REMOTE USABILITY EVALUATION OF A HEALTH COACHING APPLICATION

BY

ANDI LUSHAJ

A CAPSTONE

Presented to the Oregon Health & Science University's
Department of Medical Informatics & Clinical Epidemiology
in partial fulfillment of
the requirements for the degree of
Masters of Biomedical Informatics
May 2010

Capstone Advisor
Holly Jimison, Ph.D.

School of Medicine

Oregon Health & Science University

Master of Biomedical Informatics

CERTIFICATE OF APPROVAL

This is to certify that the Master's Capstone Project of

Andi Lushaj

"Remote Usability Evaluation of a Health Coaching Application"

Has been approved

Holly Jimison, PhD
Capstone Advisor

TABLE OF CONTENTS

ABSTRACT

BACKGROUND: Computer-mediated educational applications can provide a self-paced, interactive environment to deliver educational content to individuals about their health condition, resulting in improved overall health and reduced health care costs.

OBJECTIVE: There is a need to adequately evaluate health information systems to show evidence that justifies investment in such systems. We evaluated the feasibility of remote usability testing of the health coaching software, as an easier to arrange and cheaper method of testing users in their own environment and more frequently.

METHODS: One evaluator carried out a heuristic evaluation of the Automated Health Coaching System interface based on Nielsen's heuristics, and generated a list of usability problems with severity ratings. For the usability testing, five participants used the interface and scenarios to complete 11 tasks. We measured the elapsed testing time on task, the number of errors made per task, and the task accuracy. Sessions were recorded to capture the participant's screen and mouse movement, sound, and webcam for later analysis.

RESULTS: The remote usability evaluation found a total of 146 usability problems. The heuristic evaluation violations mostly consisted of lack of consistency and standards, lack of an aesthetic and minimalist design, failure of the system to match the real world, and lack of visibility of system status. The usability testing showed problems in areas like viewing and interpreting the performance graphics, editing of issues, locating a subject's active topic or contacts, or making the correct selections within contacts and issues.

CONCLUSION: The findings of this study demonstrate that the heuristic evaluation and remote synchronous usability testing are efficient methods for interface evaluation and design improvement.

ACKNOWLEDGEMENTS

My greatest thanks go to my advisor Dr. Holly Jimison, for her guidance and support, and for giving me the opportunity to work on this project. I am extremely grateful for her patience and encouragement throughout the progress of the project.

I sincerely thank the group of talented individuals working with Dr. Jimison for their invaluable contribution in making this project possible. Michael Chan was especially instrumental in every aspect of this project and was always available to help, particularly in selecting the testing software and setting up the testing environment; Michael Shapiro provided a great assistance with materials, insight and feedback from the very start; Chia Hua Yu helped with references and the selection of testing software. I would like to also acknowledge Nicole Larimer and Susan Butterworth for their assistance, feedback and time.

Most importantly, I thank my family, my wife Entela who has always been there for me, my daughter Emma who is the joy of my life, and Chufo for the unconditional love.

INTRODUCTION

Health Coaching System

The Automated Health Coaching System is a home-based cognitive health coaching system developed at the Oregon Center for Aging & Technology (ORCATECH) that provides computer-based health coaching interventions intended to promote improved health behaviors for elders. Coaching interventions include cognitive games and cognitively supportive activities such as physical exercise, sleep quality, brain games, socialization, and medication management. The Automated Health Coaching System provides a daily assessment of cognitive abilities of patients, which may lead to early discovery of illness where cognitive loss is a primary indicator.

Health coaching has been described as the practice of health education and health promotion within a coaching context to enhance the well-being of individuals and to facilitate the achievement of their health related goals.[1] Butterworth et al. defined health coaching as a behavioral health intervention that facilitates participants in establishing and attaining health-promoting goals in order to change lifestyle-related behaviors, with the intent of reducing health risks, improving self-management of chronic conditions, and increasing health-related quality of life.[2]

The Pew Research Center found that 34 percent of adults 65 years and older used computers in 2006, and broadband use for this group increased from 19 percent in May 2008 to 30 percent in April 2009.[3, 4] Older adults use computers for various activities such as communication with family and friends, entertainment, and information. These activities help them feel more confident and less isolated socially.[5]

Computer-mediated educational applications can provide a self-paced, interactive environment to deliver educational content to individuals about their health condition. Studies show that patients would like to receive health coaching regarding physical exercise, diet, cognition and mood.[6] Another study found that cognitive training interventions (training for memory, reasoning, or speed of processing) produced an immediate effect on its corresponding in cognitive abilities and daily functioning in older adults living independently, durable to 2 years.[7]

Usability Evaluation

The International Standards Organization (ISO) 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". There is a need to adequately evaluate health information systems to show evidence that justifies investment in such systems. In a case that illustrates a return on usability, Spool reported an instance when an e-commerce company shifted the registration step so that it came after the checkout steps. This modification resulted in an extra $300 million in the first year.[8]

There are different methods of measuring usability: inspection methods, testing methods, and inquiry methods. The different techniques used to measure usability are further broken down for each method. Inspection methods consist of heuristic evaluation, cognitive walkthrough, pluralistic walkthrough, standards inspection, and guidelines checklist. Testing methods include the thinking aloud protocol, co-discovery, performance measurement, and in-field studies. The inquiry method includes questionnaires and interviews.[9]

Usability testing involves recruiting target users to participate in the evaluation of the usability of a web page or application. Formal usability testing is usually done in a usability lab where a

moderator observes the test participant while the test is administered, with computers outfitted with screen and keystroke capture software, and a video camera to record participant actions and facial expressions.

Remote vs. In-Person

In-person usability testing has been the standard for evaluating interface software usability. Remote Usability testing is the assessment of the usability of software with users (participants) who are not in the same location as the researchers. It presents an opportunity to address some of the in-person weaknesses and it could be an effective alternative of evaluating product usability at a reduced cost and effort, with the benefit of testing globally distributed user audiences in their native environments. The usability testing done in a lab environment is limited to testing users on location and requires more travel and other resources.

One study that compared traditional lab-based vs. remote Web-based usability testing of Web sites, showed that both the lab and remote tests appear to capture similar information about the usability of a site, and the most critical usability issues with the sites were identified by both techniques.[10] Another study found no significant differences in the number of usability issues identified when remote and local studies were compared, suggesting that evaluators of expert interfaces can choose to do remote or local studies and obtain comparable results.[11] Andreasen et al. compared remote synchronous and asynchronous testing with conventional laboratory-based testing. The results showed that the remote synchronous testing is virtually equivalent to the conventional testing and equally effective in identifying usability problems.[12]

Synchronous vs. Asynchronous

Remote usability evaluation can be moderated (or synchronous) and automated (or asynchronous). Synchronous remote usability testing is conducted in real time and it is moderated by using a facilitator who works with participants during the study, but the facilitator is separated spatially from the subjects (the user and the facilitator are in different places). One or multiple tools are used to establish direct communication between the parties, and to record the interaction of the user with the application. The recorded output file may include recording of audio, video of facial expressions via a web camera, and recording of screen and mouse movements. Figures 1 and 2 below present a conceptual model of synchronous and asynchronous remote usability evaluation developed by Fidas at al.[13]

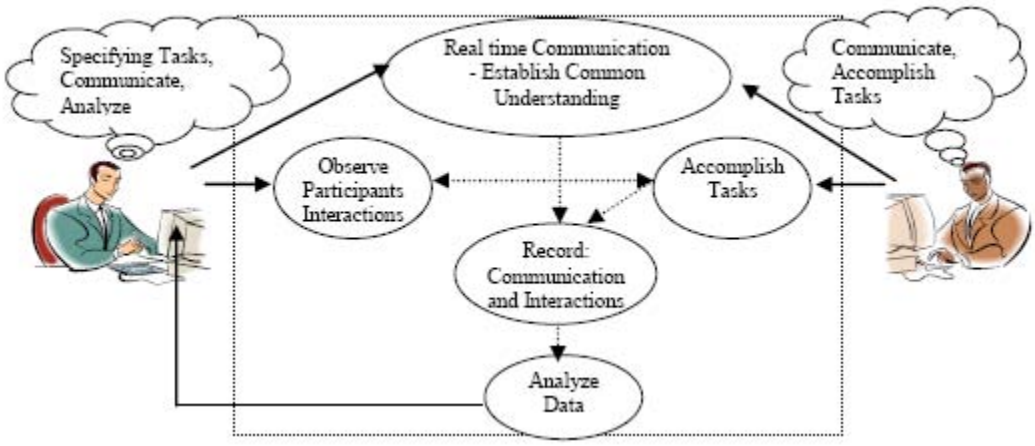

Figure 1: Synchronous remote usability illustration

Asynchronous remote usability testing is automated (unattended); there is no facilitator to guide the participant through the study, and no real-time communication with participants during the study. The monitor and the subjects are separated spatially and temporally. Typically this involves automated logging to quantify participant behavior, and survey software to obtain qualitative feedback. Asynchronous testing allows for collection of quantitative usage data from

a large number of participants, and it also provides a wealth of quantitative data about task completion rate and user clicking behavior.

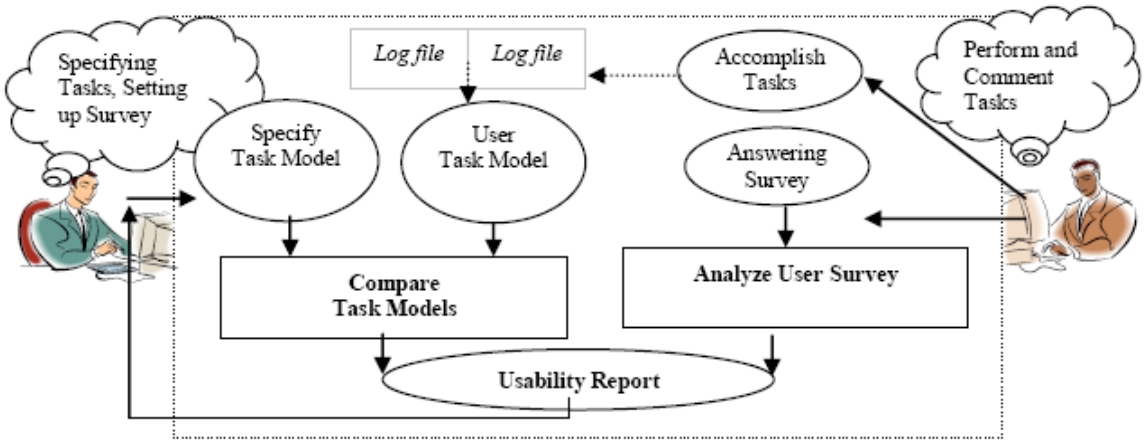

Figure 2: Asynchronous remote usability illustration

Tool Evaluation and Selection

Many tools are available that enable remote usability testing, such as web conferencing tools, survey tools, web analytics tools, meeting recording tools, and participant recruiting tools. For the purpose of this project, we did not try to identify the best usability tool in the market but rather we searched for an easy-to-use, cheap and effective single tool or a combination of tools that would allow us to establish a remote synchronous interaction with the test participants, to watch the participant's experience and to have a quality recording of the testing sessions. E.g., the screen sharing, audio and recording capabilities of web conferencing tools make them desirable solutions for remote usability testing, though they may not be labeled as usability testing tools. We aimed to have one recorded digital movie output file capturing the audio, participants' on-screen activity, and facial expression via a web camera.

We researched over thirty individual products by gathering information from vendors, and after narrowing the list of choices down based on this information, we also did two rounds of actual pre-test tool trials. The tools that I reviewed for the remote usability testing are summarized in Table 1 below.

