a list of previously selected options when users encounter a space limit and then to allow users to deselect any option in the list.

These recommendations and other changes will be addressed in the next version of the Wikibreathe tool.
7.4 Reverse Engineering as a Usability Evaluation Method

The purpose of this study was to evaluate whether a usability test method based on the concept of “Reverse Engineering” can be a viable way to assess the usability of systems. Some interesting insights were attained through the results of an experiment comparing the method of reverse engineering with that of the traditional task and scenario-based method of testing.

1. Reverse Engineering Uncovers Usability Problems

The results of the experiment showed that tasks performed using the reverse engineering method generally took longer to complete and users also committed more errors. Since many of these errors are related to design flaws in the system, reverse engineering is able to uncover many of the usability problems in the tool.

2. Reverse Engineering Testing is Easy to Prepare

The task of preparing for the reverse engineering section of this experiment simply consisted of creating an asthma action plan using the Wikibreathe tool. In comparison, a detailed set of instructions and a moderator script were required as preparation for the task-based test.

3. Reverse Engineering is an Enjoyable Task

Even though the reverse engineering task took longer and caused users to commit more errors, their attitudes toward the tool and usability method were consistently positive. One user even commented that she enjoyed the visual aspect of being able to look at something and then to recreate it.

4. Reverse Engineering may be a Good Learning Tool

Some learning effects were observed during the experiment (i.e. users seem to be faster at using the Wikibreathe tool to do the task the second time) but results were inconclusive due to the fact that only one version of the tool was available when designing the experiment. A suggestion for an experimental design to reduce the learning effect through two versions of the tool is proposed in Chapter 8.
7.5 Mapping of Usability Methods and Requirements

The following table gives a mapping of the advantages and disadvantages of each usability evaluation method used in this research study. Note that in any given study, the preferred approach would be to use a combination of methods from each of the three categories of methods (Inquiry, Inspection, and Formal Usability Testing) so that results can be duplicated and validated against other methods.

<table>
<thead>
<tr>
<th>Focus Groups and Questionnaires</th>
<th>Task-based Usability Test Method</th>
<th>Reverse Engineering Method</th>
<th>Heuristic Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low cost to implement</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Usable for different systems</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Quick to prepare test</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Quick to administer test</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Able to find usability problems due to user error</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>User involvement</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Exploration of new features</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Can be used in any phase of a project</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Does not need to involve user</td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

Table 7-2: Mapping of Usability Methods and Requirements
Chapter 8

8 Conclusion

This study examined whether reverse engineering can be a viable usability evaluation method in comparison with traditional scenario-based techniques. This concept of reverse engineering stems from software engineering but its idea of taking a completed product, analyzing its parts, then attempting to use a tool to recreate it applies as a way of evaluating the ease at which novice users can use a tool to recreate a product. Design flaws in the tool that cause difficulties with users are identified as usability problems. An additional advantage of using this method is the simplicity in how the test can be set up – one would simply need to create a product using the tool that is to be tested then ask a subject to recreate it using that same tool.

An experiment was conducted in which the usability of the Wikibreathe tool was tested using both the reverse engineering and task-based methods. Twelve subjects participated in this experiment, each creating two asthma action plans in a randomized order of which method is used first. Results showed that although the task using the reverse engineering method took a longer time to complete and with more errors committed, the users’ attitudes toward the test process remained positive. Furthermore, the reverse engineering task was able to find more usability problems overall compared to the task-based approach.

In addition to usability testing, focus groups and questionnaires were also used to gather qualitative feedback and a heuristic evaluation method was used by the researcher to assess the usability of the tool. In all, the techniques employed in this study spanned the three main categories of usability evaluation methods as defined by Battleson (2001): Inquiry (focus groups and questionnaires), Inspection (heuristic evaluation), and Formal Usability Testing (reverse engineering and task-based usability testing).


8.1 Contributions

In this section, the primary contributions of this research are outlined.

First, this study introduces the concept of Reverse Engineering as a viable usability evaluation method. This is a novel approach in the field of human factors engineering and has a great deal of potential as a technique to evaluate products in a fast and simple way.

Second, results of the experiment showed that the reverse engineering method was able to uncover more usability problems in the Wikibreathe tool than the task-based method.

Finally, empirical observations were reported based on the experience of participants in the experiment. The results showed that users had a positive attitude toward using the reverse engineering method to recreate an asthma action plan and some, in fact, preferred this method over the task-based approach due to the fact that they can visualize the final product.

8.2 Limitations

A few limitations that may have impacted the conclusion of this study are noted in this section.

First, due to technical limitations, only one version of the Wikibreathe tool was used in the experiment. Because of this, each participant performed the task twice using the same tool and therefore the learning effect cannot be properly measured. The task itself was also quite simple and the slight differences in the two tasks might have affected the results (e.g. The Reverse subtasks were inherently more difficult and therefore the total time was higher compared to the Forward subtasks). The small sample size (N=12) was a practical limitation and may pose difficulties when generalizing results. There was also limited functionality that can be tested in the Wikibreathe tool since it was considered a prototype tool.

