Published in final edited form as:

Patient Educ Couns. 2009 June ; 75(3): 295–301. doi:10.1016/j.pec.2009.02.020.

Bilingual health literacy assessment using the Talking Touchscreen/la Pantalla Parlanchina: Development and pilot testing

Kathleen J. Yost^{1,2}, Kimberly Webster^{1,3}, David W. Baker⁴, Seung W. Choi^{1,3}, Rita K. Bode^{1,5}, and Elizabeth A. Hahn^{1,2,3}

¹ Center on Outcomes, Research and Education (CORE), NorthShore University HealthSystem, Evanston, Illinois, United States

² Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois, United States

³ Institute for Healthcare Studies, Northwestern University Feinberg School of Medicine, Chicago, Illinois, United States

⁴ Division of General Internal Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois, United States

⁵ Department of Physical Medicine and Rehabilitation, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, United States

Abstract

Objective—Current health literacy measures are too long, imprecise, or have questionable equivalence of English and Spanish versions. The purpose of this paper is to describe the development and pilot testing of a new bilingual computer-based health literacy assessment tool.

Methods—We analyzed literacy data from three large studies. Using a working definition of health literacy, we developed new prose, document and quantitative items in English and Spanish. Items were pilot tested on 97 English- and 134 Spanish-speaking participants to assess item difficulty.

Results—Items covered topics relevant to primary care patients and providers. English- and Spanish-speaking participants understood the tasks involved in answering each type of question. The English Talking Touchscreen was easy to use and the English and Spanish items provided good coverage of the difficulty continuum.

Conclusion—Qualitative and quantitative results provided useful information on computer acceptability and initial item difficulty. After the items have been administered on the Talking Touchscreen (la Pantalla Parlanchina) to 600 English-speaking (and 600 Spanish-speaking) primary care patients, we will develop a computer adaptive test.

Address correspondence to: Elizabeth A. Hahn, Center on Outcomes, Research and Education, (CORE), 1001 University Place, Suite 100, Evanston, Illinois 60201, United States. Phone: +1-224-364-7373. FAX: +1-847-570-8033. E-mail: e-hahn@northwestern.edu.

Disclosure Statement

The authors do not have any actual or potential conflicts of interest that could inappropriately influence (bias) our work.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Practice Implications—This health literacy tool will enable clinicians and researchers to more precisely the level at which low health literacy adversely affects health and healthcare utilization.

Keywords

literacy; multimedia; ethnic groups; psychometrics

1. Introduction

Low health literacy is associated with health disparities such as reduced access to health information, poorer understanding of illness and treatment, poorer health status, less understanding and use of preventive services, and increased hospitalizations [1]. A better understanding of patients' health literacy may allow for targeted interventions to improve health outcomes. However, current instruments used to assess health literacy in clinical practice and research either take too long, are imprecise, or have questionable equivalence of English and Spanish versions. We are using modern psychometric methods and novel health information technology in a project that will result in a tool to measure health literacy in less time and with greater precision than existing tests.

Interactive computer-based assessment provides an innovative method for gathering and using patient self-report data [2]. Computers provide opportunities to conduct personalized in-depth interviews without the expense and potential bias of interviewers. More recently, audio components have been added to create a multimedia testing platform, the "Talking Touchscreen" ("Pantalla Parlanchina"), that has been shown to validly assess health status in English and Spanish for patients with both low and high literacy with a range of computer skills [3,4]. This multimedia technology was selected for use in this health literacy study.

The purpose of this paper is to describe the development of health literacy items, creation of the Talking Touchscreen administration platform, and pilot testing of the items and computer interface.

2. Methods

This manuscript reports on the first three phases of our research project: 1) psychometric analysis of data from studies that administered the Test of Functional Health Literacy in Adults (TOFHLA) [5]; 2) development of new items to measure health literacy using a multimedia program; and 3) pilot testing the new items in the target population.

The development of the new health literacy assessment tool followed several steps: define the construct "health literacy," identify health-related topics to include in the item pool, create items based on real-world patient information, obtain expert review of items, and translate items into Spanish (Figure 1). Following these steps, we developed a multimedia program for self-administration of the items on a touchscreen computer. Qualitative and quantitative research methods were used in pilot testing to assess computer feasibility and acceptability and to obtain preliminary information on item difficulty. We partnered with primary care providers and community organizations to recruit participants from underserved populations for the purpose of testing and refining the item pools. Each of these steps is described in detail below.