Tool	Description	Company	Cost*
BB FlashBack	Records screen, sound, webcam, and mouse	www.bbsoftware.co.uk	Express: FREE Standard: $89 Pro: $199
BB TestAssistant	Similar to BB FlashBack; new release due later in 2010 with additional features: real-time tracking (real-time clock stored while making the movie), log integration, automatic removal of inactive periods, in-movie note-taking	www.bbsoftware.co.uk	$225
GoToMeeting	Screen sharing application, screen and audio recording	www.gotomeeting.com	$39 - $49/month
Dimdim	Audio, video, web conferencing, recording of all of them in one Flash file	www.dimdim.com	FREE Pro: $25/month
UserVue	Used for remote usability testing to observe and record participant's screen and audio. With integrated phone and chat, you can ask participants questions and give tasks throughout the session. Mark important moments.	www.techsmith.com	Monthly Pass: $149
VSee	Audio, video via web camera, screen sharing, recording or audio and video	vsee.com	FREE Premium: $50/user/month
Camtasia Studio	Captures participants' on screen activities, audio and webcam video. Can record picture in picture (web camera and onscreen capture).	www.techsmith.com	$299
Morae	PC-based tool that provides remote, simultaneous observation. Records screen and web cam of user's face, plus audio via microphone. Automatically can enter tasks/directions for participants and prepares surveys for tasks. The bundle includes: • Recorder – Capture user experience (audio, video, on-screen activity and keyboard/ mouse input) • Observer – Watch and hear user experience [faces, voices, interactions on screen], mark important moments, and flag tasks in real time • Manager – Analyze recordings, auto-calculate standard usability metrics (like effectiveness, efficiency, and satisfaction), generate graphs, and create highlight videos	www.techsmith.com	Recorder: $195 Observer: $195 Recorder/Observer Bundle: $349.00 Manager: $1,295 Whole bundle: $1,495 or $1,120 (education)
ScreenToaster	Web-based screen recorder to capture screen activity in real-time. You can add audio and an embed web cam while you are recording.	www.screentoaster.com	FREE
WebEx	Online meeting service to share desktop applications on screen and control of the cursor with a remote test participant. Live video – view up to six camera-enabled meeting participants simultaneously, each in their own window.	www.webex.com	$49/month
Skype	Skype provides free or low-cost video and voice calling, instant messaging (IM) and short message service (SMS), and file sharing	www.skype.com	Skype-to-Skype: FREE Other features: pay as you go, or pay monthly
Adobe Acrobat Connect	Web conferencing, screen sharing, real-time audio and live web cam video, and chat. Only the Acrobat Connect Professional allows recording.	www.adobe.com	$39.95/ month $395/ year Pro: $45 - $55 per host per month
CamStudio	Records screen and audio activity on computer and creates AVI (Audio Video Interleaved) video	camstudio.org	FREE
Pixetell	Supports screen recording, web cam recording, and voice card recordings	www.pixetell.com	$19/month
MindCanvas	Online research service to gather insights about customers' thoughts; online surveys.	www.themindcanvas.com	Priced by project
	Tool to quantify and measure usability and user experience.		

UserZoom	Gathers qualitative and quantitative data, such as: effectiveness ratios, efficiency ratios, click-stream paths, click-mapping, and users' feedback.	www.userzoom.com	Varying yearly subscription fee
ClickTale	Tracks user keystrokes, mouse clicks and moves and the time it takes for users to move around a web page. Records video of user interaction with application, heat maps of user clicks, where they scroll. Link analytics shows every interaction, hesitation time.	www.clicktale.com	FREE version Bronze: $99/ month; $474/ 6 months; $708/ year Silver: $290/ month; $1,392/ 6 months; $2,088/ year Gold: $790/ month; $3,792 / 6 months; $5,688 / year
RelevantView	Gathers customer insights by capturing both behavior and opinion.	www.relevantview.com	Fixed pricing based on services chosen
ActivityLens	Extracts information from collected data like video and sound audio files, log files, images and text files	hci.ece.upatras.gr	FREE
ClickHeat	Tracks user clicks via heatmaps	www.labsmedia.com	FREE
Clixpy	User movement recorder that tracks what users do on a web site including mouse movements, clicks, scrolling and form inputs	clixpy.com	Price per recorded sessions: $5 for 100 $10 for 200 $20 for 600 $30 for 1,000
Ethnio	Tool for finding and recruiting real web site users for a live remote test	www.ethniodev.com	Price per recruits: FREE – 20 $400 – 200 $800 – 2,000
Loop11	Tracks user interaction with a website and provides a report of task completion rate, time on task, common fail pages, and path analysis for each user. Used for unmoderated remote usability testing.	www.loop11.com	$350 per project
Usabilla	Tracks user clicks, collects user feedback about a web page, and provides result report of heatmaps or scatterplots of clicks. Used for unmoderated remote usability testing.	usabilla.com	5 pages: FREE 10 pages: $49 50 pages: $199 250 pages: $950
Open Hallway	Browser-based screen and audio recording. Videos can be up to 10 minutes long. Used for remote usability testing. Monthly plans: Basic – 3 hours storage; Plus – 10 hours storage; Premium – 30 hours storage.	www.openhallway.com	Basic: $49/ month Plus: $99/ month Premium: $199/ month
UserFly	Captures user's mouse movements, clicks and other screen interactions. Used for unmoderated remote usability testing. Monthly plans: FREE – 10 captures, 30 day storage; Basic – 100 captures, 30 day storage; Pro – 1,000 captures, 60 day storage; Business – 1,000 captures, 60 day storage; Enterprise – 10,000 captures, 90 day storage.	userfly.com	FREE Basic: $10/ month Pro: $25/ month Business: $50/ month Enterprise: $200/ month
Chalkmark	Captures user clicks and the time each click takes, and displays a heatmap showing the location of the clicks. Plans: FREE – surveys with up to 3 tasks; Monthly – unlimited surveys and tasks; Annual – unlimited surveys and tasks.	www.optimalworkshop.com	FREE $109/ month $559/ year
Hypercam	Records screen and sound as user interacts with system, and saves it as AVI digital movie	www.hyperionics.com	$39.95 No longer offered for sale
HyperCam 3	Captures screen, video and sound being played back and a microphone input	www.solveigmm.com	29.95 Euro
Total Screen Recorder	Records screen, audio activity and mouse cursor's movement. Can record web cams from AIM, ICQ, MSN Messenger, and Yahoo Messenger.	www.totalscreenrecorder.com	$29.95
FlashDemo Studio	Records screen activities and sound narration	www.flashdemo.net	$119.00 Lite: $99
ScreenCam (formerly Lotus/IBM ScreenCam)	Virtual camcorder for screen recording	www.smartguyz.com	$199.99
			Basic: $9/ month Standard: $19/ month

CrazyEgg	Click tracking tool displaying heatmaps of locations of clicks on a page	crazyegg.com	Plus: $49/ month Pro: $99/ month
Silverback	Mac-based software that captures screen activity, video of participant's reactions, and participant's voice	silverbackapp.com/	$49.95
Simple Mouse Tracking	Click tracking software that can be used to capture and analyze the clicks users make while visiting a web	smt.speedzinemedia.com	FREE

Table 1: Remote usability testing tools

*** Cost information may be subject to change after the date of this study**

As discussed above, for the purpose of this project we searched for a tool that would allow us to establish a remote synchronous interaction with the test participants, and to observe and record the testing sessions. For this reason, we left out tools that fell into the following categories:

- Web page analytics tools: they track where users click or scroll while visiting web pages, gather usage statistics and provide a heat map. Used particularly in eCommere web sites.
- Tools used to find and recruit real web site users.
- Tools that track user keystrokes (outside the scope of this project)
- Tools that define online survey questions to collect user feedback about a web page.

The first round of tool testing was done to understand the ease of use, features, quality and reliability of the software, and to weed out product bugs before selecting the tool of choice for the actual testing. We based our analysis on the presence or absence of features like audio, video (via web camera), screen sharing, integrated chat, and recording capabilities for audio, video and screen. In addition, we considered the ease of use, features, quality, reliability, product cost, and whether it was appropriate for asynchronous or synchronous testing. Once we identified reliable tools in a particular category, e.g. web conferencing tools like GoToMeeting or WebEx, we didn't attempt to find other similar products. Table 2 below shows the results of the remote usability testing tool evaluation.

	Audio	**Video (web cam)**	**Screen sharing**	**Recording audio**	**Recording video**
BB FlashBack	* Yes * Quality: good	* Yes * Quality: good	No	* Yes * Quality: good	* Yes * Quality: good
BB TestAssistant	Same as BB FlashBack	Same as BB FlashBack	Same as BB FlashBack	Same as BB FlashBack	Same as BB FlashBack

GoToMeeting	* Yes * Quality: good	No	* Yes * Quality: good	* Yes * Quality: good	No
Dimdim	* Yes * Quality: good; some background hiss at OHSU	* Yes * Quality: good, didn't work at OHSU on one instance but OK in later testing	* Yes * Quality: good, not as good as GoToMeeting, response time is fine	* Yes * Quality: good	* Yes * Quality: didn't get a recording of the web cam at OHSU on one instance but OK in later testing
Uservue	* Yes * Quality: good	No	* Yes * Quality: good, response time viewing participant's screen is excellent and fast	* Yes * Quality: good	No
Vsee	* Yes * Quality: good; some static/ distortion at OHSU	* Yes * Quality: good	* Yes * Quality: good	* Yes * Quality: good but records 2 separate audio/video files for user 1 and 2 (see comments)	* Yes * Quality: good but records 2 separate audio/video files for user 1 and 2 (see comments)
Camtasia Studio	* Yes * Quality: good	* Yes * Quality: good	No	* Yes * Quality: good but can't record audio and screen simultaneously but only one after another	* Yes * Quality: good, records web cam and audio together but can't record audio/cam and screen simultaneously but only one after another
ScreenToaster	* Yes * Quality: good	No	* Yes * Quality: good	* Yes * Quality: good	No
Morae	See comments column	See comments column	See comments column	See comments column	See comments column
Skype	* Yes * Quality: OK	* Yes * Quality: OK	* Yes * Quality: OK	No	No
Adobe Acrobat Connect	* Yes * Quality: OK	* Yes * Quality: OK, video was jerky in one session, and there were issues getting a session started	* Yes * Quality: good	N/A (see comments column)	N/A (see comments column)
BB FlashBack + GoToMeeting	* Yes * Quality: good	* Yes * Quality: good	* Yes * Quality: good	* Yes * Quality: good	* Yes * Quality: good
Vsee + GoToMeeting	* Yes * Quality: good	* Yes * Quality: good	* Yes * Quality: good	* Yes * Quality: good	* Yes * Quality: good but records 2 separate audio/video files for user 1 and 2 (see comments for Vsee)
Morae Recorder + Skype	* Yes * Quality: OK	* Yes * Quality: OK	* Yes * Quality: OK	* Yes * Quality: good	* Yes * Quality: good