Secondly, the learning styles of the individual users were not taken into account when they were randomly assigned to one of the two groups. For example, those who were visual learners may do better at the Reverse task whereas others who learn faster through written instructions may do better at the Forward task. Therefore, the results may be skewed depending on the separation of learning styles in the two groups.
Finally, the type of system that can be evaluated by the reverse engineering method is limited to those that can produce a final product. For example, the reverse engineering method can be used to test a word or graphics editor but cannot be extended to, say, flying an airplane. Also, the method of reverse engineering can only be used when the tool has matured to a point where a final product can be developed and therefore it is limited to the latter part of the development cycle.
8.3 Future Work

This section discusses the next steps in the evaluation of reverse engineering as a usability test method.

**Experimental Design**

The next study should improve upon the experimental design used in this study in order to reduce the impact of learning on the results. For example, instead of having only one tool, it would be preferable to get two different versions of the tool, randomized so that users will use one version with each asthma action plan. The design of a proposed experimental design is outlined in Appendix H.

**Evaluating Other Types of Products**

In addition to improving the experimental design, the study of reverse engineering can also be tested with other types of products such as word processing and graphic design editors.

**Closing Remarks**

In closing, the initial results of this research suggest that the technique of reverse engineering as a usability evaluation method poses great potential in uncovering usability problems. Further research can be done in this area to solidify the test model and validate the results in this study.
References


Appendix A: Consent Form

Consent form for participation in the experiment

“Designing Asthma Action Plans using the Wikibreathe Tool”

Investigators: Flora Wan and Professor Mark Chignell

**Introductory Information:** I have been invited to participate in the experiment “Designing Asthma Action Plans using the Wikibreathe Tool”. The purpose of this study is to examine the usability of the Wikibreathe Tool for designing asthma action plans and the differences in methods of learning.

**What Will I be Asked to Do:** I understand that, as a participant in this experiment, I will be required to design two asthma action plans based on a set of requirements. Upon conclusion of the experiment, I will answer some questions concerning the interface and methods used in the experiment.

**Risks and Benefits:** I understand there is minimal risk undertaken by participants in this experiment. I also understand that my participation in this experiment will provide me with minimal direct benefit other than the opportunity to explore an online tool, and a small payment.

**Compensation:** I understand that I will be given a $15 Starbucks or Tim Hortons gift card for participating and that there will be no academic consideration for participation in the experiment. The experiment should take about an hour to complete.

**Access to Information:** The research team, Flora Wan and Professor Mark Chignell, will have access to the data. Questionnaires and experimental data will be kept for the period of approximately 1 year.

**Confidentiality and Publication of Results:** I understand that the results of my participation in this experiment will be kept confidential and that there will be no possibility that I can be identified from or associated with any results or summaries of results, as presented or published subsequent to this experiment.

**Contact Information:** Please contact Flora Wan for further information.

Phone: (905) 731-9647   Email: flora.t.wan@gmail.com
I have read the information provided to me on this experiment and I hereby consent to participate in the experiment “Designing Asthma Action Plans using the Wikibreathe Tool”. The objectives, methods, tasks and procedures have been thoroughly explained to me and all of my questions and concerns of the experiment have been answered completely to my satisfaction. I have the right to withdraw from this experiment at any point in the experiment without penalty, and to request that my data be destroyed.

Participant’s name: __________________________________________

Participant’s signature: _______________________________________

Date: _________________________________________________________
Appendix B: Focus Group Survey Results

Participants

It was decided that in order to clearly evaluate the preferences of the stakeholders and to avoid bias, each stakeholder group would be evaluated separately. Moreover, due to privacy concerns, patients were evaluated individually in the form of an interview instead of a focus group. In total, four respirologists, seven primary care physicians / asthma educators and five patients participated in the focus group studies.

Methodology

The focus group began by giving the participant an overview of the Wikibreathe project and some brief training on the important features of the tool. The participants were then given 45 minutes to create an asthma action plan using the Wikibreathe tool based on their own preferences of what they would like to be included on such a plan.

After creating the action plans, each participant was asked to fill out a questionnaire about the usability of the tool and was then led in an open discussion by the moderator.

Plans created during focus group

The following are examples of plans created by: 1) Respirologists, 2) Primary Care Physicians or Asthma Experts, and 3) Patients.
1. Respirologists (N=4)

Figure B-1: Examples of Asthma Action Plan designed by Respirologists

2. Primary Care Physicians or Asthma Experts (N=7)

Figure B-2: Examples of Asthma Action Plans generated by a Primary Care Physician or Asthma Educator

3. Patients (N=5)
Figure 8-3: Examples of Asthma Action Plans generated by Patients

The plans generated by each user group were quite different. In general, doctors and experts tended to favour adding many quantitative details (e.g. the number of times to take a reliever) and medical terminology whereas patients typically stayed away from specifics and instead selected more general statements to create a simpler asthma action plan.

