

Original Research

Heuristic Evaluation and Usability Testing of a Computerized Patient-Reported Outcomes Survey for Headache Sufferers

Diane M. Turner-Bowker, Ph.D.,¹ Renee N. Saris-Baglana, Ph.D.,¹ Kevin J. Smith, Ph.D.,² Michael A. DeRosa, M.A.,¹ Christine A. Paulsen, Ph.D.,³ and Sarah J. Hogue, B.A.¹

¹Datacorp, Smithfield, Rhode Island.

²j3 QualityMetric Incorporated, Lincoln, Rhode Island.

³Concord Evaluation Group, Concord, Massachusetts.

Abstract

Objective: The aim of this study was to evaluate usability of a prototype tablet PC-administered computerized adaptive test (CAT) of headache impact and patient feedback report, referred to as HEADACHE-CAT. **Materials and Methods:** Heuristic evaluation specialists (n = 2) formed a consensus opinion on the application's strengths and areas for improvement based on general usability principles and human factors research. Usability testing involved structured interviews with headache sufferers (n = 9) to assess how they interacted with and navigated through the application, and to gather input on the survey and report interface, content, visual design, navigation, instructions, and user preferences. **Results:** Specialists identified the need for improved instructions and text formatting, increased font size, page setup that avoids scrolling, and simplified presentation of feedback reports. Participants found the tool useful, and indicated a willingness to complete it again and recommend it to their healthcare provider. However, some had difficulty using the onscreen keyboard and autoadvance option; understanding the difference between generic and headache-specific questions; and interpreting score reports. **Conclusions:** Heuristic evaluation and user testing can help identify usability problems in the early stages of application development, and improve the construct validity of electronic assessments such as the HEADACHE-CAT. An improved computerized HEADACHE-CAT measure can offer headache sufferers an efficient tool to increase patient self-awareness, monitor headaches over time, aid patient-provider communications, and improve quality of life.

Key words: e-health, technology, telehealth, headache, health-related quality of life, patient-reported outcomes

Introduction

Global lifetime prevalence of the two most common headache disorders has been estimated at 14% for migraine and 46% for tension-type headaches.¹ Migraine ranks in the top 20 of the world's most disabling medical illnesses, based on the World Health Organization's Global Burden of Disease

survey.² Headache sufferers often report missed or reduced work and/or household productivity and interruptions in family and social activities.³⁻⁵ Still, headache is often underdiagnosed and undertreated,^{6,7} patients often do not seek medical treatment,⁷ healthcare providers may be unaware of how headaches affect patients' lives,⁸ and patient-provider communication could be improved for more effective headache treatment.⁹

Patient-reported outcomes (PRO) measures that assess headache impact¹⁰ may improve the communication gap between patients and care providers and inform patient-centered care.⁹ Patient-centered treatment approaches improve health outcomes¹¹ and information regarding headache-related disability (e.g., reduced work productivity) can improve physicians' judgments of patient migraine severity and treatment management needs.¹²

Technology innovations have the potential to improve use of PRO data, but human factors (e.g., human experience interacting with a particular technology) must be considered during the development of new or improved outcomes assessments.

This research assessed a prototype tablet PC-administered computerized adaptive test (CAT) of headache impact on health-related quality of life (HRQOL) and an associated patient feedback report, collectively referred to as HEADACHE-CAT, through heuristic evaluation and usability testing.

Heuristic evaluation involves the examination of a user interface (UI) by expert evaluators to identify problems and determine whether the UI conforms to established usability principles, or "heuristics." Usability testing, a key component of product evaluation, focuses on measuring a product's (e.g., device, Web application, and document) ability to meet its intended purpose by providing evidence on how real users interact with it. Usability testing involves end users and realistic testing scenarios.¹³ Through systematic observation, errors and areas of improvement can be discovered by observing how well test users perform tasks, the accuracy of product use, what a person remembers after use, and the person's emotional response.