Tool evaluation results con't

	Recording screen	**Synchronous/ asynchronous**	**Chat**	**Comments**
BB FlashBack	* Yes * Quality: good	asynchronous	No	* Records screen, audio + web cam activity and mouse cursor's movement; it saves the recording output file in .fbr proprietary format that can only be played back in BB FlashBack or BB TestAssistant, it doesn't delay the PC and outputs minimum file size. * As stand alone, good tool for asynchronous testing

				since doesn't have screen sharing capabilities. * Can use in addition to screen-sharing software for synchronous testing (e.g. GoToMeeting). * Single user sessions are saved as a movie with a large round circle around the user's cursor so it's easier to see. * Good customer support provided for product
BB TestAssistant	Same as BB FlashBack	Same as BB FlashBack	No	Similar to BB FlashBack
GoToMeeting	* Yes * Quality: good	synchronous	Yes	* Screen sharing application, screen and audio recording * Easy to use, send invitation to participants, share screen, press record button when meeting starts * Can add a web cam from another application e.g. Skype, Vsee. GoToMeeting records anything in the screen, including the web cam. File is small and saved as windows media video. * The main problem with this nice application is reliability when recording. I have used it extensively and reliably so, but sometimes recording overloads the system and it could result in a lost session.
Dimdim	* Yes * Quality: good	synchronous	Yes	* Web cam, audio and screen recording are all in one file * Web-based, web conferencing tool * Had camera malfunctions on one end, but worked well when retested * Recording output is a Flash Video File (FVL) and it's made available for immediate web playback or can be downloaded in the PC.
Uservue	* Yes * Quality: not good, hard to read the details in the recorded screen	synchronous	Yes	* Starting a web session is simple, response time viewing other screen is fast; looks OK and readable but recording was not as clear to read; very easy and fast to record when session ends; saves file as WMV and Morae output for analysis using Morea's Manager. * UserVue records up to 1 hour in each session. The free trial is rather short – 14 days.
Vsee	No	synchronous	Yes	* Nice free tool, good for communication but not for recording interactions * Can be used with additional software to record screen interactions (e.g. Camtasia, GoToMeeting) * Good support, direct Vsee communication with company representative * Recording output is small, manageable AVI file * This is a great product but one problem with it is the fact that the video of the interaction is saved as two separate output video files. In my web cam recording I could only hear myself but not the participant. The participant's web cam recording had the audio for both of us.
Camtasia Studio	* Yes * Quality: good, records screen and saves file, then can record narration of that screen capture separately but can't record screen and audio simultaneously	asynchronous	No	* Can't record audio and screen simultaneously but only one after another * Records web cam and audio together but can't record audio/cam and screen simultaneously but only one after another * Records web cam and audio as WMV (audio/video file) * It's good for recording screen only and possibly combined with other software like Vsee
ScreenToaster	* Yes * Quality: good	synchronous	Yes	* Web based, no login to record, login is required to save as MOV file; makes it vulnerable because can lose the whole session without saving it. We lost one recording session because the PC froze when I attempted to save the file. * Immediate web playback of recording * Web cam, audio and screen recording are all in one file * Saving file from web takes time (saving the file consists of 'save', 'achieve', zip, and 'unzip').
				We were not able to establish a successful connection between the PC with the Morae Recorder and the PC with the Morae Observer. The two PCs we tried to connect were in a University network, because we knew that the PCs had to be in a network or have VPN. We called Techsmith for assistance but weren't able to address the

Morae	See comments column	synchronous	See comments column	connection problem. We also reached out to two usability experts who suggested a solution: to string an Ethernet cable between the two machines. This wouldn't work for us because we needed a tool for remote testing. Morae works only for PCs, and it saves the recording output file in .rdg proprietary format that can only be played back in Morae Manager. (Other usability products use standard formats like AVI, WMV, WAV, FLV). While Morae has a lot of potential, though it's expensive and it requires a fair amount of familiarity to use this product.
Skype	No	synchronous	Yes	Skype is a nice free communication tool but not reliable enough for the purpose of this project. We experienced dropped calls, problems with the video frame rate, or distorted audio.
Adobe Acrobat Connect	N/A (see comments column)	synchronous	Yes	Only the Acrobat Connect Professional allows recording.
BB FlashBack + GoToMeeting	* Yes * Quality: good	synchronous	Yes	This is a combination that worked very well and we get all recordings in one file. We recorded everything using BB FlashBack (record audio, video, and screen) and used GoToMeeting only to establish real-time communication. Output file was small. Testing was very reliable because GoToMeeting was not used for recording, so there were no system delays. We transferred the saved output files via a web storage account (Dropbox) from the participant's PC to that of the facilitator.
Vsee + GoToMeeting	* Yes * Quality: good	synchronous	Yes	This is a combination that works but is not reliable because we recorded audio and screen from GoToMeeting and get video from Vsee. GoToMeeting is recording so there may be system overload problems.
Morae Recorder + Skype	* Yes * Quality: good	synchronous	Yes	We had Morae Recorder running on the facilitator's computer, and then used Skype video conference to stream the remote desktop. The video feed was recorded via the Morae Recorder. Skype's video quality was not the best (video file which was sufficient in image quality, but the frame rate left something to be desired), but the recording of the local desktop was excellent.

Table 2: Remote usability testing tool evaluation results

I selected the BB FlashBack Express combined with GoToMeeting for our usability testing because of the availability of all our features of interest, cost, and ease of use. This combination offers recording of audio, video, and screen including mouse movements, all in a very simple way and with one output file that includes all the recordings. The BB FlashBack Express version is a free product. We completed a second round of pilot testing with the BB FlashBack Express before the actual usability testing as described in the methods section.

Research Question

This work will explore the effectiveness of the Remote Usability Evaluation and is based on the following research question:

> *Is remote audio/video communications software useful for discount usability testing of a coach's interface for a health coaching communications / delivery tool?*

Objectives

Specific Aim 1: To evaluate the interface based on the heuristic criteria.

Heuristic evaluation is the most commonly used expert-based usability evaluation. This will be accomplished using Nielsen's 10 usability principles.

Specific Aim 2: To design sample coaching tasks with measurable quantitative outcomes.

The scenarios will consist of typical tasks completed by health coaches. Measurable outcomes include performance data:

- Time to complete a task (mean, median, range)
- Number of errors (not recovered and self-recovered)
- % participants performing correctly (with and without assistance)

Specific Aim 3: To remotely test usability of interface with 3-5 coaches.

This aim will be achieved by using software for screen-sharing and observation, and recording of audio, video, and on-screen activities. The following approaches will be employed:

- Use of scenarios (assign coaching tasks and observe use)
- Use of thinking aloud protocols

- Write up notes, collapse into main findings

METHODS

We measured the usability of the Automated Health Coaching System through the evaluation of the interface by employing a combination of inspection methods and testing methods. We selected the heuristic evaluation (inspection method), followed by thinking aloud protocol and scenarios (testing method).

Participants

One evaluator conducted a heuristic usability evaluation of the Automated Health Coaching System interface based on Nielsen's ten usability principles. I carried out the evaluation on a Dell Optiplex 755 computer running Windows XP operating system.

As a trained evaluator, I am a graduate student in biomedical informatics with significant experience in medical informatics and clinical information systems, and I have completed a graduate course in human-computer interaction evaluation methods, as well as corporate training in usability evaluation. I am neither an intended user of the system, nor have I used it regularly before the heuristic evaluation. I have had a one-hour demo of the system by a health coach who was an experienced user of this system.

For the purposes of the usability testing, 5 five participants were recruited from the Automated Health Coaching System Project. Four of the participants were familiar with the coaching health behavior method, but even though they had interacted with the ORCATECH Coaching Console,

three were unfamiliar with the new technology that the Automated Health Coaching employed and one was somewhat familiar with it. One participant was a health coach who was currently using the system to coach twenty patients. The participants included a graduate student in biomedical informatics, three full-time employees in the roles of research assistants (one of them a senior research assistant), and a health coaching consultant. The health coach population was defined as 1) not requiring clinical training; 2) had experience in being health coaches; and 3) they were familiar with communicating with and coaching subjects on-line.[14]

Heuristic Evaluation

Definition of the heuristic evaluation:

> Usability experts evaluate a user interface using a set of guidelines and noting the severity of each usability problem and where it exists.

I evaluated the Automated Health Coaching System interface based on Jacob Nielsen's ten usability heuristics listed in Table 3 below, which are general principles for user interface design.

1	**Visibility of system status**	The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.
2	**Match between system and the real world**	The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.
3	**User control and freedom**	Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.
4	**Consistency and standards**	Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.
5	**Error prevention**	Even better than good error messages is a careful design which prevents a problem from occurring in the first place. Either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action.
6	**Recognition rather than recall**	Minimize the user's memory load by making objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.
7	**Flexibility and efficiency of use**	Accelerators – unseen by the novice user – may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent

		actions.
8	**Aesthetic and minimalist design**	Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.
9	**Help users recognize, diagnose, and recover from errors**	Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.
10	**Help and documentation**	Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.

Table 3: Nielsen's Ten Usability Heuristics

This included the review of the interface and providing feedback. The review of the interface consisted of using the system by going through the interface to get an overview of its functionality. Providing feedback consisted of generating a list of usability problems with severity ratings and recommended solutions for the identified problems. I used the system to complete some typical tasks and I documented the usability problems that were identified. The tasks included:

- Send a message to a patient
- View patient background
- Review action plan
- Review progress on goals
- Edit the machine generated message
- Quick view of patient panel to see if there are any life events (emergencies)

I used Nielsen's severity rating for each usability problem that I encountered. Nielsen rates the severity of usability problems using the 0 to 4 rating scale described in Table 4 below.

Ranking	**Description**	**Definition**
4	Catastrophe	Imperative to fix this before product can be released
3	Major	Important to fix, so should be given high priority
2	Minor	Fixing this should be given low priority
1	Cosmetic	Need not be fixed unless extra time is available on project
0	None	I don't agree that this is a usability problem at all

Table 4: Nielsen's severity rating in heuristic evaluation

The severity of a usability problem according to Nielsen is a combination of three factors:

- The frequency with which the problem occurs: Is it common or rare?
- The impact of the problem if it occurs: Will it be easy or difficult for the users to overcome?
- The persistence of the problem: Is it a one-time problem that users can overcome once they know about it or will users repeatedly be bothered by the problem?

Usability Testing

The purpose of this study was to evaluate the usability of the health coaching software using a protocol that combined the coaching task scenarios and the talk aloud protocol.

Coaching task scenarios:

- Scenario-based testing of health coaching system, each task has a specific goal, user observation during completion of tasks.

Talk-aloud protocols:

- Usability tests can include talk-aloud protocols, where participants are encouraged to express their thinking about their experience while using an application.

Each participant's computer was equipped with a web camera, internet connection and the BB FlashBack Express Recorder free software that was used for recording the testing sessions. The evaluator was located in the state of Wisconsin while all participants were located in the state of Oregon. Some participants underwent training on the health coaching software; the training was competed a few days before the testing sessions. The evaluator did a "dry run of testing equipment" with a Professor of Medical Informatics and Clinical Epidemiology at Oregon

Health & Science University one day before the usability testing started, and the task scenarios were finalized to include only tasks that represented realistic user goals in a clear and unambiguous way.

After being consented for the study, the evaluator set up an individual online appointment with each participant to evaluate the usability of the ORCATECH Coaching Console. The evaluator sent an e-mail invitation for the meeting and once the online meeting started at the time of the appointment, the participant shared the computer screen with the evaluator via the GoToMeeting web conferencing software which also enabled real-time audio communication. The task scenarios were distributed one at a time to each participant at the time of the usability testing, and participants were instructed to read the scenarios out loud themselves and to ask questions before beginning the testing, in order to minimize the interaction during the testing session and monitor the session impartially. The testing protocol consisted of four scenarios, each with one or more tasks. Each one of the four scenarios was recorded separately and consecutively.

Once the testing session was underway, each participant was asked to perform the task scenarios while using the Health Coaching Interface, and the BB FlashBack Express Recorder software recorded the participant's screen and mouse movement, sound, and webcam for later analysis. The evaluator remotely observed the participant's screen and interactions on screen, and listened to participant's voice during the testing session. Upon completion of usability testing sessions, the saved output files were transferred from the participant's computer to the evaluator's computer via a web storage account (www.Dropbox.com). BB FlashBack Express Player was then used by the evaluator to analyze the recordings.