Most participants were able to complete the design of their asthma action plan in the allotted time (45 minutes). However, many were unable to include all the options they preferred to have on the plan due to space limitations.

Questionnaire Results

Each participant in the focus group was also asked to fill out a questionnaire after they have created the action plan. Each statement was rated on a Likert scale where 1 = Strong Disagree and 5 = Strongly Agree. The following tables show the average rating given by the three stakeholder groups: Respirologists (N=4), Primary Care Physicians or Asthma Educators (N=7), and Patients (N=5).

System Usability Scale

Overall, users liked the Wikibreathe tool and thought that it was relatively easy to use. One interesting discrepancy was the score for “I think that I would need the support of a technical person to be able to use this system.” Both the Respirologist and Physician groups disagreed
strongly with this statement whereas the Patient group rated this statement closer to the Neutral/Agree side.

<table>
<thead>
<tr>
<th>Statement</th>
<th>Respirologists</th>
<th>Physicians</th>
<th>Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>I think that I would like to use this system frequently.</td>
<td>3.75</td>
<td>3.86</td>
<td>4.20</td>
</tr>
<tr>
<td>I found the system unnecessarily complex.</td>
<td>2.25</td>
<td>2.43</td>
<td>1.80</td>
</tr>
<tr>
<td>I thought the system was easy to use.</td>
<td>4.00</td>
<td>3.86</td>
<td>4.40</td>
</tr>
<tr>
<td>I think that I would need the support of a technical person to be able to use this system.</td>
<td>1.50</td>
<td>1.71</td>
<td>3.20</td>
</tr>
<tr>
<td>I found the various functions in this system were well integrated.</td>
<td>4.00</td>
<td>3.86</td>
<td>4.00</td>
</tr>
<tr>
<td>I thought there was too much inconsistency in this system.</td>
<td>1.75</td>
<td>2.14</td>
<td>2.60</td>
</tr>
<tr>
<td>I would imagine that most people would learn to use this system very quickly.</td>
<td>4.00</td>
<td>3.86</td>
<td>3.00</td>
</tr>
<tr>
<td>I found the system very cumbersome to use.</td>
<td>2.50</td>
<td>1.86</td>
<td>1.80</td>
</tr>
<tr>
<td>I felt very confident using the system.</td>
<td>4.00</td>
<td>3.86</td>
<td>4.20</td>
</tr>
<tr>
<td>I needed to learn a lot of things before I could get going with this system.</td>
<td>1.75</td>
<td>1.71</td>
<td>2.40</td>
</tr>
</tbody>
</table>

Table 8-2: System Usability Scale

Available Functionality

All user groups found that they were able to make the appropriate content and format changes needed for their plans, as well as to change previous choices. The last statement “I would be able to do this from my home computer, over the internet” measured how easily they feel that they can use the tool without immediate technical support at home and all user groups felt confident that they can do this.

<table>
<thead>
<tr>
<th>Statement</th>
<th>Respirologists</th>
<th>Physicians</th>
<th>Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>I was able to make the content changes that I wanted to.</td>
<td>3.50</td>
<td>4.14</td>
<td>4.40</td>
</tr>
<tr>
<td>I was able to make the format/visual changes that I wanted to.</td>
<td>2.75</td>
<td>3.29</td>
<td>4.00</td>
</tr>
<tr>
<td>I was able to easily change previous choices if I wanted to.</td>
<td>4.25</td>
<td>4.29</td>
<td>4.60</td>
</tr>
<tr>
<td>I would be able to do this from my home computer, over the internet.</td>
<td>4.50</td>
<td>4.43</td>
<td>4.40</td>
</tr>
</tbody>
</table>

Table 8-3: Available Functionality
**Navigation Structure**

Despite initial concerns over the navigation structure, all user groups found it straightforward and easy to use. Users also understood the categorization of content into “Basic Choices” and “Content Options.”

<table>
<thead>
<tr>
<th>Statement</th>
<th>Respirologists</th>
<th>Physicians</th>
<th>Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>I found the navigation structure straightforward and easy to use (tabs, drop-down menus and scroll bars).</td>
<td>4.50</td>
<td>4.29</td>
<td>4.40</td>
</tr>
<tr>
<td>I found that dividing options into “Basic Choices” and “Content Options” was a good way to organize the options.</td>
<td>4.50</td>
<td>4.14</td>
<td>4.20</td>
</tr>
</tbody>
</table>

**Table 8-4: Navigation Structure**

**Special Features**

During the focus group, it was observed that respirologists and physicians used the comment feature more than patients, which might explain the higher rating they gave to the statements regarding the usefulness of this feature. Users also appreciated the ability to preview the asthma action plan as a PDF file. The Zoom feature was unavailable due to technical issues for the majority of the focus group sessions, hence the neutral response.

