Tec-Ed
Connecting People and Technology

Web Tool for Health Insurance Design by Small Groups: Usability Study

Laurie Kantner
Tec-Ed, Inc.
4300 Varsity Drive, Suite A
Ann Arbor, MI 48103 USA
laurie@teced.com

Susan Dorr Goold, MD, MHSA, MA
Associate Director for Ethics and Health Policy
University of Michigan Medical School
Ann Arbor, MI 48109-0429
sgoold@med.umich.edu

Marion Danis, MD
Department of Clinical Bioethics
National Institutes of Health
Bethesda, MD 20892-1156
mdanis@nih.gov

Mike Nowak
Center for Health Communication Research
University of Michigan
Ann Arbor, MI 48109
mnowak@umich.edu

Lesa Monroe-Gatrell
Center for Health Communication Research
University of Michigan
Ann Arbor, MI 48109
eponymous@umich.edu

CHI 2006, April 22–27, 2006, Montréal, Québec, Canada.
ACM 1-59593-298-4/06/0004.

Abstract
This Experience Report describes the challenges of evaluating the usability of a Web-based collaborative health insurance benefits planning application. The application was created by researchers at the National Institutes of Health and the University of Michigan.

Keywords
computer mediation, usability evaluation, longitudinal studies, collaborative applications, CSCW

ACM Classification Keywords
Categories and subject descriptors: J.3 Life and Medical Sciences—Health; H.5.3 [Information Interfaces and Presentation]: Group and Organization Interfaces—Evaluation/methodology, Web-based interaction

General terms: human factors

Additional key words and phrases: group health insurance benefits decision-making

Introduction
A software application originally designed for in-person group decision-making was redesigned as a Web application to broaden its use in research and practical applications. Funded by the National Institutes of

Health, CHAT (Choosing HealthPlans All Together) was created to address the problems of conflicting priorities and limited resources in designing health insurance. The application presents choices in a graphical format, providing the "big picture" of value and tradeoffs [2, 3, 4]. CHAT is copyrighted by the Board of Regents of the University of Michigan.

The original CHAT design relied on a human-facilitated setting. In that context, groups of about 9 to 12 people meet in the same room, using laptops to read about benefits and make choices. Descriptions are written for comprehension at the sixth grade reading level. The group facilitator ensures that all participants have a fair opportunity to express their health care priorities as input to the decision-making process.

As designed for the Web, CHAT retained similar features, although modified somewhat for Web presentation:

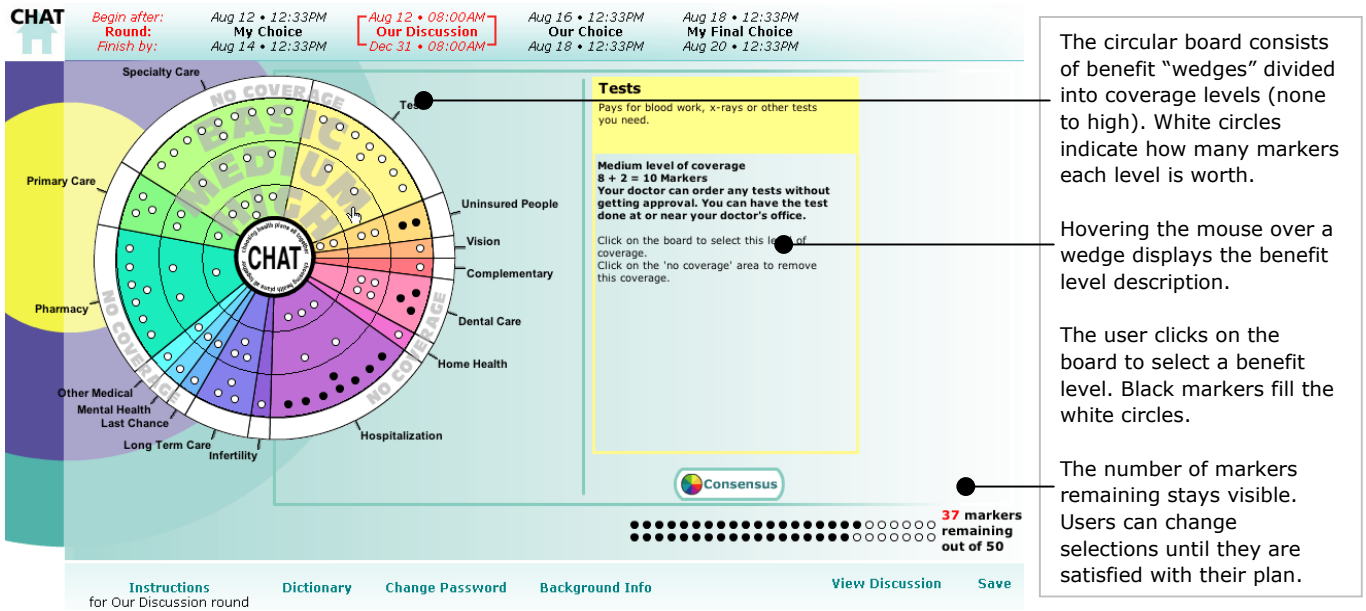
- CHAT presents a circular board divided into as many as 17 benefit areas (see Figure 1).
- CHAT participants are given 50 or 100 "markers" to spend, which allows them to choose a subset of the benefit options.
- Participants must read about the levels of coverage available and weigh various tradeoffs, such as more or less convenience, flexibility of services, and cost sharing, as they consider benefit options.
- Participants view randomly presented "health events" depicting potential illnesses and accidents that could befall them. These events explain the consequences of participants' coverage decisions.
- Instead of reaching consensus through a process of human-facilitated negotiation and compromise, participants are encouraged through the user interface to visit a "discussion board" space and share their thoughts.