RESULTS

Heuristic Evaluation

The Automated Health Coaching System interface problems are described below, along with recommendations for interface improvements. For each usability problem, Appendix B lists the results of the heuristic evaluation that include a problem identifier, the location of the interface problem, a description of the problem, Nielsen's severity rating, recommendations and the violated heuristic principle for each of the ten categories.

Nielsen recommends that the evaluator should try to be as specific as possible and should list each usability problem separately instead of lumping them together because there is a risk of repeating some problematic interface aspects unless one is aware of all its problems, and it may not be possible to fix all usability problems but it could be possible to fix some of the problems if they are all known.[15] Therefore, when several separate problems affected a specific aspect of the interface, I defined all the problems individually and offered recommended solutions. Table 5 summarizes the total number of usability violations and their severity ratings for all ten of Nielsen's usability heuristics.

Usability heuristic	Cosmetic	Minor	Major	Catastrophic	Total
Visibility of system status		1	9	1	**11**
Match between system and real the world		2	12	1	**15**
User control and freedom		1	1	2	**4**
Consistency and standards	1	10	19	4	**34**
Error prevention		1			**1**
Recognition rather than recall			4	2	**6**
Flexibility and efficiency of use	1	4	2		**7**
Aesthetic and minimalist design	3	11	14	2	**30**
Help users recognize, diagnose, and recover from errors			1		**1**
Help and documentation				1	**1**
Total	**5**	**30**	**62**	**13**	**110**

Table 5: Violations and severity ratings by usability heuristics

The heuristic evaluation identified a total of 110 usability violations across Nielsen's ten usability heuristics. More than 80% of all violations fell into four of the heuristic heuristics: principle #4: consistency and standards (e.g. inconsistent naming convention, inconsistent use of double click functionality, inconsistent location of icons, inconsistent font size); principle #8: aesthetic and minimalist design (lack of tooltips to help user, date format uses too much real estate, navigation difficulties due to disorganized heading layout, data duplication); principle #2: match between system and the real world (e.g. failure of the system to build a consistent model in user's mind, and non-standard use of the checkbox or the hyperlink); and principle #1: visibility of system status (e.g. failure of the system to inform participants that an action took place, failure of the system to provide visual confirmation – not changing the cursor shape for clickable items in the interface). The remaining violations fell into the other six usability heuristics. Overall, the usability violations were present in all ten of Nielsen's usability heuristics.

More than 50% of all the violations were major (severity = 3), 27% were minor (severity = 2), and the remaining were cosmetic (severity = 1) or catastrophic (severity = 4) violations.
The results of the heuristic evaluation were provided to the interface development team in the form of a usability evaluation findings report and a series of screen capture illustrations, in order for the team to address the usability problems and redesign the interface. I also provided the development team with a list of positive design features they had implemented in the existing interface, which are shown in Table 6.

Feature	**Positive**
Progress clock	System feedback
Blue highlight when moving cursor	Informs users where they are
Red X is a good and standard metaphor for "Delete"	Consistency and standards
Customizable font size	Accommodates different categories of

	users
Pop-up windows can be moved	It does not obscure the area beneath
User can move between entries using the TAB key	Consistency and standards
User-configurable settings/preferences	Empowers users
Menu bar (title bar) at top	Shows where you are, in what bin
In messages, there is "Reset" to set fields to defaults	Saves time and allows users to start over
Loading in start screen	System feedback
Automatic time out	System feedback
Warning before completing delete action	System feedback
Sorting of columns	Necessary functionality

Table 6: Positive user interface design features in the Automated Health Coaching Interface

Usability Testing

Usability testing of the Automated Health Coaching interface followed the heuristic evaluation. Although the results of the heuristic evaluation were provided to the interface development team in order to address the usability problems and redesign the interface, the recommended changes could not be implemented prior to the usability testing. Therefore, the usability testing results exclude the usability problems already identified by the heuristic evaluation and they consist of a distinct set of usability problems that were not picked up by the heuristic evaluation. The BB FlashBack Express Recorder software recorded the participant's screen and mouse movement, sound, and webcam without any technical problems. A description of the task scenarios used for the purpose of usability testing is provided in Appendix C. Scenario #1 has two tasks (1.1 and 1.2), scenario #2 has one task (2), scenario #3 has three tasks (3.1 – 3.3), and scenario #4 has five tasks (4.1 – 4.5). In this study, I measured performance data such as the elapsed testing time on task, the number of errors made per task, and task accuracy.[16,17] The performance data was generated by observing the usability testing session in real-time and also from analyzing the session recordings after the testing session finished.

The elapsed testing time on task is the time that the participant spent on each task regardless of success in completing the task. I measured the mean time on task (average time that all

participants spent on the task), the median time on task (the time in the middle position in the list of all participants' times spent on the task), and the high/low time range (highest and lowest time that all participants spent on the task). All times were reported in minutes and seconds. There was no specific allotted time for completion of tasks during the usability testing, and success was not measured for completion within the expected time but rather for completing the task regardless of time. The task timing performance data from all the usability testing sessions is summarized in Table 7, and the elapsed task times of all participants are shown in Table 8. Two of the participants who had more experience with the interface had lower task test times than the other three who didn't have as much training.

Some participants typed extensively during the scenario testing sessions, while others did so quite briefly. Since the typing during a scenario merely served the purpose of accomplishing the task of locating a specific area in the interface, the elapsed testing time on task was normalized for all participants. The normalization was done for the (a) typing time, (b) the time during participant-evaluator interactions not related to the tasks of the scenario being timed, and for (c) the transition time between different tasks within the same scenario (e.g. transition from 1.1 to 1.2).

Task	Mean Time	Median Time	Time Range
1.1	2:38	3:14	1:22 – 3:37
1.2	2:57	3:15	1:16 – 4:24
Total scenario 1	**5:35**	**5:46**	**2:58 – 7:20**
Total scenario 2	**3:02**	**4:32**	**4:25 – 5:58**
3.1	3:11	3:21	1:39 – 4:13
3.2	3:01	3:01	2:14 – 4:08
3.3	1:30	1:27	0:49 – 2:17
Total scenario 3	**7:42**	**7:34**	**6:07 – 9:26**
4.1	3:03	3:36	1:01 – 4:31
4.2	0:41	0:46	0:14 – 1:09
4.3	0:55	0:59	0:28 – 1:13
4.4	0:51	0:50	0:28 – 1:43
4.5*	0:32	0:30	0:28 – 0:39
Total scenario 4	**6:02**	**6:39**	**2:52 – 8:47**

Table 7: Elapsed testing time on task (minutes:seconds)

* Task 4.5 was given to four participants (added after the 1st testing session)

Task	Participant 1	Participant 2	Participant 3	Participant 4	Participant 5
1	2:58	6:30	5:19	7:20	5.46
2	4:30	4:32	5:44	4:25	5.58
3	7:34	6:07	6:43	8:38	9:26
4	2:52	4:59	6:50	8:47	6:39
TOTAL	**17:54**	**22:08**	**24:36**	**29:10**	**29:49**

Table 8: Elapsed time by task by participant (in minutes: seconds)

The task accuracy performance data consisted of the percentage of participants performing correctly (without and with assistance) and the number of errors made per task (errors where the participant did not recover, the errors where the participant recovered, and the total number of errors per task). Task errors include both errors resulting from the interface making it difficult for participants to complete their tasks, and participants -errors of omission and commission. The percentage of participants performing correctly is the number of participants completing the task divided by the total number of participants.

All participants successfully completed some of the tasks, e.g. locate and access subjects; locate issues; locate life events; add or copy life events; review subject's performance against goals for a previous week; and set new goals for the game targets. The usability testing identified problems with some of the tasks, e.g. locate contacts; select the correct choice for contact duration, contact type and contact reason; locate the active topic; add a comment in an existing issue; view and interpret a performance graphic; verify creation of new contacts, issues or messages. Participants consistently had problems with performance graphics (task 3.2) such as: not displaying current play counts though this exists in "Game targets", even when users correctly selected the date for the week in question; unclear how to read the graphics and when the patient played. Additionally, participants consistently had problems adding a comment in an existing issue (task 4.1) because most participants logically attempted to do so by editing the

issue “Description” field which is not an editable field; the “Comment” field was largely ignored due to its obscure location. Locating a subject’s active topic (task 4.2) presented challenges because it’s not obvious where this is positioned in the interface. When asked to document interactions with subjects (task 1.1) some participants consistently had trouble finding where to do so in the interface, exposing the ambiguity of using the word “Contact” for this purpose. Task 1.2 showed that selecting the correct issue type is a consistent problem for participants, warranting a revisit of this area of the interface; also, newly-created issues were not displayed in issue history, forcing participants to search for workarounds or to make errors (e.g. to re-create the issue being mislead to thinking that it didn’t work the first time). Appendix D shows a qualitative description of the problems identified via the usability testing of Automated Health Coaching interface (excluding problems identified by the heuristic evaluation). I provided recommendations only for the usability problems tied to errors related to the interface; no recommendations were provided for user-errors that were exclusively due to user oversight or due to user unfamiliarity with the interface.

The statistics for the usability testing errors and task accuracy are shown in Table 9, while the usability testing task errors by participant are shown in Table 10. This study suggests that interface training is needed and it appears to make a difference. The participants with more interface experience committed less errors during testing compared to the less experienced participants.

Task per Scenario	% Participants Performing Correctly (without Assistance)	% Participants Performing Correctly (with Assistance)	# Errors (not recovered)	# Errors (self-recovered)	Total # errors
1.1*	60 %	60 %	7	5	12
1.2	80 %	100 %	13	3	16
Total scenario 1	**70 %**	**80 %**	**20**	**8**	**28**
Total scenario 2	**80 %**	**100 %**	**15**	**2**	**17**
3.1	100 %	100 %	3	1	4
3.2	40 %	40 %	6	1	7

3.3	100 %	100 %	2	1	3
Total scenario 3	**80 %**	**80 %**	**11**	**3**	**14**
4.1	40 %	80 %	16		16
4.2	60 %	80 %	3		3
4.3	100 %	100 %	2		2
4.4	100 %	100 %	1		1
4.5	100 %	100 %	3		3
Total scenario 4	**80 %**	**92 %**	**25**		**25**
TOTAL	**77.5 %**	**88 %**	**71**	**13**	**84**

Table 9: Task Accuracy and number of errors

* No assistance was offered in Task # 1.1 (became apparent after the 1st testing session)

Task	Participant 1	Participant 2	Participant 3	Participant 4	Participant 5
1	5	6	7	4	6
2	4	2	4	4	3
3	4	2	3	2	3
4	0	3	6	9	7
TOTAL	13	13	20	19	19

Table 10: Number of errors by task by participant

DISCUSSION

Our study showed that conducting actual pre-test tool evaluation trials is a useful step that provided an opportunity to understand the ease of use, features, quality and reliability of the software. The tool chosen for the usability testing - BB FlashBack Express Recorder - recorded the participant's screen and mouse movement, sound, and webcam without any technical problems. I then analyzed the recordings with BB FlashBack Express Player that similarly performed flawlessly.

Even though there were many products available for remote usability testing, a combination of two products satisfied our tool requirements better than a single one. This is consistent with another study reporting that the combination of different tools in moderated remote usability activities seems to be a good solution for most research designs.[13] I selected the BB FlashBack

Express combined with GoToMeeting for our usability testing because this combination offered all our features of interest (recording of audio, video, and screen including mouse movements) in a simple and reliable way, without system delays and all the recordings were automatically grouped into a single output file. The BB FlashBack Express version is a free product that was sufficient for the purpose of this study, while the Professional version would be considered for more features and editing capabilities. The individual products tested for this study produced some of the following issues: didn't offer all the required features (e.g. couldn't record audio and screen simultaneously), were vulnerable to technical problems (e.g. unable to establish connection between PCs), saving file from web took too much time, some products were quite expensive and required familiarity with them, some tools were not reliable for testing (e.g. dropped calls, distorted audio, or an entirely lost testing session).

