

# Development and initial validation of a computer-administered health literacy assessment in Spanish and English: FLIGHT/VIDAS

Raymond L Ownby<sup>1</sup>  
 Amarilis Acevedo<sup>2</sup>  
 Drenna Waldrop-Valverde<sup>3</sup>  
 Robin J Jacobs<sup>1</sup>  
 Joshua Caballero<sup>4</sup>  
 Rosemary Davenport<sup>1</sup>  
 Ana-Maria Homs<sup>1</sup>  
 Sara J Czaja<sup>5</sup>  
 David Loewenstein<sup>5</sup>

<sup>1</sup>Department of Psychiatry and Behavioral Medicine, Nova Southeastern University, Fort Lauderdale, FL, USA; <sup>2</sup>Center for Psychological Studies, Nova Southeastern University, Fort Lauderdale, FL, USA; <sup>3</sup>Nell Hodgson Woodruff School of Nursing, Emory University, Atlanta, GA, USA; <sup>4</sup>College of Pharmacy, Department of Pharmacy Practice, Nova Southeastern University, Fort Lauderdale, FL, USA; <sup>5</sup>University of Miami Miller School of Medicine, Miami, FL, USA



Correspondence: Raymond L Ownby  
 Room 1477, 3200 South University  
 Drive, Fort Lauderdale, FL 33314, USA  
 Tel +1 954 262 1481  
 Email ro71@nova.edu

**Abstract:** Current measures of health literacy have been criticized on a number of grounds, including use of a limited range of content, development on small and atypical patient groups, and poor psychometric characteristics. In this paper, we report the development and preliminary validation of a new computer-administered and -scored health literacy measure addressing these limitations. Items in the measure reflect a wide range of content related to health promotion and maintenance as well as care for diseases. The development process has focused on creating a measure that will be useful in both Spanish and English, while not requiring substantial time for clinician training and individual administration and scoring. The items incorporate several formats, including questions based on brief videos, which allow for the assessment of listening comprehension and the skills related to obtaining information on the Internet. In this paper, we report the interim analyses detailing the initial development and pilot testing of the items (phase 1 of the project) in groups of Spanish and English speakers. We then describe phase 2, which included a second round of testing of the items, in new groups of Spanish and English speakers, and evaluation of the new measure's reliability and validity in relation to other measures. Data are presented that show that four scales (general health literacy, numeracy, conceptual knowledge, and listening comprehension), developed through a process of item and factor analyses, have significant relations to existing measures of health literacy.

**Keywords:** cognition, disparities, item response theory

## Introduction

Health literacy, defined as an individual's ability to obtain health-related information and use it to make decisions,<sup>1</sup> is increasingly recognized as an important factor in patient health. Several reviews show that individuals' health literacy is related to their health status, function, and use of services<sup>2,3</sup> and it has even been related to increased risk of mortality.<sup>4,5</sup> The existence of effective interventions to improve health literacy<sup>6,7</sup> highlights the possibility that improving this may be a strategy for improving health outcomes and addressing race- and ethnicity-related health disparities.<sup>8,9</sup>

Commonly used measures of health literacy include the Test of Functional Health Literacy in Adults, or TOFHLA,<sup>10</sup> the Rapid Estimate of Adult Literacy in Medicine, or REALM,<sup>11</sup> and the Newest Vital Sign.<sup>12</sup> Each measure has strengths and weaknesses. The TOFHLA, for example, assesses patients' ability to understand what they read as well as their numeracy skills. However, limitations of the TOFHLA are the requirement that the clinician administering it be trained and the time required for the clinician to individually administer and score it, typically at least 30 minutes – the time required for administration thus limits its use in clinical and research settings. A shorter version,

the S-TOFHLA, is available<sup>13</sup> but suffers from ceiling effects (many people achieve high scores) that limit its use in research since a limited range of scores affects the ability to detect its relations to other variables.