<table>
<thead>
<tr>
<th>Statement</th>
<th>Respirologists</th>
<th>Physicians</th>
<th>Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>I was able to post comments.</td>
<td>5.00</td>
<td>4.50</td>
<td>3.40</td>
</tr>
<tr>
<td>Having the ability to post comments was useful.</td>
<td>5.00</td>
<td>4.71</td>
<td>3.80</td>
</tr>
<tr>
<td>Having the ability to preview how the plan would look when printed (the “preview” feature) was useful.</td>
<td>4.50</td>
<td>4.57</td>
<td>4.40</td>
</tr>
<tr>
<td>Using the preview feature, the final format screen accurately reflected changes I made in the active screen.</td>
<td>5.00</td>
<td>4.57</td>
<td>4.40</td>
</tr>
<tr>
<td>The zoom feature was useful.</td>
<td>3.00</td>
<td>3.57</td>
<td>3.00</td>
</tr>
</tbody>
</table>

**Table 8-5: Special Features**
**Appearance**

Most users were satisfied in terms of the appearance of the Wikibreathe tool (colour, size of text, font type, and overall layout), however, there were a few comments which indicated that the font size might be a bit too small to read on screen.

<table>
<thead>
<tr>
<th>Statement</th>
<th>Respirologists</th>
<th>Physicians</th>
<th>Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>The colours were appealing.</td>
<td>4.00</td>
<td>4.29</td>
<td>4.40</td>
</tr>
<tr>
<td>The size of the text was easy to read.</td>
<td>4.00</td>
<td>4.00</td>
<td>4.40</td>
</tr>
<tr>
<td>The font (lettering) selections were easy to read.</td>
<td>4.00</td>
<td>3.86</td>
<td>4.40</td>
</tr>
<tr>
<td>The overall appearance of the tool was appealing.</td>
<td>4.25</td>
<td>4.00</td>
<td>4.80</td>
</tr>
<tr>
<td>The layout of the document was easy to read.</td>
<td>4.50</td>
<td>4.29</td>
<td>4.40</td>
</tr>
</tbody>
</table>

**Table 8-6: Appearance**

**Default Options**

An alternate option to having “Don’t Show” as the default choice would have been to include a checkbox next to each option where selecting the checkbox would enable the option and not selecting the checkbox would be equivalent to “Don’t Show.” Both design options were considered but due to time limitations in the development cycle, it was decided that we would go with the first option. Fortunately, users seemed to be satisfied with the current design and also thought that the option of “Show, but phrasing doesn’t matter” made sense as well.

<table>
<thead>
<tr>
<th>Statement</th>
<th>Respirologists</th>
<th>Physicians</th>
<th>Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>The setup with the default choice for each specific option set to “Don't Show” made sense to me</td>
<td>3.75</td>
<td>4.29</td>
<td>4.60</td>
</tr>
<tr>
<td>It made sense to include the option “Show, but phrasing doesn't matter” in addition to the specific phrasings</td>
<td>4.00</td>
<td>4.00</td>
<td>4.40</td>
</tr>
</tbody>
</table>

**Table 8-7: Default Options**

**Terminology**

The Respirologist group found the terminology used throughout the Wikibreathe tool a bit more confusing compared to the Physician and Patient groups; however, all groups found it useful to
have the feature that would show a definition of the more complex terminology when hovering over the text.

<table>
<thead>
<tr>
<th>Statement</th>
<th>Respirologists</th>
<th>Physicians</th>
<th>Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>The terminology used throughout the Wikibreathe tool was confusing.</td>
<td>2.50</td>
<td>1.71</td>
<td>1.80</td>
</tr>
<tr>
<td>The feature to hover over a term to show its definition was useful.</td>
<td>4.00</td>
<td>4.29</td>
<td>4.20</td>
</tr>
</tbody>
</table>

Table 8-8: Terminology

**Number of Options**

Curiously, the Patient group seemed more satisfied with the number of available options whereas comparatively, the Respirologist and Physician groups noted that there were too many format and content choices. Overall, however, all ratings were still within the positive range.

<table>
<thead>
<tr>
<th>Statement</th>
<th>Respirologists</th>
<th>Physicians</th>
<th>Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>There were too many format choices (ex. too many different choices for formats to describe the frequency of puffer use).</td>
<td>2.50</td>
<td>2.43</td>
<td>1.80</td>
</tr>
<tr>
<td>There were too many content choices (ex. too many different choices for how “Chest tightness” could be phrased).</td>
<td>2.25</td>
<td>3.00</td>
<td>1.80</td>
</tr>
</tbody>
</table>

Table 8-9: Number of Options

**Paper Version of Asthma Action Plan**

Finally, the focus group participants were asked to compare their plans with 1) a printed copy of the plan in both colour and black/white and 2) the plans created by the other participants in the group. Note that the Patient group was evaluated one-on-one so they were not able to compare their plans with others.

The most striking difference is in the responses to the statement “I am completely satisfied with the asthma action plan that I have created.” The Patient group rated their satisfaction with the plan quite high (4.60) whereas the Respirologist group rated their satisfaction much lower (2.75). A possible explanation could be related to the fact that patients did not see plans other than their own whereas respirologists and physicians were able to compare their plans with others in the same group. Another possible explanation for this discrepancy is that clinicians may simply
have a better understanding of the information needed on an asthma action plan and therefore have a higher standard for the content that the plan should contain.