Although heuristic evaluation based on usability specialists' review can efficiently identify usability problems for improved accessibility, alone it may not uncover all potential issues that users may experience when interacting with the system. Usability testing is often used in combination with expert heuristic testing, since it may identify problems that are otherwise impossible to detect by expert evaluators (who emulate users), regardless of their skill and experience. This testing ensures the application of basic design principles to the HEADACHE-CAT, and evaluates the survey and technology platform for interface design, navigational elements, accessibility, and user preferences.

The goal of this study was to evaluate the practical functionality of the HEADACHE-CAT, through heuristic evaluation and usability testing before its implementation in clinical research and practice.

Materials and Methods

HEURISTIC EVALUATION

Two American Institutes for Research (AIR) usability specialists independently reviewed and evaluated the HEADACHE-CAT (see Figs. 1 and 2 for initial UI and patient feedback report). Problems were noted when system characteristics violated constraints known to influence usability, based on general usability principles¹⁴ or criteria derived from and validated by human factors research.¹⁵ AIR experts then compared individual findings and formed a consensus opinion on the application’s strengths and opportunities for improvement across several usability categories (conceptual model, navigation, visual design/layout, content, patient feedback report, and administration mode).

USABILITY TEST

Sample. The sample included adult self-reported headache sufferers from a 3,000-member research panel covering the southern New England area. Members were eligible for study enrollment if they were over the age of 18, able to read and write in English, and reported current headache activity (past 4 weeks). Efforts were made to include an even distribution of participants with mild, moderate, and severe headache. Members who had participated in a focus group or usability test in the last 6 months, as well as those who reported

headaches as a result of a cold, the flu, head injury, or hangover, were excluded.

Materials and equipment. HEADACHE-CAT is an electronic survey that includes generic and headache-specific HRQOL surveys (e.g., SF-12v2TM Health Survey,¹⁶ DYNHA[®] Headache Impact Test [HITTM],^{17,18} or short-form HIT-6^{TM19}), optional modules (e.g., socio-demographics and chronic conditions checklist), and a patient feedback report.

HEADACHE-CAT was presented on a tablet PC. Participants could hold the tablet PC in their lap (stylus and on-screen keyboard used for data entry; manual keyboard not accessible) or place it on the desk (stylus and manual keyboard used for data entry), and choose a portrait or landscape screen orientation, 14-point or 18-point font size, and settings that autoadvance them from one item to the next and allow them to skip items. Default configuration settings were landscape orientation, 14-point font, double-click response (requiring the respondents to select “Next” before proceeding to a subsequent question), and forced response (no skip allowed). Participants were shown options for survey presentation, including background color, reverse print box (light font on dark background) for item numbering, single versus multiple items per page, and horizontal versus vertical layout of response options. Participants were also offered the option to complete a training module that describes how to use the tablet PC stylus and onscreen keyboard.

Procedure. AIR usability specialists conducted one-on-one structured interviews with users at a Concord, MA, laboratory. The specialist and participant were observed

by project staff through a one-way mirror, while closed circuit cameras recorded participant verbal and non-verbal communications. The specialist observed how the participant interacted with the tablet PC, navigated the survey, and responded to the survey questions. Through structured interview, they gathered user feedback on the interface layout and visual design; navigational elements; welcome page, instructions, and tablet PC training information; survey item content and patient report; and user preferences. Each interview session was videotaped and transcribed. The study protocol was approved by AIR’s Institutional Review Board, and participants signed consent forms prior to the study.