This study reaffirms the heuristic evaluation as an efficient discount usability engineering method for easy, fast and inexpensive interface evaluations early in the development cycle, which is consistent with what other authors have reported.[15,18,19] The heuristic evaluation successfully identified a total of 110 usability violations across all ten of Nielsen's usability heuristics, particularly violations of consistency and standards, aesthetic and minimalist design, match between system and the real world, and visibility of system status.

We measured the usability of the Automated Health Coaching interface through the heuristic evaluation of the interface followed by usability testing because in addition to overlapping problems identified by both methods, heuristic evaluation and usability testing identify distinct sets of usability problems not picked up by the other method, thus complimenting each-other. This study first employed a heuristic evaluation to identify and eliminate the initial layer of

interface problems without extensive use of resources. This method identified 110 interface usability problems. Usability testing was then employed to identify usability problems overlooked by the heuristic evaluation, and this resulted in an additional set of 36 interface usability problems. For this reason it is recommended that both methods be used.[15,20,21]

Five participants were recruited in this study, consistent with Nielsen's model that places the number of users required for the maximum cost-benefit ratio in usability testing between three and five, who find most of the usability problems.[22,23]

This study successfully identified a wide range of usability problems while the evaluator worked remotely, suggesting the effectiveness of the remote usability evaluation of a health coaching interface via remote audio/video communications. This is consistent with other studies on effectiveness of synchronous usability evaluation. [11-14] Even though the heuristic evaluation identified 110 interface usability problems, the remote synchronous usability testing identified 36 more new usability problems that were not evident during the heuristic evaluation.

There were some noteworthy observations regarding novice vs. experienced participants. While for some tasks most or all participants encountered problems, the experience of participants with the interface revealed problems of different nature, e.g. the experienced participants had developed workarounds for some interface problems whereas the novice participants exposed such problems during testing. For this reason I would argue that it is better to have different levels of participant competency with the interface for usability testing, both novice and experienced. This study suggests that interface training is needed and it appears to make a

difference. The two experienced participants had lower task test times and committed less errors during testing than the other three who didn't have as much training.

One limitation of this study is that only one evaluator was involved with the heuristic evaluation. Nielsen[15] argues that multiple evaluators are needed to do a heuristic evaluation in order to achieve a better performance, and he recommends using three to five evaluators because a single evaluator finds only 35 percent of the usability problems (ranged from 19 percent to 51 percent). Multiple evaluators also provide more reliable severity ratings because the quality of the mean severity rating increases.

Another limitation was that although the results of the heuristic evaluation were provided to the interface development team in order to address the usability problems and redesign the interface, the recommended changes could not be implemented prior to the usability testing. Therefore, we were not able to take advantage of the improved interface design for the usability testing that followed the heuristic evaluation. A study by Joshi et al[18] that the benefits of implementing the recommended changes before the next evaluation cycle include higher ease of use and greater ease of navigation, while minimizing errors.

Also, following up with a posttest questionnaire could have generated valuable participant feedback.

Future work could include evaluating the cost-effectiveness of remote usability testing because no cost analysis was done in this study. I would also like to analyze the usability testing data at a finer level of granularity to learn more about where the time was exactly spent during testing, e.g. time on task, time to read, time to search, interaction time with evaluator, etc. Adding an inquiry method like a post-test questionnaire may be of interest in exploring participants' views

on satisfaction with the interface, ease of use, efficiency, etc. Another area of interest could be to explore new remote usability tools and to conduct a systematic review of the effectiveness of such tools in remote usability testing. It would be of interest to conduct a follow-up usability evaluation study after the usability problems have been addressed.

SUMMARY AND CONCLUSIONS

The heuristic evaluation of the interface revealed a large number of usability violations mostly concerning the lack of consistency and standards, lack of an aesthetic and minimalist design, failure of the system to match the real world, and lack of visibility of system status. The usability testing identified problems with some areas of the interface particularly with viewing and interpreting the performance graphics, editing of issues, locating specific items in the interface like a subject's active topic or contacts to document interactions with subjects, or making the correct selections (e.g. for contact duration, contact type, contact reason, and issue type).

The results of this study suggest that:

- It is helpful to conduct pre-test tool evaluation trials before choosing the tool for the usability testing.
- Though many products are available for remote usability testing, a combination of two products satisfied our tool requirements better than a single one.
- The heuristic evaluation is an efficient and inexpensive method for interface evaluations early in the development cycle.
- Combining the heuristic evaluation of the interface with usability testing is more effective than using only one of them because they compliment each-other. In addition to

overlapping problems identified by both methods, heuristic evaluation and usability testing identify distinct problems not picked up by the other method.

- Remote usability evaluation and specifically the remote synchronous usability testing is an effective method to improve interface usability problems.
- It is better to recruit both novice and experienced participants because they reveal problems of different nature.
- Interface training would benefit the intended users of the Automated Health Coaching System. The more experienced participants had lower task test times and committed less errors during testing than those who didn't.

The findings of this study demonstrate that the heuristic evaluation and remote synchronous usability testing are efficient methods for interface evaluation and design improvement. Remote usability testing appears to be a feasible method for the evaluation of health coaching software, allowing the testing of users in their own natural environment and more frequently via iterative development cycles. This type of usability testing was easy to arrange and inexpensive.

APPENDIX A: Consent Form – Coach Usability Testing

Oregon Health & Science University
Consent Form

IRB#: ____3751_____

Protocol Approval Date: 12/25/2009

[*Ensure the initial/annual approval date is inserted into the stamped approved consent form from the IRB***]**

OREGON HEALTH & SCIENCE UNIVERSITY

Consent Form – Coach Usability Testing

TITLE: Automated Health Coaching

PRINCIPAL INVESTIGATOR: Holly Jimison, Ph.D. (503) 418-2277

SPONSOR: National Institute on Aging, Alzheimer's Association, Intel

PURPOSE:

You have been invited to be in this research study because you have interacted with the ORCATECH Coaching Console. The purpose of this study is to evaluate the usability of this coaching software.

This study requires 1-2 online sessions with a researcher who will remotely view your computer while you are performing some tasks using the coaching software. 5 coaches will be asked to participate in this study.

PROCEDURES:

After being consented for the study, the researcher will set up an appointment to evaluate how you use the ORCATECH Coaching Console. This appointment will take place over the

computer. You will receive an e-mail to connect to an online meeting and will share your computer screen with the researcher. During the session, your screen and audio will be recorded for later analysis.

Once you are connected to the online meeting with the researcher, you will be asked to complete 3-5 common coaching tasks (such as assigning a new week of activities or updating a contact with a participant).

If you have any questions regarding this study now or in the future, contact Holly Jimison, (503) 418-2277.

RISKS AND DISCOMFORTS:

There are no risks associated with this study.

BENEFITS:

You may or may not personally benefit from being in this study. However, by serving as a subject, you may help us learn how to benefit coaches using the ORCATECH Coaching Console in the future.

ALTERNATIVES:

You may choose not to be in this study.

CONFIDENTIALITY:

We will not use your name or your identity for publication or publicity purposes.

The screen capture video and audio recording from this study will be kept until the end of the study. Once the study is completed, the files will be deleted. No identifiable audio, video or photographs from this study will be displayed in a public setting.

COSTS:

There is no cost for participating in this study.

LIABILITY:

If you believe you have been injured or harmed while participating in this research and require immediate treatment, contact Holly Jimision PhD, (503) 418-2277.

You have not waived your legal rights by signing this form. If you are harmed by the study procedures, you will be treated. Oregon Health & Science University does not offer to pay for the cost of the treatment. Any claim you make against Oregon Health & Science University may be limited by the Oregon Tort Claims Act (ORS 30.260 through 30.300). If you have questions on this subject, please call the OHSU Research Integrity Office at (503) 494-7887.

PARTICIPATION:

If you have any questions regarding your rights as a research subject, you may contact the OHSU Research Integrity Office at (503) 494-7887.

You do not have to join this or any research study. If you do join, and later change your mind, you may quit at any time. If you refuse to join or withdraw early from the study, there will be no penalty or loss of any benefits to which you are otherwise entitled.

The participation of OHSU students or employees in OHSU research is completely voluntary and you are free to choose not to serve as a research subject in this protocol for any reason.

If you do elect to participate in this study, you may withdraw from the study at any time without affecting your relationship with OHSU, the investigator, the investigator's department, or your grade in any course.

We will give you a copy of this form.

SIGNATURES:

Your signature below indicates that you have read this entire form and that you agree to be in this study.

OREGON HEALTH & SCIENCE UNIVERSITY

INSTITUTIONAL REVIEW BOARD

PHONE NUMBER (503) 494-7887

CONSENT/AUTHORIZATION FORM APPROVAL DATE

Apr. 7, 2010

Do not sign this form after the Expiration date of: 12/24/2010

_____

Participant Name (printed)

_____ _____

Signature of Participant Date

_____ _____

Principal Investigator

APPENDIX B: Results of Heuristic Evaluation of Interface

ID	Location	Problem	Severity	Recommendation	Violated heuristic
1	Home page	Home page is wasted space, mostly blank	3	Space could be used to welcome and inform users	#1
2	Home page	C2 in center of screen is technical jargon that should be avoided	3	Spell out the name of the application	#2
3	Home page	Build: 2.2.607 is technical jargon, not relevant to coaches	3	Move this away from center of screen	#2
4	Main menu	Inconsistent font size of main menu items (issues, subjects vs. notices, preferences)	2	Should use the font size consistently for all menu items in interface	#4
5	Main menu	Inconsistent font color of main menu items (issues, subjects vs. notices, preferences)	2	Should use the font color consistently for all menu items in interface	#4
6	Main menu	Inconsistent use of gray highlight in main menu (issues, subjects vs. notices, preferences)	2	Menu items need to be consistent and uniform	#4
7	Main menu	Main menu items too far from each other (issues, subjects vs. notices, preferences)	2	Should display as siblings in menu, next to each-other	#8
8	Subjects submenu	Some tooltips are label duplicates, they do not serve their intended purpose	2	Elaborate in tooltips, do not duplicate labels	#8
9	Subjects submenu	Some tooltips are useless because the only difference between them and labels is the word 'tab', they do not serve their intended purpose	2	Elaborate in tooltips so they serve the purpose of informing the user, do not duplicate labels	#8
10	Subjects submenu	Tooltip is an abbreviations of label	2	Tooltips should be extensions of labels, never abbreviations of labels	#8
11	Issue details	Red X "Delete" metaphor is missing tooltip, user may wonder what this is	3	Needs tooltip (e.g. tooltip that reads 'delete item') to help user by explaining the intended action of the metaphor	#8
12	Preferences submenu	Some menu items are missing tooltips	2	Add tooltips	#8
13	Main menu	Main menu tooltips do not serve their intended purpose, they are redundant with labels and don't help user by providing information on what's inside the menu items	2	Elaborate in tooltips, inform user what content is hidden in each menu tab	#8
14	Application	A currently selected item of menu/submenu is grayed out. Gray out is traditionally used for functionalities that are inactive, so this is counterintuitive and confusing	3	Follow convention chosen in 'Issue' list or "Subject' list, use highlight to indicate current step	#1
15	Subjects/ Preferences	Inconsistent font color of submenu items (e.g. for a subject, if 'personal' is clicked, 'personal information' is black; in 'preferences' → 'notifications', 'notification options' is displayed in blue font)	2	Should use the font color consistently for all submenu items in interface	#4
16	Issue details	Blue highlight to indicate current step is excellent but if I click an issue and then move the cursor up/down the list, the issue I clicked remains highlighted and my current cursor location is also highlighted, so I get two highlighted issues. This is confusing and it defeats the original paradigm of telling the user where he is.	3	Consider using highlighting only to track user's current selection (only one highlight at a time)	#1
17				Allow one open window at a time	