The REALM also must be administered, scored, and interpreted by a trained clinician. Also, this measure only assesses health literacy as patients' ability to correctly pronounce a series of health-related words (eg, anatomical terms and the names of diseases and conditions) and thus does not directly assess their ability to understand what they read. The REALM does not assess numeracy skills, consistently shown to be an important aspect of health literacy. The Newest Vital Sign only assesses patients' comprehension of a single food label and thus only taps a very narrow range of skills; it may have limited use except perhaps for the purpose of detecting whether a patient has poor reading comprehension skills.

Baker noted the limitations of existing measures of health literacy some time ago,<sup>14</sup> and problems encountered in assessing health literacy were summarized more recently by Pleasant and McKinney<sup>15</sup> and in an empirical review by Jordan et al.<sup>16</sup> Existing measures have been criticized for assessing a limited set of skills and for their development using patients drawn from single racial, ethnic, age, or socioeconomic groups. Other criticisms have noted the limited content and face validity of the measures and limited demonstrations of the measures' construct validity.<sup>16</sup> Further, although both Spanish and English versions of several measures are available, they were not developed using psychometric procedures that establish their equivalence across languages, making comparisons difficult.

An issue limiting the usefulness of the TOFHLA is the response format it uses in evaluating reading comprehension. The TOFHLA uses the "cloze" procedure<sup>17</sup> to assess reading comprehension. In this approach, comprehension is tested by asking the person assessed to supply a word missing in a sentence (eg, "The sky is \_\_\_\_"). This approach may create items that are differentially more difficult for older persons. Cloze procedure performance has been related to information processing speed and verbal fluency, which are reduced in older persons,<sup>18</sup> and data presented in a paper currently submitted for publication from our group suggest the presence of differential item functioning (DIF) on a significant number of items from the reading comprehension subtest of the TOFHLA, for people over 50. Item DIF occurs when individuals from different groups, such as men or women or racial groups, who have the same level of ability, have different probabilities of answering an item correctly. The empirical finding of this kind of difference is usually

interpreted as evidence that some factor besides the person's actual ability affects their performance, perhaps cultural, linguistic, or some other bias.<sup>19</sup> The finding of age-related DIF on the TOFHLA reading comprehension subtest suggests that other item formats (eg, multiple choice questions) may be more appropriate for use in assessing health literacy in older persons.

Almost all existing paper-and-pencil measures require hand scoring, making them time- and effort-intensive. Clearly, a computer-administered and -scored measure of health literacy would make an assessment more accessible in both the clinical and research settings by reducing demands on clinician or researcher time, while better standardizing the measure's administration. Integration of such a measure into an electronic health record might allow for inclusion of health literacy scores into patients' health records. This would transmit information about patients' level of health literacy directly to treating clinicians, allowing them to better understand patients' information needs. The automated assessment of health literacy might also allow for the automated tailoring of disease-related information, a strategy previously shown to be effective in influencing patient behavior<sup>20,21</sup> and which may be effective in reducing health disparities.<sup>22</sup>

Pleasant et al<sup>23</sup> have argued that new measures of health literacy should be multidimensional and assess health literacy as a latent construct. A multidimensional approach would recognize that functional health literacy comprises a number of distinct skills or abilities, such as reading, listening, and performing quantitative operations.<sup>1</sup> Evaluation of latent constructs is frequently used in psychological assessment to study an ability or trait that cannot be directly measured. Multiple test items believed to be related are administered, and then, what they have in common is statistically extracted, usually with factor analysis. Many item response theory (IRT) models approach the measurement of abilities as latent constructs and have been used to develop assessments of health literacy.<sup>24–26</sup> Pleasant et al<sup>23</sup> also suggested that assessments should recognize that measures are most likely to be accurate when they are similar to the context in which the actual behavior occurs. Assessment using a video simulation of a clinical encounter, for example, may be more accurate than asking for responses to written questions.

Jordan et al<sup>16</sup> reported a review of existing health literacy measures and found many of them lacking in important measurement characteristics. These authors noted the great variability in the content assessed by measures and the lack of a coherent conceptual model underlying them. Although measures such as the TOFHLA and the REALM

provide descriptive score categories, such as “adequate” or “inadequate” to assist in interpretation, the rationale for them is not clear. Jordan et al<sup>16</sup> also reported limitations in the construct validity of the different measures, noting that the correlations of the measures with other measures of health literacy and reading are quite variable. These findings imply that different measures may actually evaluate different abilities and skills, calling into question the typical interpretation of measures of the same construct, an issue also raised by Haun et al.<sup>27</sup> Finally, Jordan et al<sup>16</sup> assessed the feasibility of actually using the measures they reviewed, noting that the need for time, individual administration, and scoring was a substantial limitation that may limit their use.