Analyses. The specialists independently reviewed transcript data and evaluated participant feedback as it related to the survey interface, navigation,

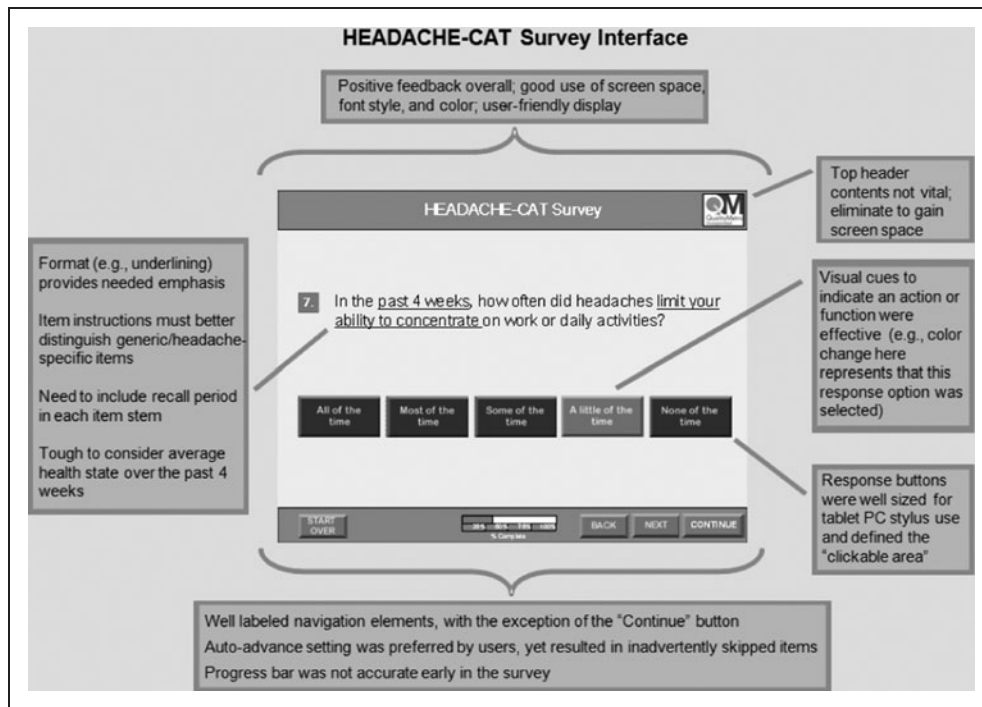


Fig. 1. HEADACHE-CAT survey interface. CAT, computerized adaptive test.

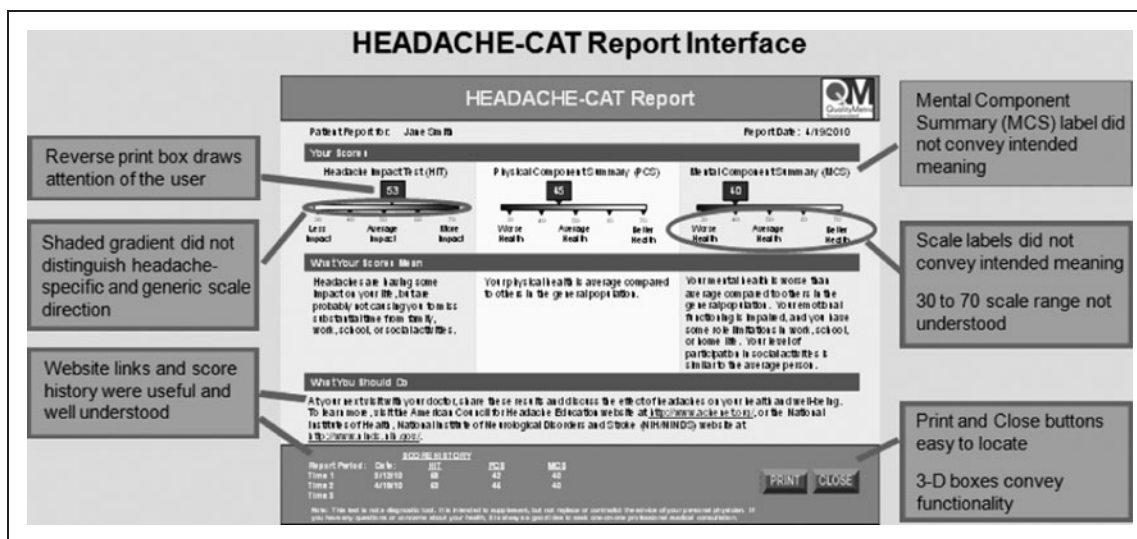


Fig. 2. HEADACHE-CAT feedback report interface.

content, and user preferences. Findings were specified by item (e.g., pros and cons for each item) and then summarized by emerging themes.