2.1. Phase 1: Psychometric Analysis of TOFHLA data

The purpose of the first phase of this project was to evaluate the psychometric properties of the two components of the TOFHLA (Numeracy and Cloze/reading comprehension) [5] in order to identify potential gaps where new items might be needed. We obtained three large datasets: 1) Numeracy data for 2,659 English- or Spanish-speaking emergency department

patients (principal investigator: David Baker [6]); 2) Numeracy and Cloze data for 902 English- or Spanish-speaking primary care patients (principal investigator: Barry Weiss); 3) Cloze data for 97 English-speaking cancer patients (principal investigator: Elizabeth Hahn) and 203 English-speaking patients with asthma or heart failure (principal investigator: Arthur Evans). For each dataset we implemented Rasch analysis [7] to evaluate internal consistency reliability (coefficients > 0.80 were considered acceptable), item fit to a unidimensional measurement model (infit mean square statistics < 0.70 and > 1.40 were considered evidence of misfit) [8], and item difficulty (the percent answering the item correctly). A unidimensional measurement model specifies that one latent trait is estimated and that each item contributes in a meaningful way. Fit statistics help to ascertain whether the assumption of unidimensionality holds up empirically [8]. Differential item functioning (DIF) analyses were conducted to determine whether patient characteristics affected performance on individual Numeracy items, controlling for overall literacy [9,10]. In this study, DIF was examined by ethnicity (Hispanic versus non-Hispanic respondents who completed the English version), language (Hispanic respondents who completed the English versus Spanish versions), and combinations of ethnicity and language (non-Hispanics who completed the English version versus Hispanics who completed the Spanish version). Two methods were used to identify DIF: ordinal logistic regression as implemented in *OLRDIF* [9] and comparison of Rasch item calibrations as implemented in *Winsteps* [7]. An item was considered to exhibit DIF if significant differences were identified by both methods ($p < 0.01$ to adjust for multiple comparisons). DIF analyses were not possible for the Cloze items because the English and Spanish items are not comparable.

2.2. Phase 2: Development of new items

Before developing the assessment tool, we first clearly defined the construct we wanted to measure. We considered previous definitions offered by others [1] and then tailored it to specifically fit what could reasonably be measured using a multimedia assessment tool. We obtained input from a panel of experts on health literacy.

Our initial assumption is that health literacy is a unidimensional construct. Although no item set will ever perfectly meet strictly defined unidimensionality assumptions, one wants to assess whether scales are “essentially” or “sufficiently” unidimensional to permit unbiased scaling of individuals on a common latent trait [11,12]. The 1992 National Adult Literacy Survey (NALS) [13] and the 2003 National Assessment of Adult Literacy (NAAL) [14] created three different literacy scores. However, analyses showed strong correlations between the three scores and few practical distinctions among them, suggesting the usefulness of creating one overall composite score [15,16,17,18]. Specific analyses will be conducted to evaluate the dimensionality of our new health literacy tool.

2.2.1 Creation of items in English and Spanish—We proposed a stepwise methodology for item format, topic and concept development with a goal of deriving pools of three different types of items commonly used to measure literacy. The item types as defined for the NALS and NAAL are 1) prose, 2) document, and 3) quantitative. For our health literacy assessment tool, we adopted item-type definitions based on the NALS/NAAL framework and applied them to health-related materials.

Building upon this framework and the work documented by the developers of the TOFHLA [5], we developed standard formats for item administration. A modified native cloze technique was used for the *prose* items, which consist of a brief text passage of health-related information (approximately 40–60 words) followed by a single-sentence comprehension item with a missing word. A respondent must choose the correct missing word from four multiple choice options which are syntactically similar but semantically different. *Document* items are

administered with a related image (e.g., table, graph, prescription label). The item asks a question about information that can be located in the image. A respondent must choose the correct response from four multiple choice options. Typically, the foils (incorrect responses) for these items include data presented in the image that do not correspond to the question being asked. *Quantitative* items are often coupled with an image, but can also be in text format only. All quantitative items present the participant with a question that requires some type of arithmetic computation to determine the correct answer, which is chosen from four multiple choice options. Typically, the foils for these items include numbers/percentages that are the result of common computational errors or that are already provided in the image or text.

We circulated a list of topics to eight members of our advisory panel (seven of whom are practicing clinicians). Panel members identified topics most likely to be of interest and importance to both clinicians and patients. They were also asked to consider whether there is likely to be sufficient source material for the topic and whether they expect the source material to be fairly fixed (i.e., not likely to change dramatically in the near future due to advances in medicine or technology). They were allowed to write in suitable topics that were not on the list.