Several researchers have addressed these limitations in developing new assessments of health literacy. In work reported by Hahn et al<sup>24</sup> and by Yost et al,<sup>26</sup> researchers created a health literacy assessment using a touch screen computer format they called the “Health Literacy Assessment Using Talking Touchscreen Technology,” or Health LiTT. Their measure, developed in Spanish and English<sup>26</sup> allows for automated administration and scoring and was developed using IRT methods. Data on this measure’s development in Spanish is limited, however, and the sample of Spanish-speaking adults used in the development efforts was not clearly characterized with respect to bilingualism or linguistic preference. How the Spanish-speaking participants were chosen to be tested in Spanish or English was not clearly described nor was their level of acculturation. The measure is not equivalent in Spanish and English, as the test stimuli differ in the two versions of the measure, limiting its usefulness in research, and both the Spanish- and English-speaking groups were patients in primary care with low levels of educational attainment. Although clearly a relevant population, the development of the measure with persons likely to have a limited range of ability may indicate that the measure will not function well in assessment of persons with higher levels of ability. The importance of understanding patients’ English competence when information is delivered to non-native speakers has been shown in studies by several investigators.<sup>28,29</sup> Thus, the measure may replicate the commonly observed ceiling effect from the TOFHLA. The Health LiTT measure is not based on a coherent theory or conceptual model of health literacy, although the authors link its development to an existing descriptive definition of health literacy.<sup>26</sup> Further, the reading comprehension section of this measure continues to rely on the cloze procedure, which as noted above, may result in items that are differentially more difficult for older participants. Finally, in a 2011 publication, the authors

reported that the Health LiTT would be available through The Assessment Center (<http://www.assessmentcenter.net>), a free online resource that allows investigators to access a standardized set of measures for use in research. At the time of this writing, however, it is not available on this site. The actual availability of this measure for use is not clear.

Lee et al<sup>25</sup> developed an instrument based on the REALM, selecting items based on analyses of DIF between Spanish- and English-speaking patients seen in a primary care clinic. As with other measures of health literacy, this measure continues to tap a narrow range of content and has limited demonstrations of its relation to other measures that might help establish its validity. It does not assess numeracy at all and only provides a limited assessment of comprehension. It thus suffers from some of the limitations others have criticized, including sampling a limited range of content, uncertain relation to actual health behaviors, and development on a small population of clinic patients.

A group at the Research Triangle Institute led by Lauren McCormack has also developed a new measure of health literacy, the Health Literacy Skills Instrument (HLSI).<sup>30,31</sup> This measure was developed using a rigorous psychometric approach and can be computer administered and scored. The development population was broader than that used to create most other measures (research volunteers versus clinic patients in many other studies), and the development population was large (several thousand). Analyses of its validity have been presented.<sup>30,32</sup> The manual for this measure, however, does not provide directions for administering the measure, either in person or by computer, raising questions about whether the measure could be reliable without a standard approach to administration.<sup>31</sup> The authors have suggested that the measure can be administered as a paper and pencil test, but it uses an audio recording to assess listening comprehension and requires access to several web pages to answer two questions, thus raising questions about how this might be possible. Such an administration would, again, not be standardized, raising questions about the validity of this form.

Although the article describing the measure stated that five subscales could be constructed from the 25 items in the final measure, the final test manual provided by the authors does not describe how to score them, nor does either of the articles describing its development and use provide data on the psychometric characteristics (reliability, validity) of the subscales. The manual for the measure states that it assesses oral health literacy, and this statement appears to refer to two items on the measure that require that the person assessed listen to a telephone menu recording and determine which

button on the telephone to press. It thus assesses listening comprehension with two items. Similarly, the measure states that it assesses “information seeking” skills on the Internet, but this statement refers to two items in which a hyperlink is provided that takes the person assessed directly to a web page.<sup>31</sup> The two pages include calculators for calories burned during exercise and risk for heart attack. While these are useful information skills, these items do not actually assess an individual’s ability to locate information on the Internet, an increasingly important skill.<sup>33,34</sup> Finally, the HLSI is not available in Spanish, so it may have limited usefulness with the most rapidly growing minority population in the United States.