Results

HEURISTIC EVALUATION

HEADACHE-CAT was found to have a professional, familiar “look and feel” with customary item formats and fairly straightforward language text. The feedback report was appealing and contained useful information, and the tablet PC was a manageable size and easy to hold. Priority areas for enhancement included improved user orientation and instructions, tablet PC user training, and improved navigational features; increased font size for readability; consistency in instructions, definitions, and acronyms; simplified item and response wording; formatting to emphasize key text; improved page setup to avoid scroll; and simplified score reporting.

USABILITY TEST

The sample ($n = 9$) included six women and three men (age range, 27 to 54 years; mean, 38 years) and headache severity (22% mild, 56% moderate, and 22% severe) (see Table 1).

Overall, HEADACHE-CAT was found to have a user-friendly design. Although most participants reported no experience using tablet PCs, all were readily able to complete the survey.

User preferences. Five participants positioned the tablet PC upright and used the stylus to answer questions and the manual keyboard to type their text responses. Remaining participants ($n = 4$) chose to keep the tablet PC flat on the table or in their lap and used the stylus and onscreen keyboard. Participants did not indicate a preference for screen orientation. Two participants requested a larger font size. The majority chose the single click, autoadvance feature ($n = 6$), and the option to skip questions ($n = 5$). Most preferred the off-white back-

ground ($n = 7$), reverse print item number ($n = 7$), single-item-per-page layout ($n = 6$), and horizontal-response-option layout ($n = 6$). Three participants chose to complete the training module.

Survey. Participants reported that the stylus was “easy” or “extremely easy” to use. However, typing numbers and text into required fields was problematic for those using the onscreen keyboard. Participants who opted out of the training module ($n = 6$) reported that it was difficult to find the writing pad, and, once found, keep it on the screen long enough to use it. The specialist also observed that these participants had difficulty using the onscreen keyboard for response entry.

Participants who completed the training module ($n = 3$) reported that it helped them understand how to use the stylus and onscreen keyboard, but noted that the writing pad did not stay on screen long enough and blocked instructional text.

All participants reported that the survey instructions were clear, both at the beginning and embedded within the questions. However, based on participants’ think aloud responses to the opening instructions, “The first set of questions will ask about your health in general. Then, you will be asked about the impact of headaches on your health,” many missed the distinction between the first (generic) and second (headache-specific) set of questions until later in the survey. Further, throughout the survey, some questioned when to answer an item based on their headache symptoms, overall physical health, emotional health, or a combination of headaches and their overall health.

When shown a question screen, participants indicated that they first noticed the item number and response option buttons. Visual cues to indicate an action or function were effective. Participants recognized that the three-dimensional graphic for response option and navigational buttons conveyed functionality, and that a color change in the response button registered their selection. The button

Table 1. Participant Characteristics

	N
Ethnicity	
Hispanic or Latino	2
Not Hispanic or Latino	7
Race	
American Indian/Alaska Native	0
Asian	1
Black or African American	2
Native Hawaiian/Pacific Islander	0
White	6
Headache severity	
Mild	2
Moderate	5
Severe	2
Education level	
Some high school	1
Graduated high school/GED	1
Some college or technical school	3
College graduate	2
Graduate or professional degree	2
Tablet PC experience	
None	7
Familiar	2

size and shape defined the clickable response area. Color schemes provided sufficient contrast.

Generally, participants could answer survey items, and underlined text helped them to understand important elements within each question. Some participants overlooked item-level instructions, and recommended that those instructions be presented as part of the item itself.