<table>
<thead>
<tr>
<th>Statement</th>
<th>Respirologists</th>
<th>Physicians</th>
<th>Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>I am completely satisfied with the asthma action plan that I have created.</td>
<td>2.75</td>
<td>3.20</td>
<td>4.60</td>
</tr>
<tr>
<td>The appearance of my paper AAP is the same as what was displayed on the Wiki tool.</td>
<td>4.75</td>
<td>4.40</td>
<td>4.40</td>
</tr>
<tr>
<td>I prefer the colour version of the paper plan compared to the black and white version.</td>
<td>4.75</td>
<td>4.40</td>
<td>5.00</td>
</tr>
<tr>
<td>When compared to the AAPs created by the other participants, I feel that my plan is the best in terms of appearance.</td>
<td>3.25</td>
<td>3.00</td>
<td>n/a</td>
</tr>
<tr>
<td>When compared to the AAPs created by the other participants, I feel that my plan is the best in terms of relevant content.</td>
<td>3.00</td>
<td>3.00</td>
<td>n/a</td>
</tr>
</tbody>
</table>

Table 8-10: Paper Version of Asthma Action Plan
Appendix C: Focus Group Moderator Script

OCTAPUS Study
Standardized Script – Development Stage
Focus Group Moderator

Participant type: **Respirologist**

**INTERVIEWER COPY**

*Notes to interviewer in BOLD*

HAND OUT NAME TAGS AND ASK PARTICIPANTS TO WEAR.

1. **Introduction (25 minutes)**

- Hello, my name is “x”, I am a research associate working with the OCTAPUS Study, looking at asthma action plans.

Participant Introductions
This is a focus group session for respirologists. I’d like to welcome you and to thank you for taking the time to participate in this study. Before we begin, I’d like to go around the table quickly and ask you to tell us your name and please tell us where you currently work.

GO AROUND THE TABLE ONE BY ONE.

As you know, this focus group will take about 2 hours. We will start with a brief introduction, then you will use the Action Plan Tool on the laptop in front of you for about 45 minutes, and then you will complete an online questionnaire, and we will have a 45-minute group debriefing session and go through the tool together.

Background (SLIDE 2) (5 MINUTES)

• HAND PARTICIPANTS THE 5 LAMINATED AAP’s (PASS AROUND).

Study Goals (SLIDE 3) (1 MINUTE)

Study Process (SLIDE 4) (1 MINUTE)

Tool Tutorial (SLIDE 5) (15 MINUTES)

• I will run through the tool with you to show you how it works. Feel free to ask questions as I go.
• GO THROUGH ALL POINTS IN FAQ BY DEMONSTRATING LIVE ON THE TOOL:
  o LOGON – tell each user that logon is their last name, password: “password” – have them log on
  o TOOL SECTIONS (1-3) WITH THEIR FUNCTIONS
  o HOW TO NAVIGATE (REMIND PARTICIPANTS TO CLICK ON EVERY TAB LEFT TO RIGHT AND GO THROUGH EVERY MENU) + EXPLAIN APPEARANCE OF NEW MENUS
  o HOVER FEATURE
  o BASIC CHOICES VS CONTENT OPTIONS
EXPLANATION OF DEFAULT SETUP (“DON’T SHOW” AND “SHOW, BUT PHRASING DOESN’T MATTER”)

TERMOWOLOGY SELECTION

POSTING COMMENTS (REMNID PARTICIPANTS THAT IF THERE WERE CHOICES MISSING THAT THEY WOULD LIKE TO ADD, TO ADD THEM HERE. ALSO, THEY CAN USE THIS SECTION TO COMMENT ON A GLITCH THEY ENCOUNTERED OR THEIR OPINION REGARDING THE OPTION)

ZOOM FEATURE

PREVIEW FEATURE

SAVING CHANGES

RESET FEATURE

MAXIMUM NUMBER OF OPTIONS

MENTION THAT THE FAQ IS AVAILABLE BY CLICKING THE LINK IN THE TOOL AT ANY TIME.

- ARE THERE ANY QUESTIONS ABOUT THE SETUP OR FEATURES OF THE TOOL?

Objectives (SLIDE 6) (1 MINUTE)

TELL THEM THAT WE WILL BE OBSERVING THEM AND TAKING NOTES AS THEY USE THE TOOL, AND THEY CAN ASK QUESTIONS AT ANY TIME.

WE WILL LET THEM KNOW WHEN 30 AND 15 MINUTES ARE LEFT.

2. Individual Tool Use (45 minutes)

3. System Usability Feedback

- This session is being audio-recorded, but all of your answers and these recordings will be kept confidential and stored without your name on them.