This measure thus addresses many of the issues previously raised in critical evaluations of measures of health literacy. Its psychometric characteristics have been established, and it taps a wider range of content than do other measures. It includes two items tapping listening comprehension, although it appears likely that these items cannot be used as a separate scale. The measure can be computer administered and scored, but no standard format for this administration is provided, raising questions about the reliability of test scores that might result from diverse approaches to administration. It can be administered by computer, but as described, it appears likely that this administration format requires the use of a computer mouse to answer questions and to navigate hyperlinks. Given the difficulty many elders and others with little computer experience have in using a computer mouse (and especially the fine psychomotor skills required to click the small dots used by this instrument, to select a preferred answer), it is likely that the format of administration as developed, may place the very groups in which health literacy is most important at a disadvantage when responding to its questions. Although it includes items that mimic health-related calculators that might be found on the Internet, they do not actually evaluate Internet information search. As the web pages are on an external server, the ability to administer them requires an Internet connection and limits the ability of users to administer the measure by paper and pencil. Finally, the measure is brief, does not have separate subscales for skills such as reading, numeracy, and listening, and is not available in Spanish.

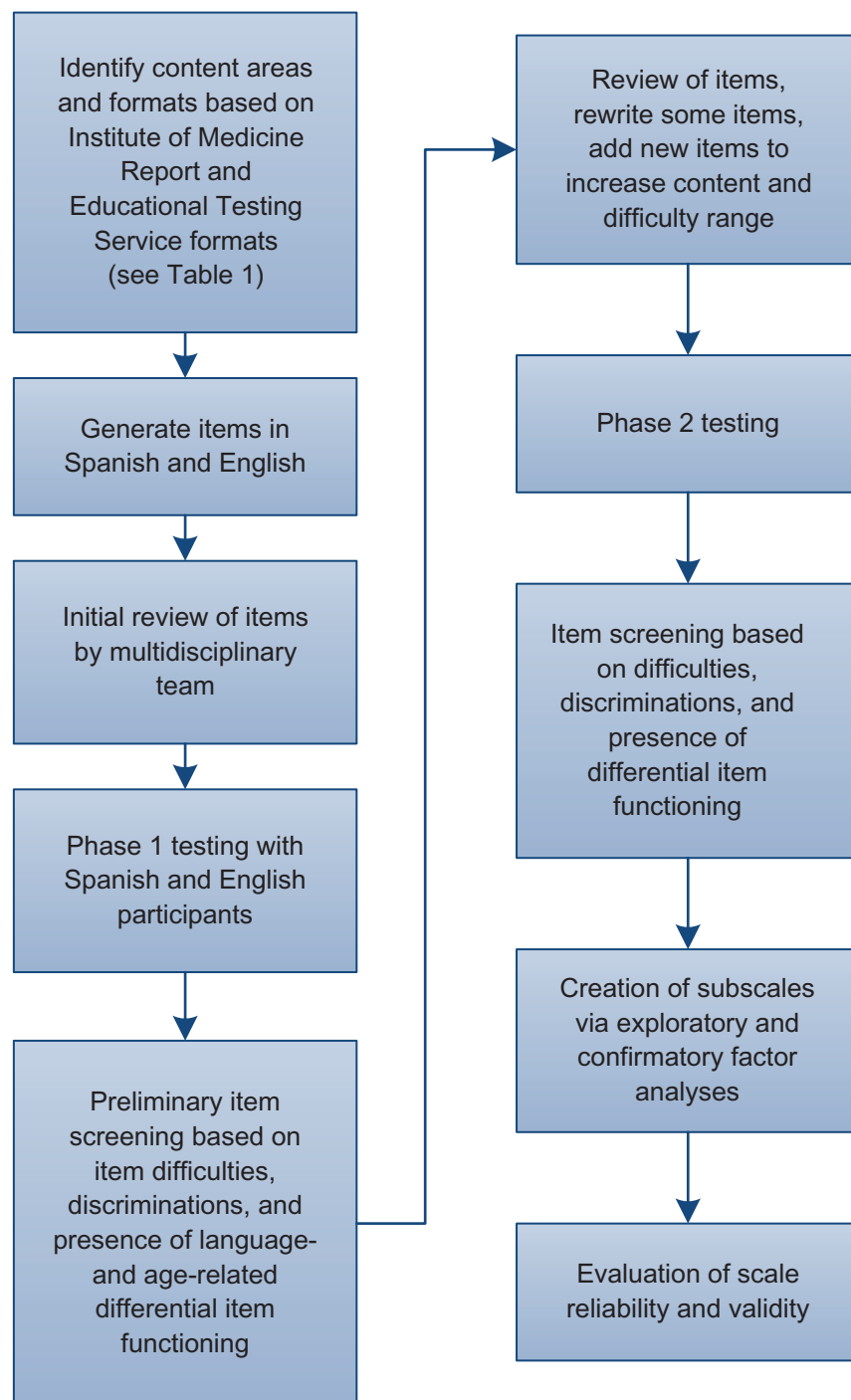
As summarized by Pleasant and McKinney<sup>15</sup> and suggested by Jordan et al,<sup>16</sup> many workers in the field have argued that new measures of health literacy should be developed that broaden the range of content assessed, are based on diverse groups, and have better demonstrated psychometric characteristics. We are currently engaged in the development