Participants were easily able to determine how to begin the survey again, close the application, return to a previous question, and proceed to a subsequent question. Those who chose the autoadvance feature encountered a few navigation issues. For example, the delay between questions was too long for some, who attempted to click the “Next” button to proceed more quickly and inadvertently skipped some questions. Those who used the default setting (double-click option—first click on response, then click “Next” button to advance) did not have problems progressing through the survey.

When asked to return to a previous question, participants were surprised by a pop-up message containing navigation instructions

prompting them to select the “Continue” button to return to that last question answered. Initially, some respondents thought they had done something wrong or were not allowed to return to previous questions. Even after reading the pop-up message, the majority of participants expressed confusion regarding the “Next” and “Continue” buttons. Five participants did not understand the pop-up information and three misinterpreted the text. In general, pop-up messages containing instructions or prompts tended to be ignored by users. For example, participants dismissed or ignored the “Start Over” and “Exit” pop-up messages without reading them.

Participants reported mixed responses to the progress bar. Whereas some indicated that it was helpful, others ignored it or reported that they found it frustrating. The specialist observed that the progress bar did not seem to be functioning properly during certain survey module administrations, which may have led participants to feel that they were progressing more slowly than they actually were.

Report. The specialist asked participants what type of information they would expect to see in a feedback report based on their responses. Expectations included statistics; an indicator of headache severity; impact of headaches on daily life; feedback on their health; time spent on the survey; a summary of their responses to share with a doctor; and recommendations for how to cope with headaches.

Next, participants were shown a sample report (see Fig. 2) that included two scale scores (physical component summary [PCS] and mental component summary [MCS]) based on the generic SF-12v2 Health Survey, one scale score based on the headache-specific survey (HIT), and interpretive text. When asked what they first noticed, the majority said numerical scores (shown on reverse print box).

Participants were confused by the scales and numerical scores, and none were able to confidently explain their scores without reading the “What Your Scores Mean” section. Participants expected to see a 0–100 scale score (rather than the 30–70 score range shown in the report).

Scale labels were ignored or misinterpreted. Many thought the PCS and MCS scores were specifically related to headache impact rather than a reflection of their overall health. A few participants did not understand what physical health meant in relation to their headaches. Two participants reacted negatively to the MCS score; they interpreted the scores to indicate a need for “psychiatric help.”

Although graphics provided anchors for the score continua (e.g., worse health and better health), a few participants misunderstood the direction of the scale scores. The use of shading on the scales to indicate scoring direction was not effective. Lastly, it was unclear to participants whether the “general population” or “average person” included as a point of reference in a report included only headache sufferers, the entire world, the U.S. population, or was generalized from a sample from another study. Therefore, they were unsure who they were being compared with when normative groups were referenced.

Participants liked the score history section and Web site links, and were able to easily print the report.

Discussion

HEADACHE-CAT was found to have a professional and user-friendly design. Although most participants reported no experience using tablet PCs, all were readily able to complete the survey, indicating that the stylus was easy to use. Overall impressions of the survey and report were positive; generally, participants reported that instructional and item content was clear and easy to follow. On the basis of these findings, many existing features of the HEADACHE-CAT will be maintained; for example, the training module, off-white screen that minimizes glare/contrast for extended use, item font size options, look/feel of response buttons (e.g., reverse print and three-dimensional), and most navigational buttons were sufficient for users. However, this evaluation uncovered some usability issues that hindered participants' optimal use of the HEADACHE-CAT, including, but not limited to, using the onscreen keyboard, navigational difficulties using the autoadvance option and "Continue" button, understanding how to answer general health versus headache-specific questions, and difficulties understanding and interpreting score output provided in feedback reports.

On the basis of these findings, some specific application improvements will be made, including (1) revision of instructions and functionality for the onscreen keyboard features, autoadvance, and the "Continue" navigational button; (2) introductory instructions to better distinguish the generic and headache-specific survey modules; (3) elimination of instructional pop-up messages and inclusion of more detailed survey and item-level instructions; and (4) simplified representation and improved labeling of scores and scales on the feedback report.