For each identified topic, we developed a set of prose, document, and quantitative items and images which covered a range of subtopics. Whenever possible, real world, copyright-free sources were used in the development of an item (e.g., prescription labels, web-pages, patient education materials, consent forms, instructions for diagnostic tests). Considerable effort was made to ensure that only one concept was clearly articulated and that the information was accurate. Modifications from the original source material were made to almost all items in an effort to remove the influence of confusing text or instructions and poorly illustrated tables or graphs on the ability to understand and respond to the question/task.

Items were developed to cover a wide range of reading levels with the majority targeted to the 6th to 11th grade level. Very few items above or below this range were developed because precise measurement of very low or very high health literacy is not needed. A combination of the Flesch-Kincaid index [19], the Lexile Framework[®] for Reading in English and Spanish [20], qualitative assessments of item difficulty levels from the advisory panel, and pilot testing data were used to inform item difficulty level and range.

Spanish translation of items and images was conducted by a certified team of language coordinators and translators with knowledge of medical terminology, using a validated multi-step translation methodology [21,22,23,24]. The goal of the translation was one of semantic not syntactic equivalence; that is, the goal was to capture the meaning of the item rather than perform a literal translation.

Our item development process included several iterations of internal and external expert review. English and Spanish Item Development and Translation Advisory Panels (IDTAP) were assembled to provide ongoing input on item content, difficulty and quality. The English IDTAP was comprised of 12 investigators and consultants with expertise in medicine, literacy assessment, cross-cultural research, and psychometrics. The Spanish IDTAP consisted of seven fully bilingual clinicians and researchers. Each item was rated for difficulty [Easy (<6th grade), Somewhat easy (6th–9th grade), Somewhat difficult (10th–12th grade), Difficult (>12th grade)], accuracy of the information provided in the text or image (accurate/not accurate), quality of item for measuring item-type skill (poor, fair, good, very good, excellent), and overall quality of item for measuring health literacy (poor, fair, good, very good, excellent). Items were also reviewed by psychometricians with expertise in reading level measurement. Feedback from the IDTAP and psychometricians was used to further refine items.

2.2.2. Development of a multimedia computer program—We adapted our bilingual multimedia “Talking Touchscreen” (“Pantalla Parlanchina”) [3,4] to administer health literacy items. Screens and audio files were created to closely mimic the current method used for the TOFHLA [5]. For example, participants can read the prose passages on their own and choose the missing word from the multiple choice responses by touching their chosen word on the touchscreen. For items that include an image, instead of having an interviewer hand the participant a card to read and then ask a question about it, the image appears on the screen along with a question and a set of response buttons. For document and quantitative items, the participant touches the screen if s/he would like to hear the question read out loud.

2.3. Phase 3: Pilot testing

Participants for the English pilot study were recruited from primary care clinics at Northwestern Medical Faculty Foundation and John H. Stroger Jr. Hospital, and from two community organizations: Albany Park Community Center Adult Education Program and Literacy Chicago. Participants for the Spanish pilot study were recruited from primary care clinics at Evanston Hospital and Access Community Health Network, and from two community organizations: Albany Park and National Able Network, a job training organization. All sites are in the Chicago metropolitan area. Participants were eligible to participate if they were at least 21 years of age, spoke English or Spanish, and had sufficient vision, hearing, cognitive function and manual dexterity to interact with the touchscreen. All participants completed a subset of health literacy items, provided informed consent in accordance with institutional review board requirements and received \$20 as compensation for their time.

2.3.1. Qualitative methods—We conducted cognitive interviews [25] with a subset of participants to elicit feedback on design aspects of the health literacy tool. The purpose was to identify potential problems with item formats and instructions, and to incorporate preferences into the design of the computer interface. The cognitive interviews focused on 1) understandability of the task required to answer different types of items, 2) the extent to which answering the literacy questions were uncomfortable or anxiety-provoking, 3) acceptability of the computer interface, and 4) preferences for an audio symbol (“Which of these symbols do you think would be the most clear for letting people know that a question can be read out loud?”).

2.3.2. Quantitative methods—Although the item types differ in format and task, all items are essentially dichotomous because the responses are scored as correct or incorrect. Analyses were conducted to obtain a rough estimate of item difficulty and quality. We calculated an adjusted point-biserial correlation, r_{pb} , for each health literacy item to evaluate the relationship between item responses (correct/incorrect) and the total aggregated score for all other administered items. The maximum possible value for r_{pb} is about 0.80, and we set a criterion value (<0.20) to identify poorly performing items [26].