of a new measure of health literacy that addresses these criticisms. It samples a wide range of content chosen from the domains listed in the 2004 Institute of Medicine (IOM) report<sup>1</sup> on the competencies needed for adequate health literacy. It is based on a coherent conceptual model of health literacy that has been developed based on our own and others’ research. It includes items that assess prose, document, and quantitative literacies in each of the domains. It has been developed through a rigorous two-stage process in which items have been pilot tested, assessed for equivalence in both Spanish and English as well as in younger and older individuals, and has been subjected to assessments of construct and concurrent validity. The measure assesses, not only reading and quantitative skills, but also uses video simulations of health care–related encounters to assess listening comprehension and to provide test stimuli that bear a close relation to the actual situations in which health literacy skills might be applied. By asking questions that assess expressive writing skills (asking, for example, where certain kinds of information would be placed in a form), the measure also indirectly assesses expressive written language skills. The purpose of this paper is to describe the initial development and testing of this new measure and provide preliminary data on its validity and reliability. The new measure utilizes a broad range of item formats and contents, includes listening comprehension, and has been developed in Spanish and English. The English project has been named Fostering Literacy for Good Health Today (FLIGHT), and the related Spanish project has been named Vive Desarrollando Amplia Salud (VIDAS).

## Methods Overview

This section provides an overview of study procedures (Figure 1). Test items were developed to sample a broad range of health-related content in Spanish and English. The sample on which the measure is validated was purposely drawn from a range of abilities and backgrounds, as evidenced by participants’ occupations and educations. In order to accurately characterize Spanish-speaking participants, we developed a procedure to assess language dominance in Spanish-English bilinguals. Items in both languages were created to minimize the impact of regional usage, and data analyses employ a combination of classical test theory<sup>35–37</sup> and IRT<sup>19</sup> techniques.

In phase 1, a pool of candidate items was administered to Spanish and English speakers, with approximately one-half of each group aged 50 years or older. Items were screened for difficulty and discrimination (correlation with total score)



**Figure 1** Item development and testing process.

and for age- and language-associated DIF. The original pool of items was reduced, and some new items were written to enhance the total scale's range of content and difficulty. In phase 2 (currently in progress, with initial results presented here), the items developed in phase 1 are administered to an age-stratified sample of community-dwelling Spanish and English speakers, along with measures chosen to establish

the new scale's validity. This paper presents information on the first 93 Spanish- and 105 English-speaking participants who have completed phase 2.