Some of the difficulties in using the onscreen keyboard, most of which cannot be modified as they are part of the computer system itself (e.g., size, appearance only when user "hovers" over the keyboard), might require that users have a manual keyboard option available for response entry. Other difficulties may be resolved through a required training. In this study, many participants may have opted out of the training module that provided instruction for use of the tablet PC stylus and onscreen keyboard because they had computer experience in general. Although many participants elected to use autoadvance navigation, the moderator observed that the double-click navigation was actually more efficient, allowing users who selected this option to proceed at their own pace. In addition, the double-click navigation eliminates the potential for skipping questions. If both navigational strategies are offered, consistent and adequate delay time is required so that participants will recognize that the change from one item to the next is taking place.

Although participants navigated easily through the application, the "Continue" button, employed when a user attempted to go back to change a response and then continue by returning to the last item answered, was problematic. We will modify the label on this button function to "Return to Item #XX," to indicate that the button will bring them forward to the last item that they answered before going back to a previous item in the survey.

The generic and headache-specific modules will be better distinguished by providing explicit instructions illustrated with a sample question on individual pages that precede each question set. For

example, instructions for the generic items indicate that the first set of questions refer to one's general health, including, but not exclusive to, headache.

Generally, information contained in the feedback report was not easily understood by participants. Although efforts will be made to improve scale labeling, patients may need additional support from a healthcare provider to interpret results, in part because component scores are complex for the average person (norm-based scores representing a composite score across health domains). To alleviate the negative connotation a few users associated with "MCS," we will consider an alternative feedback report score label. Results also suggested that users will require additional information to adequately interpret the scores produced and displayed in the report. For example, the user's score could be presented in relation to a table showing possible score levels and associated interpretative text. Participants liked the score history section and Web site links, and were able to easily print the report.

Despite some usability issues, all participants responded favorably to the application, noting that it was a useful tool that they would be willing to complete again and recommend to their healthcare provider. They commented that the survey helped them to increase their awareness of their symptoms and how their headaches are impacting their life, reminded them that others experience similar pain, and would help them to describe specific experiences to their healthcare provider.

A limitation of usability testing is that it is based on a small group of end users, and additional product refinements may be needed. Efforts were made in this study to recruit a representative sample of headache sufferers to address this limitation. Qualitative research can inform the development of PRO measures²⁰ and user-friendly technology platforms.²¹ It is imperative to conduct usability testing before implementation of a new technology in broader groups of patients, as healthcare providers and researchers alike require valid data and the assurance that patient responses are not influenced by usability issues. Additionally, testing the application before clinical field testing reduces the likelihood that staff will be needed to answer usability-related questions in busy clinical settings where there are time and resource constraints. Incorporating input from patients in the early stages of application development should improve the construct validity of this PRO measure and enhance its practical application in healthcare.

Acknowledgments

This study's data were collected while Dr. Turner-Bowker, Dr. Saris-Baglana, and Mr. Michael DeRosa were employed at i3 QualityMetric Incorporated and Dr. Paulsen was employed by AIR. This work was supported in part by a National Institutes of Health-sponsored grant through the National Institute of Neurological Disorders and Stroke (Grant Number #5R44NS047763-03). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Neurological Disorders and Stroke or the National Institutes of Health.

Disclosure Statement

Kevin J. Smith, Ph.D., is employed by i3 QualityMetric Incorporated, the company that provides the HEADACHE-CAT survey. The remaining authors have no competing financial interests.