3. Results

3.1. Phase 1: Psychometric Analysis of TOFHLA data

The basic item statistics for Numeracy and Cloze showed a pronounced positive skew in the data. Many items had weak incorrect alternatives which few patients selected as the correct response, resulting in the items being very easy to answer correctly. For example, one-third of the Numeracy items were answered correctly by more than 80% of respondents in one dataset, and two-thirds of the Cloze items were answered correctly by more than 80% of the respondents in two of the datasets. Rasch model misfit was found in one Numeracy item, and in six Cloze items. In the DIF analyses of the Numeracy items, the number of items exhibiting DIF differed across comparisons; for example, one item was identified when only ethnicity (Hispanic vs.

non-Hispanic) varied, six when only language varied, and nine when both ethnicity and language varied [27]. Based on the results of these psychometric analyses, we decided to set the TOFHLA items aside and to develop a new set of health literacy items.

3.2. Phase 2: Development of new items

The resulting definition of health literacy used in this study is as follows: “*Health Literacy is the degree to which individuals have the capacity to read and comprehend health-related print material, identify and interpret information presented in graphical format (charts, graphs, tables), and perform arithmetic operations in order to make appropriate health and care decisions.*” This definition has essentially two parts: capacity and application. First, an individual must have the capacity to process and understand health-related information. S/he must then be able to apply that information in the management of her/his own health. We propose that the capacity to obtain information, which is part of previous definitions [1], is a navigation skill that cannot be included in this health literacy tool. Instead, we focus on comprehending and interpreting information provided and understanding what an appropriate health care decision based on that information should be. Whether the patient actually implements an appropriate health care decision and related behavior is also beyond the capability of this assessment tool. There may be many barriers between understanding what an appropriate decision is and actually implementing that decision in one’s own behavior.

3.2.1. Creation of items in English and Spanish—The advisory panel reviewed an initial list of topics organized into three categories: disease/health (24 topics); Medicare/Medicaid/Insurance (11 topics); and consent/HIPAA (11 topics). The final list of 13 topics is in Table 1. Mental health was not on the original topic list, but was added at the request of an IDTAP member.

We then developed 138 items in English: 58 prose, 39 document, and 41 quantitative. Ninety-eight of them were tested in this pilot study and 40 were written afterwards. We also developed 25 unique images to complement document and quantitative items, several of which were used for more than one item. Prior to expert review, one image and three items were dropped due to incompatibility with the format defined for touchscreen administration, two for lack of relevance, and one for overlap; therefore, 132 English items and 24 images were fully developed. All items and images were translated into Spanish.

3.2.2. Development of a multimedia computer program—English items, including sound files and images, were loaded onto four touchscreen laptops. Figures 2 – 4 are examples of prose, document and quantitative items, respectively as they appear on the Talking Touchscreen. A comparable multimedia computer program is currently being developed for the Spanish items.

3.4. Phase 3: Pilot testing

This manuscript presents selected data on 97 English- and 134 Spanish-speaking participants in the pilot study. Characteristics of the participants are summarized in Table 2. The English-speaking participants were predominantly female and African-American. When asked their race, the Spanish-speaking participants tended to respond with their ethnicity, “Hispanic,” rather than with a race. These participants were also predominantly female. Spanish-speaking participants were younger than the English-speaking participants.

A total of 98 English items (42 prose + 27 document + 29 quantitative) were tested. Seventy-two participants completed the items by self-administration on paper-and-pencil and 25 completed them by self-administration on the Talking Touchscreen using earphones to hear the audio component for the document and quantitative items. A total of 127 Spanish items

were tested (54 prose + 36 document + 37 quantitative). All Spanish items were completed by self-administration on paper-and-pencil. All participants answered approximately 30 items that were a mix of prose, document and quantitative items. For paper-and-pencil administration, participants were instructed that they could ask the research assistant to read aloud any text that had a sound symbol next to it.