The CHAT creators wanted to ensure that the Web version of CHAT did not pose barriers to group discussion and to learning about benefit tradeoffs. They engaged Tec-Ed to conduct usability testing of the Web version of CHAT. The findings from this usability test confirmed that improvements were needed, and two cycles of software changes and usability testing followed.

Application Structure

CHAT guides individuals through the process of selecting health benefit options for themselves and then, working with a group of 10 to 12 people, creating a health benefit plan for a larger group. The experience consists of four "rounds" of activity:

- **Round 1**: Create an individual health plan.
- **Round 2**: View consensus of individual health plans, and nominate changes for the group plan, giving reasons and comments.
- **Round 3**: View consensus based on previous round, and finalize a group plan.
- **Round 4**: Create an individual health plan again.



The circular board consists of benefit "wedges" divided into coverage levels (none to high). White circles indicate how many markers each level is worth.

Hovering the mouse over a wedge displays the benefit level description.

The user clicks on the board to select a benefit level. Black markers fill the white circles.

The number of markers remaining stays visible. Users can change selections until they are satisfied with their plan.

Figure 1. CHAT board

The Web version of CHAT required reprogramming the application using a combination of Shockwave components and HTML pages. This platform simplified interaction with the CHAT board, but introduced challenges in integrating instructional information.

Challenges in Designing the Usability Test

Usability testing of CHAT required evaluation of not only individual success but also group interaction. Groups needed to have at least nine members for meaningful results from one round to the next. The risk of attrition between rounds required consideration of how many people to recruit and the incentive for participation.

Usability test participants needed to represent diversity in age, gender, race, employment, income level, education, and computer and Web experience. Persons with the lowest computer skill level were of greatest concern to the CHAT creators. However, persons with less than a year of computer experience or who were new to the Web were excluded.

All of these challenges suggested designing a very large usability study to observe multiple groups and account for attrition. However, we also wanted to complete the study within a reasonable timeframe and budget.

To balance data collection needs with budget and schedule, we designed a study with the following parameters:

- 27 participants were recruited, diverse in age, gender, employment, income, and education.
- Participants were divided into 3 groups of 9 people each. Groups were balanced for age, gender, and income.
- To provide incentive for completing all rounds, the honoraria were given in two installments, a smaller one after Round 1 and a larger one after Round 4.
- We were selective about which sessions required in-person observation in the usability lab: all Round 1 sessions, half of the Round 2 and 3 sessions (making sure to observe each participant once), and none of the Round 4 sessions, which were identical to Round 1.
- For unobserved sessions, participants worked independently and responded to follow-up questions either at the beginning of their next round or by email.
- The user researcher observing the sessions also fulfilled the role of "CHAT administrator," sending invitation and reminder emails when rounds were beginning or about to end, and providing "tech support" by phone if participants had difficulty accessing CHAT from home.
- The CHAT exercise took place within a one-month period. Without the need for observation, the normal CHAT experience can be completed in less time.