REFERENCES

1. Stovner LJ, Hagen K, Jensen R, Katsarava Z, Lipton R, Scher A, Steiner T, Zwart JA. The global burden of headache: A documentation of prevalence and disability worldwide. *Cephalalgia* **2007**;27:193-210.
2. Leonardi M, Steiner TJ, Scher AT, Lipton RB. The global burden of migraine: Measuring disability in headache disorders with WHO's Classification of Functioning, Disability, and Health (ICF). *J Headache Pain* **2005**;6:429-440.
3. Brandes JL. Migraine and functional impairment. *CNS Drugs* **2009**;23:1039-1045.
4. Buse DC, Rupnow MFT, Lipton RB. Assessing and managing all aspects of migraine: Migraine attacks, migraine-related functional impairment, common co-morbidities, and quality of life. *Mayo Clinic Proc* **2009**;84:422-435.
5. Lipton RB, Stewart WF, Diamond S, Diamond ML, Reed M. Prevalence and burden of migraine in the United States: Data from the American Migraine Study II. *Headache* **2001**;41:646-657.
6. De Diego EV, Lantieri-Minet M. Recognition and management of migraine in primary care: Influence of functional impact measured by the headache impact test (HIT). *Cephalalgia* **2005**;25:184-190.
7. Lipton RB, Stewart WF, Simon D. Medical consultation for migraine: Results from the American Migraine Study. *Headache* **1998**;38:91-96.
8. Lipton RB, Hahn SR, Cady RK, Brandes JL, Simons SE, Bain PA, Nelson MR. In office discussions of migraine: Results from the American Migraine Communication Study. *J Gen Intern Med* **2008**;23:1145-1151.
9. Buse DC, Lipton RB. Facilitating communication with patients for improved migraine outcomes. *Curr Pain Headache Rep* **2008**;2:230-236.
10. Brandes JL. The migraine cycle: Patient burden of migraine during and between migraine attacks. *Headache* **2008**;48:430-441.
11. Stewart M, Brown JB, Donner A, McWhinney IR, Oates J, Weston WW, Jordan J. The impact of patient-centered care on outcomes. *J Fam Pract* **2000**;49:796-804.
12. Holmes WF, MacGregor EA, Sawyer JP, Lipton RB. Information about migraine disability influences physician's perceptions of illness severity and treatment needs. *Headache* **2001**;41:343-350.
13. Nielsen J. *Usability engineering*. San Diego, CA: Academic Press Inc., **1993**.
14. Molich R, Nielsen J. Improving a human-computer dialogue. *Commun ACM* **1990**;33:338-348.
15. Gerhardt-Powals J. Cognitive engineering principles for enhancing human-computer performance. *Int J Hum Comput Interact* **1996**;8:189-211.
16. Ware JE, Kosinski M, Turner-Bowker DM, Gandek B. *How to score version 2 of the SF-12[®] health survey*. Lincoln, RI: QualityMetric Incorporated, **2002**.
17. Turner-Bowker DM, Saris-Baglana RN, DeRosa MA, Maruish ME. *The DYNHA headache impact test: A user's guide*. Lincoln, RI: QualityMetric Incorporated, **2009**.
18. Bjorner JB, Kosinski M, Ware JE Jr. Calibration of an item pool for assessing the burden of headaches: An application of item response theory to the Headache Impact Test (HIT[™]). *Qual Life Res* **2003**;12:913-933.
19. Kosinski M, Bayliss MS, Bjorner JB, Ware JE Jr., Garber WH, Batenhorst A, Cady R, Dahlöf CG, Dowson A, Tepper S. A six-item short-form survey for measuring headache impact: The HIT-6[™]. *Qual Life Res* **2003**;12:963-974.
20. Turner-Bowker DM, Saris-Baglana RN, Derosa MA, Paulsen CA, Bransfield CP. Using qualitative research to inform the development of a comprehensive outcomes assessment for asthma. *Patient* **2009**;2:1-14.
21. Dillman DA, Smythe JD. Design effects in the transition to web-based surveys. *Am J Prev Med* **2007**;32:S90-S96.

Address correspondence to:
Diane M. Turner-Bowker, Ph.D.
Datacorp
 7 Thurber Blvd.
 Smithfield, RI 02917

E-mail: dtbowker@mjdatacorp.com

Received: July 20, 2010
 Revised: August 3, 2010
 Accepted: August 4, 2010