	Subjects/ Issue details	Multiple windows open simultaneously clutter the screen and it is difficult to form a mental map of the system structure (menu items remain open when navigating through them)	3	(currently open window only). When user leaves one menu item and goes to another, close the prior one automatically to keep application cleaner and simpler.	#2
18	Main menu	Initially there are four main menu items: issues, subjects, notices, preferences. After navigating them, they also display at the bottom left as 'currently open windows' and are clickable. Providing system feedback is excellent but providing different ways of doing the same thing can lead to confusion.	3	Do not duplicate menu items, consistently use highlight paradigm to provide real-time system response for currently selected items in interface	#2
19	Issue details	Issue details screen is very cluttered and busy, items are not properly aligned/indented, not pleasing to the eye	3	Change to well-defined fields that are grouped together and easy to scan down; use aesthetic and minimalist design (see proposed solution)	#8
20	Issue details	Cursor remains unchanged upon mouse-over the red X, so user doesn't know an action could take place by clicking the X	3	Cursor should change from arrow to hand when moving over the red X throughout the interface, to show action will take place upon clicking icon	#1
21	Issue submenu	Inconsistent location of deletion metaphor: Red X is located on the left in issue 'files', but on the right in issue 'details'	3	Keep red X in a small box on the right of the field for consistency throughout the interface, because other metaphors in the interface are located on the right (e.g. dropdown arrow, up/down sort arrows, calendar)	#4
22	Issue submenu	Red X is a good and standard metaphor for "Delete" it is next to file name without aesthetic boundaries	1	Separate red X from file name by putting X in a small box on the right of the field	#8
23	Issue details	Confusing location of deletion metaphor: Red X is located between the scheduled (date) field and the hour field, and it's not clear which of these two it should delete	3	Define fields clearly with enough space from each-other to avoid confusion	#6
24	Issue details	Mouse over the calendar icon does nothing, cursor remains unchanged, so user doesn't know an action could take place by clicking icon	3	Cursor should change from arrow to hand when moving over the calendar, to show action will take place upon clicking icon	#1
25	Issue details	Calendar metaphor is missing tooltip, user may wonder what this is	3	Needs tooltip (e.g. tooltip that reads 'enter date') to help user by explaining the intended action of the metaphor	#8
26	Issue details	Date format uses too much real estate	3	Change to standard compact date format. Verify whether currently date/time format is necessary or date-only format is sufficient (how is 'time entered' unnecessary to track bugs/issues?)	#8
27	Issue filter	Inconsistent naming convention; space vs. no space for filter choices (e.g. IssueId or SubType vs. 'Assigned To' or 'Last Change' or 'Opened By')	2	Use consistent language and either remove or keep the space for all items. Recommended to use space as separator because these are labels, not code variable in a programming language.	#4
28	Issue filter	"Currently opened" heading is duplicated	3	Remove header duplication, use tabs instead of headings	#2
29	Application	Menus are sometimes vertical and other times horizontal (horizontal headings not well demarcated)	3	Change metaphor to use tabs instead of horizontal headings	#2
30	Issue	Inconsistent appearance of date format	2	Interface should be uniform to be	#4

	details			easily understandable by users	
31	Issue details	Label and data field mismatch: 'Date' label for 'Date/ time' field	3	Match label and data field correctly	#4
32	Issue details	Inconsistent naming convention: "Opened by" describes who entered the issue, "Create Date" describes date issue was entered; labels need to match	3	Recommend "Entered by" and "Date entered", but any other naming is OK as long as it's consistent (e.g. "Opened by" and "Date opened"; or "Created by" and "Date created")	#4
33	Application	Cursor turns into a progress clock that indicates an action is taking place and user has to wait. Excellent feature because it provides system feedback to user but as user moves the mouse around, so does the clock and user has to look for it.	2	Consider placing it (or progress bar) in fixed more conventional location (e.g. left lower corner of screen) for consistency; user doesn't have to search for it	#1
34	Application	Progress clock that shows the system is thinking has a color that blends into the interface colors	3	Consider blue or green color for progress clock (bar)	#8
35	Issue details	Inconsistent use of checkbox: it does different actions in different locations in interface; users will wonder what checkbox does every time they click it	3	Use checkbox consistently in interface; stick with standard and traditional use of checkbox (which is to check a box to make a selection)	#4
36	Issue details	Non-standard use of checkbox: used here to open a comment box in addition to its standard use of checking a box to make a selection	3	Do not use a checkbox to enter a comment, use standard comment textbox to enter it	#2
37	Application	Error message is a command line not written in plain language	2	Error message should provide information on what action the user should take	#5
38	Issue submenu Issue details	Inconsistent font color: 'Double click to view' is in blue in issue detail, and gray in targets	2	Should use the font color consistently for similar items in interface	#4
39	Issue details	Inconsistent use of hyperlink: it does different actions in different locations in interface; users will wonder what blue-font text does every time they click it	3	Use hyperlink consistently in interface; stick with standard and traditional use of hyperlink (as clickable item)	#4
40	Issue details	Non-standard use of hyperlink: though it looks like a hyperlink and is next to another blue text that is a hyperlink, this one is not and is used here only to provide a message	3	Do not make a blue text message look like a standard hyperlink if it's not; reserve this for hyperlinks	#2
41	Issue submenu Issue details	Confusing messages at confusing locations; not clear where the user should click	3	Revisit these message paradigms; consider using standard paradigm of putting the messages into tooltips and displaying them with mouse-over	#2
42	Issue details	Mouse over the home address hyperlink remains unchanged, so user doesn't know an action could take place by clicking	3	Cursor should change from arrow to hand when moving over the home address hyperlink, to show action will take place upon clicking icon	#1
43	Issue details	Clicking the hyperlink of home address opens a blank popup	3	Complete missing information or consider removing hyperlink if it serves no purpose	#2
44	Issue details	Tooltip missing in hyperlink, doesn't inform user what to do or what to expect to find inside the hyperlink	2	Add tooltips for all hyperlinks in interface to provide information to user (to understand this is a hyperlink, and what's inside it)	#8
45	Issue details	Tooltip completely covers the dropdown list of user choices	3	Reposition tooltip not to obscure the viewing area	#8
46	Issue submenu	Too many clicks to learn that a field is required (e.g. add event in life events and	3	Keep 'required fields' displayed at all times in interface	#6

		forget to enter 'start date'). No way for user to know during data entry where the required fields are.			
47	Issue submenu	Non-standard use of tooltip; it's used to inform user and not for system response	3	Tooltip should inform user (e.g. 'enter date' here)	#2
48	Issue details	Headings are in regular font, selection choices that belong to them (children) are bold; it should be the other way around	3	Headings should be bold; choices within them should be regular font (not bold)	#7
49	Issue submenu	Inconsistent use of 'delete'; some items in the list of contacts can be deleted (have red X), others cannot (don't have the red X)	3	Use 'delete' consistently, either make it available for all items in the list or none	#4
50	Open issues	Many issues with 'Closed' status are found listed inside the 'Open issues" menu heading; confusing and leads user to believe that only currently open issues are in this list	4	Modify "Open issues" heading to remove the word 'Open' if issues of all status types will remain present, including the 'Closed' issues	#4
51	Open issues	Fixed issues that are closed cannot be easily spotted or scanned visually without using filter	3	Gray out the closed issues so user can easily identify them at high level	#1
52	Issue filter	When filtering for issue status 'Opened/Activated' and I click any of the issues in returned results, the status is set to 'Update' inside the issue. User shouldn't have to wonder whether different words mean the same thing.	4	Do not use multiple word labels to indicate the same status choice. Remove 'Opened' status choice because unless an issue is marked as 'Closed', it is understood to be opened (status choices: e.g. assigned, updated).	#4
53	Issue filter	List of issue status choices in 'Issue Filter' do not match with issue status choices in 'Status' inside the issue	4	Maintain standard list of status choices and stick with it throughout the interface	#4
54	Issue filter	No filter is available for 'Re-open' status	4	Should be able to filter for any status choice, add this status in status filter	#8
55	Issue filter Issue details	Inconsistent labels and verb tense for issue status choices in issue status filter and in issue status inside issue (past tense vs. present tense)	3	Use the same labels consistently throughout the interface for issue status	#4
56	Issue filter Issue details	One action is represented by different labels at different locations. 'Status' is one of the issue fields but this field is not present in the displayed filter columns, 'Last change' is displayed instead and it represents status. User shouldn't have to wonder whether different words mean the same thing.	3	'Status' label should be consistently represented by one word (Status) and remain unchanged. I understand 'Last changed' to mean one of the status subtypes (choices) which is a sibling of 'Assigned' or 'Closed', but not a sibling of the 'Status' category itself.	#4
57	Open issues Issue details	Issue status in the issue list doesn't match issue status inside the issue	3	A standard set of status choices should be used consistently in interface. Status displayed outside the issue should be the same as status inside the issue.	#4
58	Open issues Issue details Issue filter	There are three different issue status lists of choices and all have differences and inconsistencies. Reconcile various issue status choices into one standard list. Currently have issue status in filter, issue status in details, and 'Last change' + 'Resolution' in issue list; all represent status choices.	4	Create one standard list of 'Status' choices that follows issue tracking workflow, remove 'Resolution' field. Revisit issue tracking workflow to add a verification step before closing the issues.	#4
59	Issue details	No need to use colon (punctuation mark) in headings	2	Headings should be bold; choices within them: regular font (not bold), remove colon	#7
60	Application	System expires after 30 minutes of	3	In the 'Welcome' screen, inform user	#1

		inactivity (great feature) but user doesn't know about it and only gets a 30 second warning at far lower left corner of screen.		that system will time out in 30 minutes so they know what to expect	
61	Open issues	There shouldn't be duplicated issues. User shouldn't have to track two identical issues but just one.	2	Recommend 'Duplicate' choice should not exist. If 2 issues duplicate each-other, put a comment in issue 2 description to say this duplicates issue 1, and close issue 2.	#2
62	Issue submenu	Dropdown list of choices has blank choice available ('SubType' in a specific subject's issues)	3	Replace blank choice with a 'None' choice to inform user	#8
63	Contacts	In 'View contacts', mouse-over displays tooltip with content of contact for 10 seconds but many of contacts are very long and it's impossible to read in 10 seconds. There isn't an 'edit' function for contact and no other way to read the contact but to keep revisiting it and read in pieces where you left off.	4	Need 'Edit' functionality for contacts, similar to that in 'Issues'	#1
64	Contacts	Contradictory and confusing: message states 'click date to change it' and when I click it I get the message '...date not editable'	4	Remove 'click date to change it' message; add 'Edit' functionality for contacts	#2
65	Contacts	Cannot edit contact	4	Need 'Edit' functionality for contacts, similar to that in 'Issues'	#3
66	Open issues Subject list Contacts	Inconsistent use of standards. 'Issues', 'Subjects' and 'Contacts' are lists but have inconsistent functionalities (can edit issue but not contact; can double click to open issues or subjects but not contacts), and inconsistent headings ('Open issues' vs. 'Subject list' vs. 'Contacts' or 'Contact history'.	3	Use consistent headings: issue list, subject list, and contact list. Use consistent functionalities: use highlight to mark current item, allow double click to open items in list.	#4
67	Contacts	Inconsistent use of double click functionality. If double click in a contact at the 'date' field – get message "Coaching dates are not editable". If I click on any other field of the same contact, which are also not editable, I do not get any message and I don't know what's happening because I get no feedback at all.	3	Use double click functionality consistently, allow for entire 'Contact' similar to 'Issues' and 'Subjects'	#4
68	Contacts	Inconsistent use of 'Edit' for dates inside 'Contacts'. Some dates are not editable (and get message 'Coaching dates are not editable') while others are editable.	3	Use 'Edit' functionality consistently	#4
69	Contacts	Reorganize 'Contact' headings, tabs and fields for easier navigation	2	Consider standard buttons to provide 'Edit' functionality for contacts	#8
70	Targets	Disorganized heading layout in 'Targets'. Not easy to navigate through headings that are all over the screen.	3	Use standard submenu layout throughout interface. Instead of headings that are all over, change metaphor to use standard tabs next to each-other at the top, open one at a time. Remove duplicate headings.	#8
71	Life events	Inconsistent use of double click functionality and 'Edit' functionality. In 'Life events', double clicking a specific field in a row will edit that individual field only. Confusing because it's a different paradigm than double clicking	3	Need to use double click and 'Edit' consistently with other areas of interface. Upon double click, open popup where user can edit entire event.	#4