### Computer-delivered format

In order to ensure that the resulting measure will be inexpensive and easily deployed, it has been developed

using off-the-shelf touch screen computers that are readily available and reasonably priced (HP TouchSmart®; Hewlett-Packard Development Corporation, Palo Alto, CA, USA). These computers have large touch screens (20 inch diagonal measurement) and run the Windows® operating system (Microsoft Corporation, Redmond, WA, USA) and include self-contained speakers that allow participants to hear all items as they are presented. In another, previous project, we did extensive user testing with the touch screen interface format, iteratively testing the interface, modifying it in response to user comments, and then retesting it.<sup>38</sup> In phase 1, we used the previously developed touch screen format. Interviews conducted after each participant completed phase 1 focused on possible usability and navigation issues as well as on the format and content of the questions.

## Item development

A framework for item development was created, based on the domains of literacy skills needed for health outlined in the 2004 IOM report.<sup>1</sup> For each of the seven health-related goals listed in the report (left column in Table 1), items were

created in one of the three formats commonly used in assessing literacy – prose, document, and quantitative.<sup>39</sup> A 7 × 3 item content matrix was created and used as a guide in item development (Table 1). Candidate items were developed by individual team members and reviewed by the entire team. The team members represent a range of health professions, including medicine, nursing, social work, pharmacy, and psychology. Each team member has extensive experience in clinical work and thus is familiar with the types of clinical problems encountered by patients in obtaining health care. The lead investigator (RO) has extensive experience with psychometric scale development, with multiple publications on the subject. The team includes a psychologist with a strong background in multicultural and multilingual assessment (AA) who has also published multiple articles on psychometric assessment. Other members of the team have had extensive experience in patient education and assessment.

Some items were first created in Spanish and then translated into English, while others were created in English and translated. A guiding principle in item development was to create items that would be culturally and linguistically

**Table 1** Item content examples, by IOM report domains<sup>1</sup> and ETS formats<sup>39</sup>

Goals	Prose	Document	Quantitative
Health promotion	Read a passage on exercise and identify desirable duration of exercise	Make menu choices based on fat and sodium guidelines	Calculate the number of grams of fat in a package of a product given a per serving value
Understand health information	Read a passage on risk factors for diabetes and identify relevant behaviors that would reduce someone's risk	Given a checklist of risk factors for diabetes; be able to complete a checklist of risk factors for the disease	Given information on normal and abnormal blood glucose levels, identify normal and abnormal levels
Apply health information	After being provided with information on physical activity guidelines, identify appropriate exercise duration and frequencies	Given narrative information on exercise frequency and intensity; complete an exercise log	Calculate the number of calories used during exercise given a table of exercises, times, and values; Use Internet based calculator to calculate body mass index
Navigate the health care system	After reading an informational brochure, be able to describe how specific health care services are covered by an insurance program	Review information from a table on dates and times for applying for specific health care benefits	Calculate relative costs of two insurance plans
Participate in encounters with health care professionals	After viewing a video of a person's encounter with a physician providing a new medicine, identify information provided by the physician about dosage and schedule	After viewing a video describing how to apply for long term care insurance, fill out an application	After viewing a video that presents information on desirable weights, calculate one's own body mass index
Give informed consent	After reading information about a colonoscopy, describe the risks and benefits of the procedure	After viewing a video that presents information on informed consent for a clinical study, describe its risks and benefits	Given a graphical representation of the probability of a medication side effect, correctly identify how likely its occurrence will be
Understand rights	After reading an explanation of benefits, correctly identify the procedure to appeal a denial of benefits	Given an insurance explanation of benefits on an insurance payment statement, identify an inappropriate denial	After viewing a video presentation on patient rights, correctly determine the number of options available to access services

**Abbreviations:** IOM, Institute of Medicine; ETS, Educational Testing Service.

equivalent rather than word-for-word translations.<sup>40</sup> From the project's inception, word and item selection have focused on the use of high-frequency words and terms, to ensure that participants would understand all questions. Care was taken to use words, in both languages, that would be understood by persons of varying socioeconomic and educational levels and that were not region- or nation-specific.