Methodology for Group Process

For this iterative testing of a group decision-making application to collect useful data, we needed to make some methodology adaptations.

Balancing Data Collection with Session Continuation

In usability testing of CHAT, we needed to learn where participants were having difficulty, but we also needed them to continue meaningful participation from one round to the next. Compared to a usability test where a session may end in failure, in this study it was more important to note the failure but then provide remediation so that the participant could continue functioning as a group member.

Problems with Lack of Noticeability

Important usability metrics for CHAT were whether participants understood the tradeoffs they were making and whether participants made the selections they intended. In-person observation and use of think-aloud protocol were critical for these assessments. Because onscreen reading determined participants' level of understanding, we observed how much they read, skimmed, skipped, and missed.

In the initial CHAT design, several design elements caused participants to make unintended selections. Participants needed to click on a label to read about a benefit selection, but the label did not look clickable. Instead, participants clicked on the circular board itself, which allowed reading but also selected the benefit.

In some cases, participants noticed these selections on their own, or noticed with a mild hint. In other cases, participants did not recognize that display of black markers meant a benefit was selected. After confirming

the participant was "stuck," the user researcher started hinting how to use CHAT. Hinting and remediation ensured that each user confirmed their selections matched their decisions. The researcher noted whenever a participant required remediation and what information had been lacking.

Thus, careful tracking of steps taken, information read, "wrong turns" made, administrator hinting, and instructional explanations to the participant were key components of the data collection. The hints and instructional explanations became input to designing better instruction for CHAT in the next version.

Selection Problems Masked Group Discussion Problems
Two important aspects of the group decision-making process in CHAT were the task of "nominating" benefits in Round 2 and use of the discussion area in all rounds. The discussion area was the forum for nominating and negotiating selections for the group plan. Participants needed to engage comfortably in computer-mediated discussion with people they did not know.

In the first cycle of testing, a usability problem prevented participants from seeing a benefit description once they began the nomination process. This usability problem, and the problem with unintentional benefit selection, masked problems with the discussion process itself. The discussion area problems did not emerge fully until the third cycle of usability testing.

Unobserved Sessions: Computer Skill Issues
Originally the plan was to select participants for observation in Round 2 based on even distribution of key characteristics. However, participants with weaker computer skills were at risk of dropping out of the study if chosen to complete Round 2 on their own. We chose to observe those participants in Round 2, and the participants with stronger computer skills completed Round 2 on their own and were observed in Round 3.

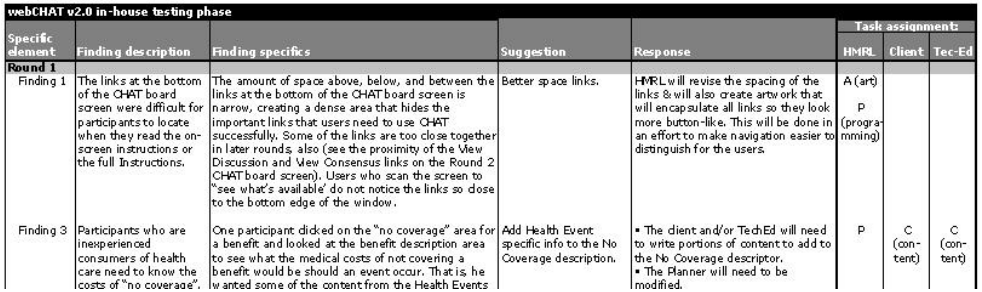
Post-round questionnaires for all participants provided additional detail, and analysis of logfiles containing Round 2 nominations, Round 3 rankings, and discussion comments provided additional insight. Selections inconsistent with participants' stated opinions (collected during observations and found in discussion comments) indicated potential usability problems.

For the unobserved Round 4 activity, we analyzed benefit selections, discussion comments, and final-questionnaire data to assess difficulty with that round. We also followed up with some participants by email and phone.