3.4.1. Qualitative data—Cognitive interviews were conducted with the first 24 English- and 14 Spanish-speaking participants. The overwhelming majority (over 90% English, 100% Spanish) correctly described the steps needed to answer each type of question (prose, document, quantitative), e.g., “*the instructions are clear, you read the question and choose the correct answer to complete the sentence;*” “*look at entire chart, look at question, then back at chart and choose answer;*” “*I read the question and looked at choices below, I used math skills.*” Participants were comfortable answering the questions; only one English- and three Spanish-speaking participants reported feeling uncomfortable or anxious: “*I wasn’t sure if I was answering correctly and I wanted to make sure I was;*” “*I felt anxious when I didn’t understand a word.*”

The majority of English- (59%) and Spanish-speaking (74%) participants preferred the symbol of the man talking (🗣️) as an indicator that audio was available for an item. Among the 25 English-speaking participants who completed the pilot test on a touchscreen computer, most reported that it was easy to use and commented favorably on the screen design and the availability of audio: “*I don’t know much about computers, but it was easy;*” “*It was fun;*” “*The words were big so I didn’t have to put glasses on. The screen was nice and bright;*” “*Simple –having the questions read to me made me feel more comfortable;*” “*I like the idea of using a touchscreen. I like the headphones and it allowed me to concentrate better.*” There were only a few negative comments about the computer: “*I prefer for an actual person to ask me the questions;*” “*I didn’t like the pen [stylus], probably because I never used a pen on a computer.*” Participants also commented favorably on the health literacy items themselves, even when acknowledging that some of them were difficult to answer: “*They were difficult, but interesting;*” “*The beginning was easy and then some information became difficult. I learned a lot;*” “*I liked the diabetes, high blood pressure and cholesterol questions, but some questions I didn’t understand;*” “*The questions were easy and understandable, but the medical words were hard;*” “*I learned something about health today.*”

3.4.2. Quantitative results—Twenty-one of the 98 English items were flagged as problematic based on low adjusted point-biserial correlations (< 0.20). Two items were too easy (all participants got them correct). These 23 and one redundant item were dropped from the item pool. The 74 items remaining in the pool showed a wide range of item difficulties (proportion correct) ranging from 0.22 to 0.96 (mean=0.65 and SD=0.16). The item pool provided good coverage of items along the difficulty continuum. The range of item difficulties (and the number of items in the range) are as follows: <0.30 (2); 0.30–0.39 (4), 0.40–0.49 (9), 0.50–0.59 (6), 0.60–0.69 (22), 0.70–0.79 (18), 0.80–0.89 (9), and ≥ 0.90 (4).

To compensate for these items dropped based on pilot data, an additional 40 English items were written using procedures described above. From this item pool, we selected the best 90 items based on pilot data results, IDTAP review, and content coverage and relevance.

Fifty-three of the 127 Spanish items were problematic. Forty-three items had adjusted point-biserial correlations below 0.20 and 10 items were too easy (all participants got them correct). We selected the best 90 Spanish items based on pilot data, IDTAP review, content and approximate comparability to the 90-item English set. This strategy resulted in retention of 14 items with adjusted point-biserial correlations below 0.20, one item that was not pilot tested in Spanish and one item with a proportion correct of 1.0. Difficulties (proportion correct) for

the 90 items ranged from 0.12 to 1.0 (mean=0.68 and SD=0.21). The range of item difficulties (and the number of items in the range) are as follows: <0.30 (4); 0.30–0.39 (6), 0.40–0.49 (8), 0.50–0.59 (11), 0.60–0.69 (11), 0.70–0.79 (20), 0.80–0.89 (16), and ≥ 0.90 (13).

4. Discussion and Conclusion

4.1. Discussion

We began this project with an evaluation of the TOFHLA [5] in order to identify gaps in the health literacy continuum where new items were needed. Although the TOFHLA has been widely validated as a measure of health literacy and is predictive of health outcomes, we found several problems. First, for many items, a very high proportion of respondents got the item correct. This is not surprising considering the first reading comprehension passage is written at a fourth grade level. Thus, for the majority of people who read well above this level, these items provide little information. Second, several items showed measurement bias (DIF) between Hispanics and non-Hispanics and between English and Spanish-speaking participants. This measurement bias means that using the TOFHLA to compare literacy levels between groups that differ on ethnicity or language may result in biased estimates of the true differences. However, since stable estimation of item parameters is a critical prerequisite for some DIF detection methods [28], the skewness of the data may have affected our ability to conduct DIF analyses. For these reasons, we decided to develop an entirely new set of items to provide better measurement precision, especially in the mid-range of reading levels, and to avoid measurement bias by ethnicity or language.

The new health literacy items have good content validity, covering a variety of topics that should be relevant to primary care patients and their healthcare providers. Results from cognitive interviewing showed that participants were able to correctly describe the steps needed to answer each type of question, that they were comfortable completing a literacy test, and that the Talking Touchscreen was acceptable and easy to use. Results of the quantitative analysis allowed us to identify and discard items that were too easy or performed poorly.