		an 'Issue' which edits entire issue and opens a popup.			
72	Life events	Ambiguous location of delete icon in 'Life Events', located under 'Start date' column but deletes entire event when clicked	4	Move red X to far right of row, or use standard buttons to provide 'Edit' functionality	#6
73	Issue details	Issue 'Type' choices are in fact technology issue types. There is no "Coach" type issue, currently these issues are classified as "Other"	3	Add a choices for 'Coach' in issue type	#8
74	Subjects submenu	Data duplication: 'Personal information' and 'Participant info'	3	Merge 'Personal information' and 'Participant info' tabs	#8
75	Issue details Subjects submenu	Incorrect terminology used in headings: SubType, Info	3	Replace with Subtype, Information. The latter should also be modified in 'Preferences'.	#8
76	Life events	Dropdown list of choices has blank choice available (e.g. Add event, secondary field is blank)	3	Replace blank choice with a 'None' choice to inform user	#8
77	Info	Inconsistent submenu layout: For every menu heading, submenu headings are either next to each other and opened one at a time, or several are opened at the same time.	3	Use standard navigation throughout the interface: main menu headings aligned vertically, their submenu tabs aligned horizontally	#4
78	Issue filter	Only some of the names present in the 'Assigned to' field in issues are also present in the 'Assigned to' filter	3	Use one standard list of names across the interface	#4
79	Issue details	Not clear what 'Remote' means in issue detail	3	Add tooltip to explain what 'Remote' means	#8
80	Issue details	'Subjects' textbox in issue details has only one name in some cases but it's too big and not pleasing to the eye	1	Fit textbox to its content	#8
81	Issue filter	No filter option in issue details for 'Remote' or 'Source'. User may wish to filter all issues where source = e-mail, or find all issues where 'Remote' is present	4	If field is in issue detail, it should be filterable	#8
82	Application	No "Help" functionality	4	Consider minimally adding FAQ for the main user tasks, or put this information in the 'Welcome' screen	#10
83	Application	Users cannot reverse their actions. When user makes a mistake, e.g. clicks 'Add entry' to add 'Target schedule' and changes mind, there isn't a way to back out	4	Consider 'Cancel' at the level of a single action, or 'Undo'	#3
84	Scheduled contacts	No way for user to know during data entry where the required fields are in 'Scheduled contacts' but after adding a 'Scheduled contact' they become visible, at a point when user is done and this information doesn't help anymore (unless user adds more than one contact).	3	Keep 'required fields' displayed at all times in interface to inform user	#6
85	Subjects submenu	Inconsistent location of icons: sometime on the right and other times on the left (e.g. Scheduled contacts, etc.).	3	Keep icons on the right side of the data field consistently in interface	#4
86	Scheduled contacts	Inconsistent use of messages in interface. In many cases there's only a red X in the individual rows and without instructions. In 'Scheduled contacts', there are instructions at the top about deletion or marking contacts as 'not done'.	3	Consider removing instructions and putting information to the tooltips of 'delete' and 'not done' icons	#4
87	Scheduled	Not easy to navigate through headings	3	Use standard submenu layout	

	contacts	that go multiple levels down (e.g. 'Scheduled contacts')		throughout interface. Instead of headings that are all over, change metaphor to use standard tabs next to each-other at the top, open one at a time	#2
88	Scheduled contacts	Instead of using different modes of delete/edit/view in different areas of interface, consider using a standard approach and layout (e.g. 'Scheduled contacts')	3	Consider using standard buttons to provide 'Edit' functionality	#4
89	Subject list	Can't add a new subject. Project Coordinator notes this is because all subjects are coming from existing studies	2	Entering data for a new patient is a basic coaching task. If you know that in the future subjects may not come only from existing studies, consider adding this functionality.	#8
90	Life events	'Move/copy' icon doesn't appear standard	2	This looks like a "Redo" icon. Reconsider, even use widely accepted 'Copy' icon and add tooltip 'Move/copy.	#2
91	Personal information	Data fields are not editable. Can't edit subject's "Personal information". Project Coordinator notes since all participants are managed through the main Living Laboratory study, their Living Laboratory RA is the one that edits personal information. The Coach view only provides viewing of the personal information.	2	If you know that in the future subjects may not come only from existing studies, consider adding this functionality.	#3
92	Application	There is a log in but no log out (exit)	3	Consider logout	#3
93	Scheduled contacts	Though each word in multi-word headings starts with a capital letter, across the interface, in 'Scheduled contacts', 'remove the current date.' That is treated as an actual sentence and has a period at the end, starts with lowercase.	2	A sentence should always start with uppercase	#4
94	Issue filter	In issue filter, 'Status' tab has a large box but only 5 choices in it	1	Fit frame to content	#8
95	Targets	User can enter a date in the past when setting a patient coaching target schedule.	4	Application should warn user this is an invalid action in order to help the user avoid errors, past dates should be disabled.	#6
96	Preferences	Labels are uppercase across the interface. In 'Preferences', choices in list are in lowercase, e.g. 'default'	2	Use case consistently across the interface	#4
97	Main menu	Some menu headings are aligned left (issues, subjects) and others are centered (notices, preferences)	2	Align menu headings in a consistent way	#4
98	Messaging	Confusing workflow when providing recommendations in 'Messaging'. Project Coordinator adds the recommendations for games in the game feedback but activity recommendations in the recommendations section.	3	Be consistent, may use one of the following options: (a) Remove 'Recommendations', provide activity feedback in the 'Activity feedback' section, provide game feedback in the 'Game feedback' section, or (b) Put all recommendations (game and activity) in the 'Recommendations' section, or © Break 'Recommendations' into "Game recommendations' and 'Activity recommendations' sections	#4
99	Messaging	Cursor remains unchanged upon mouse-	3	Cursor should change from arrow to	#1

		over the green icon, so user doesn't know an action could take place by clicking the icon		hand when moving over the green icon, to show action will take place upon clicking icon	
100	Messaging	Tooltip on default green icon is not clear; 'Reload' does not tell user that a default message will be generated.	3	Tooltip needs to help user by explaining the intended action of the metaphor and provide clear instructions. A simple tooltip that reads 'Click to create new default message' would be sufficient.	#8
101	Contacts	No way for user to know during data entry where the required fields are. They only become visible after user saves a contact (which would only help if user plans to immediately add another contact)	3	Keep 'required fields' displayed at all times in interface	#6
102	Contacts	There is s field without a label; if user tries to save contact, this field becomes required	2	A label or a tooltip 'Enter contact here' would help	#7
103	Contacts	Grammatical problem with 'Other (see Info), the two word in brackets are one lowercase and the other uppercase	1	Use lowercase and uppercase consistently	#4
104	Contacts	Not clear what 'Other (see Info)' means and where would I see info	2	Clarify label	#8
105	Contacts	Headings are in regular font while choices inside them are bold	2	Headings should be bold; choices within them: regular font (not bold)	# 7
106	Contacts	No need to use colon (punctuation mark) in headings	2	Headings should be bold; choices within them: regular font (not bold), remove colon	#7
107	Contacts	Space needed to separate adjacent words, e.g. Request(Coaching), Update(Coaching), Sensor outage(all)	1	Use space to separate words (recommend proofreading the entire application)	#7
108	Contacts	Time fields are blank and do not suggest the next action to user, it's not clear how time should be recorded and in what format	3	Use standard time format across the interface, facilitate data entry by using arrow up/down functions	# 9
109	Issue list	Issues should have priority field (that can also be filtered)	2	Consider adding priority field	#8
110	Issue list	Color codes are used for issues of different status but no legend is provided for colors	3	Use color legend to inform user	#7

APPENDIX C: Usability Testing Scenario Sample

Scenario #1

Scenario applies only to subject [Patient Name]:

(1) You had a 10-minute Skype call with subject [patient name] today at [time], regarding subject's inquiry to refer a friend to the Health Coaching study. Please document this interaction.

(2) During the call, the subject also reports having video problems recently when using Skype. Document the reported video problem so that [staff name] can follow up.

Scenario #2

Scenario applies only to subject [Patient Name]:

Please review the messages sent to this subject on [date] and the contacts with this subject on [date]. Note any recommendations given to the patient on [date] to run or jog. Then, send a message to this subject and instruct him to double the number of minutes he was recommended to walk or jog (e.g. if message or contact from [date] reads "Walk for 30 minutes", you'd instruct the subject to "Walk for 60 minutes"). Do this for both walking and jogging. Subject prefers to be called [preferred name].

Scenario #3

Scenario applies only to subject [Patient Name]:

(1) Review subject's performance against goals for the week of [date] through [date], focusing on the games, physical exercise, and social activities. If the subject has not met the goals in any of these areas, check for any emergencies or unusual events in subject's

life. If such event is present, send a brief message to the subject simply to ask how things are going following the event.

(2) Review subject's performance against goals for the week of [date] through [date], focusing on the games, novelty mental exercise, and social activities. View any graphics that are available for subject's performance during the week and interpret them while talking aloud.

(3) Set new goals for the game targets for the week of [date] through [date] as it follows: 3 FreeCell games, 5 Solitaire games, 10 Sudoku games.

Scenario #4

Scenario applies only to subject [Patient Name]:

(1) Please review the issue for [date] for this subject. Note the "Patient Id" in subject's information and add it as a comment in this existing issue.

(2) Change the active topic for this subject to "Physical exercises"

(3) Check for any emergencies or unusual events in subject's life. Delete all events where the description is blank.

(4) You just learned that the subject had a death in the family (you don't know who died). Please document this tragic event so it is present in subject's record.

(5) If this subject has an [event name] event, please copy this event from this subject to subject [patient name].