- Now that we have finished developing our individual asthma action plan templates, I would ask you to open the other web-browser window on your laptop, and to fill out the online questionnaire about the actual usability of the tool. Your user ID has already been entered.
• Fill out question 1 (System Usability Scale) and question 2 (Available Functionality)
• MODERATOR OPEN QUESTION: Do you have any comments about available functionality?
• Fill out question 3 (Navigation Structure)
• MODERATOR OPEN QUESTION: Do you have any comments about the navigation structure of the tool?
• How can the navigation structure be improved?
• Fill out question 4 (Special Features)
• MODERATOR OPEN QUESTION: Do you have any comments about the special features of the tool (comments, preview and/or zoom)?
• Are there any other functions that would make it easier to use or more useful?
• Fill out question 5 (Appearance)
• MODERATOR OPEN QUESTION: Do you have any comments about the appearance of the tool (colours, size of text, font selections and/or layout)?
• Would you change anything about the appearance of the tool?
• Fill out question 6 (Default Options)
• MODERATOR OPEN QUESTION: The current default in the drop-down menu for most of the options is “Don’t Show.” This way, options are not included in the asthma action plan unless the user picks an alternate choice from the drop-down menu. An alternative way to allow users to determine which options to include in the asthma action plan would be to add a checkbox next to each option. This checkbox would be left unselected by default, and users could then check off the options that they wish to include in the asthma action plan, and then choose among specific phrasings in each associated drop-down menu. Diagrams below demonstrate this (use SLIDE 7 to explain).

Which setup for allowing users to choose options to include in the asthma action plan would be preferable? Why?
• Fill out question 7 (Terminology)
• MODERATOR OPEN QUESTION: Do you have any comments about the terminology used in the tool?
• Were there any confusing terms that did not have a “hover” explanation? Were any of the “hover” explanations inadequate to explain the term?

4. Incident Reports

MODERATOR OPEN QUESTION: We noted a few areas where people seemed to struggle with the tool as we observed them. Let’s go over these to clarify what went wrong.
5. AAP Feedback (15 minutes)

PRINT ALL AAPs IN B&W AND COLOUR AND BRING COPIES OF EACH AAP FOR EACH PARTICIPANTS, LABEL THEM A-E

- Now that we you have finished developing our individual asthma action plans, we have printed your own AAP and those of your fellow participants for you and would like your comments.
- GIVE OUT THE AAPs AND ALLOW PARTICIPANTS A FEW MINUTES TO LOOK OVER EACH ONE
- Fill out question 8 (Paper plans)
- MODERATOR OPEN QUESTION: How closely did the printed version of the AAP resemble the version displayed on the Wiki tool? What were the differences (besides colour for the black and white ones)?
- Having looked at your colleagues’ AAPs, were there any content elements that you wish you would have included?
- If so, were you aware of these content choices in the AAP tool, or did you miss them?
- Were there any format choices that you think were better than yours?
- If so, were you aware of these format choices in the AAP tool, or did you miss them?
- Please rate the colour AAPs 1 to 5 by placing a number in the upper right hand corner, according to the best overall AAP to the worst overall AAP.
- Please place a star next to the AAP that was yours.

5. Content/Format Feedback (30 minutes)

- At this point I would like to run through the tool with you on the projector, and solicit your feedback regarding various parts of it.

- MODERATOR OPEN QUESTION: To begin with, were there any major content or format features that were missing from the tool?
- Fill out question 9 (Number of Options)

- Let’s go through the parts in some detail.
GO THROUGH EACH ITEM HERE (LISTED BELOW) AND ASK IF PEOPLE FELT THAT IT WAS RELEVANT TO INCLUDE, AND IF ALL RELEVANT OPTIONS FOR IT WERE PRESENT. IF NOT, SOLICIT WHAT OTHER OPTIONS SHOULD BE ADDED.

AAP Sections

ASK ABOUT CONTENT AND FORMAT SEPARATELY IN EACH SECTION.

a) Green Zone
b) Yellow Zone
c) Red Zone
d) Header Tab
e) Other
f) Layout/Number of Zones Box
g) Setup Tab

6. Overall Impressions (5 minutes)

MODERATOR OPEN QUESTIONS:
• Did you identify any glitches in the tool? If so, which ones?

• Overall, what did you like about the wiki tool?

• Overall, what did you dislike about the wiki tool?
Appendix D: Usability Problems

Table 1 (Protocol A: Forward > Reverse)

* Problems found during *Task 1* (Forward Flow)

<table>
<thead>
<tr>
<th>Usability Issue</th>
<th>User 1</th>
<th>User 3</th>
<th>User 5</th>
<th>User 7</th>
<th>User 9</th>
<th>User 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peak Flow</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Search</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Hidden Triggers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Basic Choices vs Content Options</td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Additional Statements</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
</tbody>
</table>

Table D-1: Protocol A, Task 1

Table 2 (Protocol B: Reverse > Forward)

* Problems found using *Task 1* (Reverse Flow)

<table>
<thead>
<tr>
<th>Usability Issue</th>
<th>User 2</th>
<th>User 4</th>
<th>User 6</th>
<th>User 8</th>
<th>User 10</th>
<th>User 11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Layout</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Peak Flow</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Terminology</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Basic Choices vs Content Options</td>
<td>1</td>
<td></td>
<td>2</td>
<td>5</td>
<td></td>
<td>5</td>
</tr>
</tbody>
</table>
Table D-2: Protocol B, Task 1