There are some limitations to the study. Because computer programming tasks were ongoing during the pilot test, some English-speaking participants and all of the Spanish-speaking participants completed the pilot test on paper. It is unknown whether this introduced any response bias due to different methods of administration. However, data from the pilot test will not be used for the final calibrations. Another limitation is that additional items were written after pilot testing and are being used in the main testing phase of this project. Given that these items underwent extensive internal and external expert review, they should be of similar quality as all other items.

We are currently administering the 90 English items on the Talking Touchscreen to 600 English-speaking primary care patients and we are preparing to administer the 90 Spanish items on “la Pantalla Parlanchina” to 600 Spanish-speaking primary care patients. These data will be used to develop item banks. Items in a well-constructed bank cover the entire continuum of the trait and are calibrated on the same measurement scale, thus simplifying scoring and interpretation [29,30]. With a calibrated item bank in place, the same pool of questions can be used to create fixed-length test instruments of various lengths and even computer adaptive tests (CAT).

4.2. Conclusion

The results of the TOFHLA analyses confirmed the need to develop a more comprehensive pool of items to measure health literacy. With properly calibrated item banks, it will be possible to administer a small targeted set of items to each person, thereby providing a brief, real-time,

tailored assessment tool for use in clinical settings. Modern psychometric methods also provide useful techniques to identify and control for potential cultural or linguistic bias, and to establish equivalent scores across English and Spanish items.

4.3. Practice Implications

A bilingual health literacy tool developed using state-of-the-science psychometric methods will enable clinicians and researchers to more precisely determine at what level low health literacy begins to adversely affect health and healthcare utilization. By using novel computer-based methods for health literacy assessment in clinical settings, this tool will also increase the access of underserved populations to new technologies, and contribute information about the experiences of diverse populations with new technologies.

Acknowledgments

The authors thank Barry Weiss, M.D., University of Arizona, and Arthur Evans, M.D., Cook County Hospital, for providing access to the literacy assessment data they collected at their clinical sites. We also thank Irina Kontsevaia, M.S. for programming the Talking Touchscreen, and Jack Stenner, Ph.D. and Eleanor Sanford-Moore, Ph.D. for consultation on reading level measurement. Finally, the authors thank the following members of the Item Development and Translations Advisory Panel for reviewing the items: Hugo Alvarez, M.D.; Yolanda Ayubi, Ph.D.; David Cella, Ph.D.; Paul Cella, B.A.; Darren Dewalt, M.D.; Emanuel Diaz, M.D.; Deborah Dobrez, Ph.D.; Sofia Garcia, Ph.D.; Yvette Garcia, M.A.; Richard Gershon, Ph.D.; Ronald Hambleton, Ph.D.; Elizabeth Jacobs, M.D.; Lee Lindquist, M.D.; Maria McCullen, R.N.; Jairo Mejia, M.D.; Joanne Nurss, Ph.D.; Veronica Valenzuela, B.A.; Priscilla Vasquez, B.A.; David Victorson, Ph.D.; and Barry Weiss, M.D.

Role of the funding source

This study was supported by grant number R01-HL081485 from the National Heart, Lung, and Blood Institute. The study sponsor had no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication.

References

1. Committee on Health Literacy. Health Literacy: A Prescription to End Confusion. Washington, D.C: The National Academies Press; 2004.
2. Cella DF, Lloyd SR. Data collection strategies for patient-reported information. *Qual Manag Health Care* 1994;2:28–35. [PubMed: 10137605]
3. Hahn EA, Cella D, Dobrez DG, Shiimoto G, Taylor SG, Galvez AG, Diaz P, Valenzuela V, Chiang HL, Khan S, Hudgens SA, Du H. Quality of life assessment for low literacy Latinos: A new multimedia program for self-administration. *J Oncol Manag* 2003;12:9–12. [PubMed: 14619989]
4. Hahn EA, Cella D, Dobrez D, Shiimoto G, Marcus E, Taylor SG, Vohra M, Chang CH, Wright BD, Linacre JM, Weiss BD, Valenzuela V, Chiang HL, Webster K. The Talking Touchscreen: a new approach to outcomes assessment in low literacy. *Psychooncology* 2004;13:86–95. [PubMed: 14872527]
5. Parker RM, Baker DW, Williams MV, Nurss JR. The test of functional health literacy in adults: A new instrument for measuring patients' literacy skills. *J Gen Intern Med* 1995;10:537–41. [PubMed: 8576769]
6. Baker DW, Williams MV, Parker RM, Gazmararian JA, Nurss J. Development of a brief test to measure functional health literacy. *Patient Educ Couns* 1999;38:33–42. [PubMed: 14528569]
7. Linacre, JM. A User's Guide to WINSTEPS/MINISTEP: Rasch-Model Computer Programs. Chicago, IL: Winsteps; 2008.
8. Bond, TG.; Fox, CM. Applying the Rasch Model: Fundamental measurement in the human sciences. Mahway, NJ: Lawrence Earlbaum Associates, Inc; 2001.
9. Zumbo, BD. A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and likert-type (ordinal) item scores. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense; 1999.