APPENDIX D: Description of Problems Found by Usability Testing

ID	Scenario	Location	Problem	Severity	Recommendation
1	1.1	Contacts	Word "Contact" not well chosen because it's confusing, contact usually used for contact information. This is used here to document an interaction and some users consistently had trouble finding where to do so in the interface.	2	Consider the word "Encounter" to document interaction with patients
2	1.1	Contacts	Having both start/end time and duration for contacts may be redundant	2	Consider removing one of these fields, possibly 'Duration' since it's deducted by start/end time
3	1.1	Contacts	Contact duration choices are limited (e.g. for a '10 minute' duration, participant selected '<= 5 min' because the next available choice is '15 min', or participant selected 'Other – see info' and had to type in the time in the description field)	2	In addition to providing a predetermined list of choices, enable user to type in a desired duration or may remove "Duration" altogether
4	1.1	Contacts	Users consistently exposed problems with the contact "Reason", anything from making an erroneous selection ("Recruitment call: follow-up to referral" instead of "Responding to query: interested in participating, friend/other referral") to making no selection at all but typing it in description (instead of selecting choice from list)	3	Revisit "Reason for contact" choices
5	1.1	Contacts	"Date" field is a standard structured date format while start/end time fields are text fields	2	Need to use standard structured "Time" fields that are easy to use
6	1.1	Contacts	Reason for contact has a list of choices that have obscure labels and there is no explanation what each means (e.g. does 'friend/other referral mean the patient is referring a friend to the study or is the patient being referred by a friend?)	3	Revisit, clarify and shorten labels, add tooltips to explain to users what these choices mean
7	1.1	Contacts	User wonders what the "Red X" means and what it does (e.g. does it mean there's an error while saving contact or is X used for deleting a contact?)	3	Needs tooltip (e.g. tooltip that reads 'delete item') to help user by explaining the intended action of the metaphor
8	1.1	Contacts	"Reason" for contact and contact "Description" have workflow ambiguity (e.g. if user selects reason "Responding to query: interested in participating, friend/other referral", then types "Inquiry for friend referral", not clear how these fields complement each other instead of just repeating each other?)	2	Revisit "Reason for contact" choices, provide tooltip and label for Description" field to help and inform user
9	1.2	Subject/ Issues	Newly-created issues are not displayed in "History", user is forced to make errors (e.g. to re-create the issue being mislead to thinking that it didn't work the first time). Users consistently had trouble with this. The workaround is to close "Subject" and re-open it in order for issue "History" to refresh.	4	Fix "Refresh" bug in "Issues"
10	1.2	Subject/ Issues	"Issue type" list of not easy to understand (forcing users to select "Other")	3	Revisit, consider add tooltips to explain to users what these choices mean
11	1.2	Subject/ Issues	Upon creating an issue, user is prompted to create another one instead of getting feedback that user actions were successful	3	Automatically take user to "History" upon creating issue (to see it's successful)
12	1.2	Subject/ Issues	Difficulty locating headings in "Issues" submenu, not immediately clear "History" and "Create Issue" are heading tabs	3	Revisit interface layout to facilitate user navigation
13	1.2	Subject/ Issues	When entering an identical issue, some users select to "Track" it to know when anything	2	May add tooltip to inform user

			occurs while other don't		
14	2	Messaging	User wondering whether to use "Generate" or "Send" to create a message and send it to a patient, lost entire message when erroneously using "Generate"	3	Inform user of the difference between these two buttons, add tooltip at a minimum
15	2	Messaging	Location of "Send" button problematic because it's surrounded by two buttons that would practically clear any typed message (namely "Generate" and "Clear"	2	Consider repositioning these 3 buttons to avoid loss of work
16	2	Messaging	"Contact history" and "Messaging history" have duplicated information, with the latter being a synopsis of what's in "Contact history". Counterintuitive that user has to leave "Messaging" and go to "Contacts" in order to see all details of "Messaging history".	2	Assess the need for such duplicate information. Do not cut off tooltips in "Messaging history"
17	2	Messaging/ Contacts	Some confusion because users expect that double clicking contact or message would open it	3	"Double click" functionality should be standardized in interface
18	2	Messaging/ Targets	Sets activity targets for current weekly schedule for jog/walk then goes to messaging to create a message for the current week. Clicks default button in "Activity feedback" but doesn't see automatically updated activity targets in it as supposed to.	4	Fix "Refresh" bug for "Activity feedback" when activity target is updated
19	2	Messaging	Automated messaging is supposed to autogenerate messages based on prior week's performance but it is not doing so	4	Fix bug to autogenerate messages based on prior week's performance
20	2	Messaging/ Targets	When sending message with activity feedback/ recommendations, there is a workflow discrepancy: some users first set the activity targets for the specific activity and then send message, expecting to see automatically updated activity targets in "Activity feedback". Other users directly sent message without setting activity targets.	2	Reconcile standard workflow
21	2	Messaging	Some confusion because when users want to review message history for a specific date for a subject, some select "Display date" for the said week, then go to message history. While quite logical, this currently has no effect on history displayed because all issues are displayed for a given subject for the current year, regardless of selecting date in "Generate" tab.	2	Users' actions are logical but start/end date in messaging "Generate" tab is unclear; need to clarify this for users
22	3.1	Targets	Scheduled date range in "Targets" is error-prone because it allows users to enter end dates that are more than 7 days away from start date, whereas this is supposed to set a weekly date range only.	3	Consider alternatives that enforce the weekly date range to prevent user errors
23	3.1	Targets	Folder icon in activity target doesn't open when user clicks directly on it	3	Standard functionality would allow user to click on folder to open it
24	3.2	Targets	Graphic is not displaying current play counts though the latter exists in "Game targets", even when users correctly selected the date for the week in question	4	Investigate for graphic bug
25	3.2	Targets	When clicking on a specific play count in "Game targets", Graphic displays a date range across several months instead of just the week corresponding to play count in question	3	For ease of use, display only week corresponding to play count in question
26	3.2	Targets	Unclear how to read the graphics and when the patient played, unable to interpret it. Users consistently had trouble with this.	3	Revisit and consider simplifying graphic, one user suggested using dots instead of lines to

					show exactly where they played
27	3.3	Targets	Newly-set goals not displayed in "Current" column of targets. The workaround is to close "Subject" window and then return to it to see updated current targets	4	Should refresh in real-time
28	3.3	Targets	Some confusion because activity targets have columns for "Progress" and "Play count" while game targets have only "play count" but not "Progress". How to document games that are in progress? Also label "Play count" not accurate for activities like sleep or social.	2	Make targets consistent, review "Play count" label in activity targets
29	3.3	Targets	When asked to look at subject performance against goals, user confused "Add entry" start/end date with "Target schedule" start/end date.	2	Layout at top in "Target schedule" is confusing, need to revisit
30	4.1	Issues	When asked to type a comment inside a specific "Issue", most users consistently attempted to do so by editing the "Description" field (not an editable field). The "Comment" field was largely ignored due to its obscure location. Attempted to edit "Description" by double-clicking or right-clicking "Description" field, or via "Double click to view" or "Update.	4	See proposed solution in heuristic evaluation
31	4.1	Issues	Issue "History" has "Issues" window at the top and "Change history" at the bottom. Items at top are editable via double-click while those at bottom aren't, confusing users who expect consistent behavior try to open them by double-clicking. E.g. when asked to type a comment inside a specific "Issue", attempted to do so by selecting the specified issue and then clicking in the "Comment" column in issue's "Change history" without success	2	Standardization discussed in heuristic evaluation
32	4.1	Issues	Double clicking to open issue either didn't take at times or it took quite long for issue to open, user thought it wasn't working due to delayed response time	2	Check response time
33	4.2	Subjects/ Info	Some users couldn't locate "Active topic" because it's not obvious where this lives.	2	See button/ heading layout suggestions in heuristic evaluation
34	4.3	Life events	User states expectation to be able to double-click a "Life event" to see details	2	Need standard approach for editing throughout the interface
35	4.3	Life events	Upon creating a life event, user is prompted to create another one instead of getting feedback that user actions were successful	3	Automatically take user to "Event history" upon creating a life event (to see it's successful)
36	4.3	Life events	Having a start date and end date for life events appears illogical because most evens have only one origin date (e.g. accident, death, etc.). Some users didn't record it because thought start/end was inappropriate.	2	Recommend only one date for events

REFERENCES

1. Palmer S, Tubbs I, and Whybrow A. Health coaching to facilitate the promotion of healthy behaviors and achievement of health-related goals. International Journal of Health Promotion & Evaluation. 2003; 41(3); 91-93.

2. Butterworth SW, Linden A, McClay W. Health coaching as an intervention in health management programs. Disease Management & Health Outcomes. 2007; 15(5); 299-307.

3. Horrigan J. Older Home Broadband Adoption 2009. Pew Internet and American Life Project. 2009; Available from: http://pewresearch.org/pubs/1254/home-broadband-adoption-2009

4. Madden M. Internet penetration and impact. Pew Internet & American Life Project. 2006; Available from: www.pewinternet.org

5. Gatto SL, Tak SH. Computer, Internet, and e-mail use among older adults: Benefits and barriers. Educational Gerontology: An International Journal. 2008; 34(4); 800-811.

6. Jimison H and Pavel M. Integrating computer-based health coaching into elder home care, Technology and Aging, eds. Mihaildis, A., Boger, J., Kautz, H., and Normie, L., IOS Press, Amsterdam, The Netherlands, 2008.

7. Ball K, Berch D, Helmers K, et al. Effects of cognitive training interventions with older adults: A randomized controlled trial. Journal of American Medical Association. 2002; 288(18); 2271-2281.

8. Spool J. The $300 Million Button. User Interface Engineering. [Published: Jan 14, 2009]. Available from: http://www.uie.com/articles/three_hund_million_button.

9. Avouris NM. An introduction to software usability. Proceedings of 8th Panhellenic Conference on Informatics. 2001; 514-522.

10. Tullis T, Fleischman S, McNulty M, Cianchette C, & Bergel M. An Empirical Comparison of Lab and Remote Usability Testing of Web Sites. Presented at the Usability Professionals Association Conference. 2002; Orlando, FL.

11. Brush, A, Ames, M, Davis, J. A Comparison of Synchronous Remote and Local Usability Studies for an Expert Interface. Proceedings of CHI 2004 Conference on Human Factors in Computing Systems. 2004; 1179-1182.

12. Andreasen MS, Nielsen HV, Schroder SO, Stage J. What happened to remote usability testing? An empirical study of three methods. Proceedings of CHI 2007 Conference on Human Factors in Computing Systems. 2007; 1405-1414.

13. C. Fidas, C. Katsanos, E. Papachristos, N. Tselios, N. Avouris (2007). Remote usability evaluation methods and tools: A survey. In T. Papatheodorou, D. Christodoulakis and N.

Karanikolas (Eds.) Proceedings of PCI 2007 (pp. 151-163), Patras, Greece, 18-20 May, 2007, ISBN: 978-960-89784-0-9

14. Shapiro, M. Can an Automated Tailored Health Messaging System Be an Effective Tool for the Health Coaching of Older Adults? [Thesis]. Portland (OR): Oregon Health & Science University; 2008.

15. Nielsen, J. "How to Conduct a Heuristic Evaluation". Available from: http://www.useit.com/papers/heuristic/heuristic_evaluation.html

16. International Organization for Standardization. ISO 9241-11: Guidance on Usability. [June 2009]. Available from: http://www.usabilitynet.org/tools/r_international.htm

17. Rubin J. Handbook of usability testing. New York: John Wiley & Sons. 1994.

18. Joshi A, Arora M, Dai L, Price K, Vizer L, Sears A. Usability of a Patient Education and Motivation Tool Using Heuristic Evaluation. Journal for Medical Internet Resesearch. 2009; 11(4):e47.

19. Thyvalikakath TP, Monaco V, Thambuganipalle H, Schleyer T. Comparative study of heuristic evaluation and usability testing methods. Studies in health technology and informatics. 2009;143; 322–327.

20. Straub K. Cleaning up for the housekeeper or why it makes sense to do both Expert Review and Usability Testing. The UPA Voice. [Published November 2004]. Available from: http://www.upassoc.org/upa_publications/upa_voice/volumes/2004/nov/cleaning_up.html.

21. Fu L, Salvendy G, Turley L. Effectiveness of user testing and heuristic evaluation as a function of performance classification. Behaviour & Information Technology. 2002; 21(3); 137–46.

22. Nielsen J. Why you only need to test with 5 users. Alertbox. [Posted on March 19]. http://www.useit.com/alertbox/20000319.html.

23. Nielsen, J. Guerrilla HCI: Using Discount Usability Engineering to Penetrate the Intimidation Barrier. 1994. Available from: http://www.useit.com/papers/guerrilla_hci.html.