Table 3 (Protocol A: Forward > Reverse)

* Problems found during Task 2 (Reverse Flow) that were not already found in Task 1

<table>
<thead>
<tr>
<th>Usability Issue</th>
<th>User 1</th>
<th>User 3</th>
<th>User 5</th>
<th>User 7</th>
<th>User 9</th>
<th>User 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default Options</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guided Prescription</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Terminology</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Submenus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

Table D-3: Protocol A, Task 2

Table 4 (Protocol B: Reverse > Forward)

* Problems found during Task 2 (Forward Flow) that were not already found in Task 1

<table>
<thead>
<tr>
<th>Usability Issue</th>
<th>User 2</th>
<th>User 4</th>
<th>User 6</th>
<th>User 8</th>
<th>User 10</th>
<th>User 11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic Choices vs</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Content Options</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Default Options</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table D-4: Protocol B, Task 2
Appendix E: Moderator Script for Experiment

Moderator Script for Usability Experiment

Protocol: ______
Participant #: ______

Date: _______________________ Start time: ______________

1. Please read this introduction and let me know if you have any questions.
   - Give three asthma action plan examples to user as reference

2. Sign the consent form

3. Please fill in the survey question #2: Background information

4. <Log on to the Wikibreathe tool for the user: userX_p1>

This is the Wikibreathe tool that you will be using to create the asthma action plan. We will spend the next 10 minutes going over some of the important features in the tool.

5. Now that you are familiar with the tool, we’ll proceed with the first part of the experiment. You may begin.
   - Hand “Experiment 1” to user
   - Reset plan
   - Turn on recorder

   Initial Setup: __________
   Header: __________
   Zone Description: __________
6. Please fill in survey question #3 (Experiment 1)

   We will now move on to the second part of the experiment. <Hand “Experiment 2” to user>

7. <Log on to the Wikibreathe tool for the user : userX_p2>

   Initial Setup: __________
   Header: __________
   Zone Description: __________
   Zone Instruction: __________
   Footer: __________

   Number of Errors: __________

8. Please fill in survey question #4 (Experiment 2)

9. I will now ask you a few questions:
Open-ended discussion questions

• <Specific questions based on observations: Why did you do ________?> when having problems

Other Questions

• You did this task twice, once to follow a set of written instructions and once to recreate a completed plan – which one did you enjoy more and why?
• Which method did you feel was better in helping you learn how the tool works? Why?
• If you were running a usability test, which method would you choose to use or think would be more effective?
• What did you like most about the tool?
• What did you dislike most or find most confusing about the tool?
• Which part was hardest to use / most difficult?
• In what ways would you improve the design of the tool?
Appendix F: Participant Protocol for Experiment

Introduction

Thank you for taking the time to participate in this study. This research study is focused on creating an online tool that will generate an asthma action plan. An asthma action plan is a paper plan given to an asthma patient by their physician. This plan tells the patient what to do when their asthma symptoms worsen, for example, what medications they need to take and how often to take it. In front of you are a few different examples of existing asthma action plans.

The purpose of this research is to study the preferences of users by giving them a way to design and generate the asthma action plan that best fit their needs. To do this, we have created an online tool called “Wikibreathe” that would let the user select various options to build their own plan.

After a brief introduction to the Wikibreathe tool, you will be asked to complete two separate tasks and questionnaires related to each task. When both tasks have been completed, we will have a short debriefing session to gather your feedback. The entire experiment should take about one hour.

Note that the session will be audio-recorded; the information will remain confidential to the researchers.
Experiment 1

Imagine that you are a physician who is designing an asthma action plan for your patients with certain preferences. Given these preferences, use the Wikibreathe tool to create the action plan that best fits the requirements. Feel free to talk aloud on what you're thinking about while you are working on the task. If you are unable to complete a task, move on to the next task. If the instructions do not specify the exact wording of a selection, pick any option that fulfills the requirement. You will have approximately 15 minutes to complete this section.

Preferences Part 1: Initial Setup

- The plan should contain three zones representing worsening asthma symptoms.
- Each time controller inhalers are mentioned, the plan should use the term **preventer** instead.

Preferences Part 2: Header

- The plan should include a space to put the patient’s name and their date of birth and the severity of their asthma symptoms.
- There should be an empty space where you can fill in your patient’s personal asthma goals.

Preferences Part 3: Zone Descriptions

- The titles of the asthma zones should be:
  - Green = Normal
  - Yellow = Caution
  - Red = Danger.
- **Each** of the three zone sections should include the range of peak flow for the patient where the physician can fill in the peak flow percentages in blank spaces.
- **Each** of the three zone sections should include descriptions on Breathlessness, Coughing, and the Frequency of Daytime Symptoms.
- In addition, the Red Zone should also mention that the patient is having difficulty speaking.