10. Teresi JA. Overview of quantitative measurement methods: equivalence, invariance, and differential item functioning in health applications. *Med Care* 2006;44:S39–S49. [PubMed: 17060834]
11. McDonald, RP. *Test Theory: A unified treatment*. Mahwah, NJ: Lawrence Earlbaum Associates, Inc; 1999.
12. Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, Thissen D, Revicki DA, Weiss DJ, Hambleton RK, Liu H, Gershon R, Reise SP, Lai JS, Cella D. Psychometric Evaluation and Calibration of Health-Related Quality of Life Item Banks: Plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med Care* 2007;45:S22–S31. [PubMed: 17443115]
13. Kirsch, I.; Jungeblut, A.; Jenkins, L.; Kolstad, A. *Adult literacy in America: A first look at the results of the National Adult Literacy Survey*. Washington, DC: National Center for Education Statistics, US Department of Education; 1993.
14. Kutner, M.; Greenberg, E.; Jin, Y.; Paulsen, C. *The Health Literacy of America's Adults*. Washington, DC: National Center for Education Statistics: U.S. Department of Education; 2006.
15. Salganik, LH.; Tal, J. *A Review and reanalysis of the ETS/NAEP young adult literacy survey*. Washington, DC: Pelavin Associates; 1989.
16. Rock, DA.; Yamamoto, K. *Construct validity of the adult literacy subscales*. Princeton, NJ: Educational Testing Service; 1994.
17. Reder, S. *Dimensionality and construct validity of the NALS assessment*. Westport, CT: Greenwood Publishing Group; 1996.
18. National Research Council. *Measuring Literacy: Performance levels for adults*. Washington, DC: The National Academies Press; 2005.
19. Kincaid, JP.; Fishburne, RP., Jr; Rogers, RL.; Chissom, BS. *Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy enlisted personnel*. Millington, TN: Naval Technical Training, U.S. Naval Air Station, Memphis, TN; 1975.
20. Stenner, AJ.; Wright, B. Uniform Reading and Readability Measures. In: Wright, BD.; Stone, MH., editors. *Making Measures*. Chicago: The Phaneron Press; 2004. p. 79-115.
21. Bonomi AE, Cella DF, Hahn EA, Bjordal K, Sperner-Unterweger B, Gangeri L, Bergman B, Willems-Groot J, Hanquet P, Zittoun R. Multilingual translation of the Functional Assessment of Cancer Therapy (FACT) quality of life measurement system. *Qual Life Res* 1996;5:309–20. [PubMed: 8763799]
22. Cella D, Hernandez L, Bonomi AE, Corona M, Vaquero M, Shiimoto G, Baez L. Spanish language translation and initial validation of the functional assessment of cancer therapy quality-of-life instrument. *Med Care* 1998;36:1407–18. [PubMed: 9749663]
23. Eremenco SL, Cella D, Arnold BJ. A comprehensive method for the translation and cross-cultural validation of health status questionnaires. *Eval Health Prof* 2005;28:212–32. [PubMed: 15851774]
24. Wild D, Grove A, Martin M, Eremenco S, Ford S, Verjee-Lorenz A, Erickson P. Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient reported outcomes(PRO) Measures: Report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value Health* 2005;8:94–104. [PubMed: 15804318]
25. Willis, GB. *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks, CA: Sage Publications; 2005.
26. Nunnally, JC.; Bernstein, IH. *Psychometric Theory*. New York: McGraw-Hill, Inc; 1994.
27. Bode, RK.; Yost, KJ.; Webster, K.; Baker, DW.; Weiss, BD.; Hahn, EA. Effects of confounding ethnicity and language in DIF studies. Poster presented at the 5th Conference of the International Test Commission, 2006; July 6–8; Brussels, Belgium. 2006.
28. Hambleton RK. Good practices for identifying differential item functioning. *Med Care* 2006;44:S182–S188. [PubMed: 17060826]
29. Hays RD, Morales LS, Reise SP. Item Response Theory and health outcomes measurement in the 21st century. *Med Care* 2000;38:28–42.
30. Hahn EA, Cella D, Bode RK, Gershon R, Lai JS. Item Banks and Their Potential Applications to Health Status Assessment in Diverse Populations. *Med Care* 2006;44:S189–S197. [PubMed: 17060827]