Iterative Changes and Testing

The user researcher's report prioritized the problem findings with CHAT, and the development group organized the findings onto a spreadsheet with implementation suggestions (see Figure 2). Stakeholders met to discuss and agree on priorities and implementation. Stakeholders also supported conducting another cycle of usability testing. The second round found additional issues, and we repeated the process of addressing high-priority issues and testing one more time.

The user researcher revised the CHAT online help for the second cycle of usability testing, while the CHAT developers added onscreen instructions. For the third cycle of testing, the user researcher revised the onscreen instructions based on usability feedback.



webCHAT v2.0 in-house testing phase							
Specific element	Finding description	Finding specifics	Suggestion	Response	Task assignment: HMRL	Client	Tec-Ed
Round 1							
Finding 1	The links at the bottom of the CHAT board screen were difficult for participants to locate when they read the on-screen instructions or the full Instructions.	The amount of space above, below, and between the links at the bottom of the CHAT board screen is narrow, creating a dense area that hides the important links that users need to use CHAT successfully. Some of the links are too close together in later rounds, also (see the proximity of the View Discussion and View Consensus links on the Round 2 CHAT board screen). Users who scan the screen to "see what's available' do not notice the links so close to the bottom edge of the window.	Better space links.	HMRL will revise the spacing of the links & will also create artwork that will encapsulate all links so they look more button-like. This will be done in an effort to make navigation easier to distinguish for the users.	A (art) P (programming)		
Finding 3	Participants who are inexperienced consumers of health care need to know the costs of "no coverage".	One participant clicked on the "no coverage" area for a benefit and looked at the benefit description area to see what the medical costs of not covering a benefit would be should an event occur. That is, he wanted some of the content from the Health Events	Add Health Event specific info to the No Coverage description.	▪ The client and/or TechEd will need to write portions of content to add to the No Coverage descriptor. ▪ The Planner will need to be modified.	P	C (content)	C (content)

From left to right, the spreadsheet refers to the finding number in the usability report, the finding statement and description, and the user researcher's suggestion, and then provides the development team's response and suggests assignments for implementation.

Figure 2. Usability results action spreadsheet

The second and third cycles of usability testing used the same protocol but were conducted on a smaller scale. Six people were recruited for each cycle. To create groups of 9 to 15 people, we added stakeholders and the user researcher to the group. The user researcher observed all participant sessions for Rounds 1, 2, and 3. Participants completed Round 4 independently.

Each cycle of usability testing validated improvements made, confirmed improvements still needed that had been deferred, and identified new issues. By the third cycle, with earlier issues resolved, the discussion area emerged as an area requiring design improvements to facilitate the group decision-making process and to improve social presence. Improvements included inserting benefit level context in discussion comments, simplifying discussion threads, and adding a place for a customized group description. Discussion using other media remains a future area for study to evaluate the effect on social presence and establishment of trust [1]. The final, improved version of the Web-based version of CHAT is now ready for use by employers, researchers, insurance companies, community-based and health care organizations, and policy makers.

References

[1] Bos, N., Olson, J., Gergle, D., Olson, G., Wright, Z. Effects of Four Computer-Mediated Communications Channels on Trust Development. Collaboratory for Research on Electronic Work (CREW), University of Michigan. http://www.crew.umich.edu/publications/02-02.pdf

[2] Danis, M., Benavides, E., Nowak, M., Goold, S.D. Development and Evaluation of a Computer Decision Exercise for Consumer Participation in Insurance Benefit Planning. *Forum for Family and Consumer Issues 10* 2 (2005). http://www.ces.ncsu.edu/depts/fcs/pub/10_2/

[3] Goold, S.D., Biddle, S.K. Klipp. G., Hall, C.N., Danis, M. Choosing Healthplans All Together: A deliberative exercise for allocating limited health care resources. *Journal of Health Politics, Policy, and Law 30* 4 (2005), 563-601.

[4] Sacramento Health Decisions, 2003. When Options Exceed Resources. http://www.sachealthdecisions.org/CHAT%20report.pdf

The opinions stated here are those of the authors and no reflection of the policies of the NIH or the US Department of Health and Human Services.