Preferences Part 4: Zone Instructions

- For the instructions section of the plan, the Green and Yellow zone sections should be in the format of a table that has columns representing the medication, puffer colour, how much of it to take and when to take it.

...continue to next page.
Preferences Part 5: Footer

• At the bottom of the page, there should be a section on the most common asthma triggers that patients should be aware of: air pollution, cats and dogs, dust and pollens. This should be displayed using both text and images.

• For the section on asthma triggers, you would like to be able to check off the ones that pertain to your patient.

• There should also be a line at the bottom of the page for the physician’s signature.

End of Experiment 1
Experiment 2

In the second part of the experiment, you will be given a completed version of an asthma action plan. Your goal is to recreate the same plan using the Wikibreathe tool. Feel free to talk aloud on what you're thinking about while you are working on the task. If you are unable to complete a task, move on to the next task. You will have approximately 15 minutes to complete this section.

End of Experiment 2
Appendix G: Results from Heuristic Evaluation

The Heuristic Evaluation technique was used by the researcher to assess the usability of the Wikibreathe tool.

The seven questions in the standard from ISO-9241-110 will be applied to the Wikibreathe tool. The responses will be ranked as Low, Medium and High where Low compliance means that the tool does not support this heuristic and High compliance means that the tool completely satisfies the heuristic and Medium compliance means that the tool is satisfactory but can be improved.

Usability Evaluation Results

1. Is the dialogue suitable for the user's task and skill level? (Suitability for the task)
   Analysis: MEDIUM
   > The dialogue, or in this case, menu for the Wikibreathe tool is somewhat complex in that there are multiple levels of tabs, sub-tabs, menus, and options. The navigation can be clarified by reducing duplicate terms (ex. ‘Basic Choices’ and ‘Content Options’) and allowing users to manipulate items directly on the asthma action plan.

Figure 8-1: The dialogue/menu for Wikibreathe
2. Does the dialogue make it clear what the user should do next? (Self-descriptiveness)
   Analysis: MEDIUM
   > It is not always clear to the user where they are in the system due to the multi-level navigation structure. It somewhat makes sense to go through the tabs from left to right and follow the menus from top to bottom but a more explicit “wizard” type of interface may help novice users navigate the tool more easily.

   ![Figure G-2: Tab and Sub-Tabs](image)

3. Is the dialogue consistent? (Conformity with user expectations)
   Analysis: MEDIUM-HIGH
   > The Wikibreathe tool mostly follow Windows conventions of tabs, expanding menus and drop-down boxes. A source of confusion might stem from the use of black arrows to indicate that a new drop-down box has appeared (Figure 6-3):

   ![Figure G-3: Arrows to indicate that a new submenu or drop-down has appeared](image)

4. Does the dialogue support learning? (Suitability for learning)
   Analysis: LOW-MEDIUM
   > There is a link to a Help File / FAQ document that is available through the tool and some text indicates when a user has reached the end of a section, but the program mainly
leaves it up to the user to learn how to use it.

You have reached the end of this section. Please continue on to the next tab.

Figure G-4: End of Section Message

5. *Can the user control the pace and sequence of the interaction? (Controllability)*
   
   Analysis: **HIGH**
   
   > The Wikibreathe tool allows users to edit any part of the action plan. There is no restriction on the pace or sequence of the interaction.

6. *Is the dialogue forgiving? (Error tolerance)*
   
   Analysis: **MEDIUM**
   
   > The main source of error in the Wikibreathe tool is when a user selects too many options and runs out of space on the asthma action plan. In this case, a dialogue box will pop up (Figure 6-5) but the program leaves it up to the user to figure out which element to remove. The program can be improved by adding an intelligent error correction feature that guides the user in removing the necessary elements in order to make more room in the action plan.

   ![Figure G-5: Space limitation alert box](image)

7. *Can the dialogue be customised to suit the user? (Suitability for individualisation)*
   
   Analysis: **LOW**
   
   > The Wikibreathe tool cannot be customized to suit individual user preferences.
Appendix H: Experimental Design

This section outlines an experimental design for an editing tool like Wikibreathe that takes into account the learning effects of using the same tool.

Experiment Design

Requirements: 8 subjects, 2 versions of the tool, and 2 versions of asthma action plans

Description of Tools: Wiki 1 is the original version of the tool which should be more basic and harder to use, Wiki 2 is the newer updated version which should be more usable

Description of the Asthma Action Plans: “Simple" is the simpler version of the two action plans and "Advanced" is the more complicated version of the action plan

Process: Each subject will reverse engineer two action plans

Design: To account for learning effects, some subjects will use the original version of the tool first, and others will use the newer version. Each of these is also paired with either the simple or advanced version of the AAP (see table below, "A" indicates that this test will be done first, "B" will be done second).

<table>
<thead>
<tr>
<th>Subject</th>
<th>Wiki 1</th>
<th>Wiki 2</th>
<th>Simple</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>4</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>5</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>6</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>7</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>8</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
</tr>
</tbody>
</table>

Table H-1: Experimental Design