Figure 1.
Item Development Procedures

After a medical test or procedure, be sure to get the results. Ask whether you will get them in person or by phone or by mail. Also ask when you will get the results. Do not assume the results are fine if you do not get them when expected. If you do not get them, call your doctor.

You should _____ all test results.

Figure 2.
Prose item screen shot from the Talking Touchscreen

Medications for Mr. Beta

Medication	Start Date	End Date	Instructions
Hanebrex: 200 mg tablets	Aug. 27	Sept. 26	1 Tablet daily
Yostatin: 250 mg tablets	Mar. 8	None	1 Tablet twice daily
Nandozol: 90 mcg per puff	Mar. 8	None	1-2 Puffs by mouth every 4-6 hours as needed
Cellacillin: 250 mg tablets	Apr. 22	Apr. 29	2 Tablets on the first day, then 1 Tablet daily after that

Look at the Medications for Mr. Beta. How many tablets of Cellacillin should he take on the third day?

1

2

3

4

Figure 3.

Document item screen shot from the Talking Touchscreen

Sample Body Mass Index Chart								
		Weight in Pounds						
		145	155	165	175	185	195	
Height in Inches	66	23.4	25.0	26.6	28.2	29.9	31.5	Underweight
	67	22.7	24.3	25.8	27.4	29.0	30.5	
	68	22.0	23.6	25.1	26.6	28.1	29.6	
	69	21.4	22.9	24.4	25.8	27.3	28.8	Normal
	70	20.8	22.2	23.7	25.1	26.5	28.0	
	71	20.2	21.6	23.0	24.4	25.8	27.2	
	72	19.7	21.0	22.4	23.7	25.1	26.4	Overweight
	73	19.1	20.4	21.8	23.1	24.4	25.7	
	74	18.6	19.9	21.2	22.5	23.8	25.0	
	75	18.1	19.4	20.6	21.9	23.1	24.4	Obese
	76	17.6	18.9	20.1	21.3	22.5	23.7

Look at the Sample Body Mass Index Chart. In what category would a body mass index of 23.2 be?

Underweight

Normal

Overweight

Obese

Figure 4.
Quantitative item screen shot from the Talking Touchscreen

Table 1
Final List of Topics

Disease/Health-Related Topics
Diabetes
Hypertension
Nutrition and Obesity
Instructions (e.g., pre-operative, obtaining an X-ray etc.)
Asthma
Risks/Benefits of Immunizations
General Cancer Information and Cancer Screening
Mental health
Medicare/Medicaid/Insurance-related Topics
Determining your Premium/Deductible/Costs
Pharmaceutical Plans/Drug Co-Pay Coverage
Informed Consent/HIPAA-Related Topics
Risk/Benefits of Participation in Research Study
Confidentiality and Acceptable uses of Personal Information
Understanding Medical Care (e.g., risk/benefits of medical procedure, consent to receive medical care)

Table 2

Sociodemographic characteristics and administration methods for pilot testing participants

	English- speaking (n=97)	Spanish- speaking (n=134)
	%	
Female	65.0	74.6
Hispanic Ethnicity	18.6	99.3
Race		
White	7.2	21.6
Black/African-American	59.8	0
Asian	13.4	0
American Indian/Alaskan Native	1.0	0
Native Hawaiian/Pacific Islander	0	0
Other	16.5	78.3
Multiple races	1.0	0
Age		
21–30	21.7	26.8
31–40	13.4	38.8
41–50	27.8	18.7
51–60	21.7	10.5
61+	11.3	5.2
Education		
Not asked/missing	27.8	0
None	1.0	0
Grades 1–8	10.3	30.6
Grades 9–11	19.6	22.4
High School Graduate/GED	22.7	24.6
Some college or college graduate	18.6	22.4
Method of administration		
Paper-and-pencil	74.2	100
Talking Touchscreen	25.8	---
Recruitment Site		
Primary Care Clinics	28.9	46.3
Community Centers/Organizations	71.1	53.7