Items developed within the 7 × 3 content matrix targeted the component skills of literacy (conceptual knowledge, listening and speaking, writing, reading, and numeracy) as outlined in the IOM report.<sup>1</sup> For example, to assess conceptual knowledge, items that tapped basic health facts were created (eg, "Hemoglobin A1C measures which of the following?"). Listening comprehension was assessed using 60–90 second videos of simulated interactions with health care providers or presentations of health information. For example, one video showed an encounter in which a patient was given a new medication and directions for its use, while another simulated a TV news presentation on finding health information on the Internet. After viewing, participants responded to multiple-choice questions. It was not possible to directly assess participants' oral expression, but questions were created that presented problems that could only be solved by communicating with providers (eg, "Arthur doesn't understand what the doctor says. What can he do?").

Written expression was assessed as "document" literacy, through questions evaluating participants' ability to complete materials, such as insurance forms. "Navigating the health care system" included interpreting hospital maps; some documents and maps included items that asked participants to respond by tapping on the appropriate area of the screen (eg, "Tap on the area where you would find information on how to use toothpaste with a 4-year-old"). Reading comprehension was assessed through questions about passages of varying difficulty levels, and numeracy was assessed through items demanding reading, arithmetic computation, and decision making based on probabilities. Approximately ten items were created for each element in the item content matrix, resulting in 208 candidate items.

## Phase I: initial item testing

This base group of items was administered to 69 Spanish- and 73 English-speaking participants. Language dominance of the Spanish-speaking participants who indicated that they also spoke English was assessed by comparing their performance on the relative proficiency indices (RPI) of the reading and listening comprehension subtests of the Woodcock–Johnson<sup>®</sup> III Diagnostic Reading

Battery (English) and the Woodcock–Muñoz Language Survey<sup>®</sup>–Revised Normative Update (Spanish) psychoeducational batteries (Riverside Publishing, Rolling Meadows, IL, USA). Level of acculturation was assessed using the Marin Acculturation Scale.<sup>41</sup> Most participants showed clear superiority in one language or the other (ie, more than one standard deviation difference in RPI scores), including those who indicated they had proficiency in both languages. Only those participants who showed clear evidence of greater proficiency in Spanish completed the Spanish assessment. The importance of actually assessing Hispanic participants' language skills is underscored by a study that showed that Hispanics who state they are fluent in English may function at lower levels compared with native English speakers.<sup>29</sup>

Almost half of each language group was 50 years of age or older (30 of 69 Spanish and 29 of 73 English speakers), allowing for the assessment of language- and age-related DIF. As discussed above, we believed that this issue was important in light of our finding that almost one-half of the items on the reading comprehension scale of the TOFHLA showed evidence of age-related DIF. After responding to all items, participants completed interviews during which items were reviewed with them for problems in clarity and to assess whether the items actually measured what was intended. Initial analyses were completed using jMetrik ([www.itemanalysis.com](http://www.itemanalysis.com)) to assess item difficulties, discriminations (item-total correlations) and the presence of DIF. The Mantel–Haenszel chi-square statistic, as well as the Educational Testing Service (ETS) grading system,<sup>42</sup> were used to evaluate DIF, supplemented by review of nonparametric item response curves.

In creating the final item pool for further study, some items were eliminated due to language-related DIF, while other items were rewritten after team consultation. No items showed substantial age-related DIF, supporting our decision to avoid using the cloze response procedure used in other measures, as it might bias items against older individuals. A number of items were either low range or midrange in difficulty; many of these were eliminated when their content or format duplicated other items. Data from interviews were used to rewrite items when participants indicated that an item was confusing or when the interviews showed that the item did not actually assess its target skill. Several new items were created in this phase, in order to broaden the range of content covered and provide items with greater difficulties. Although these items were not subjected to the same developmental testing as the others from phase 1, these will be assessed for psychometric characteristics prior to inclusion

in a final measure. This procedure resulted in the 98 items used in phase 2.

## Phase 2: further development and testing

The purpose of this phase is to validate the new health literacy measure by assessing its relations to other measures of health literacy and of participants' health. A purposive sample with a range of abilities (based on education and occupation) is being recruited over specific age ranges. This strategy is likely to be the most efficient approach to obtain optimal item statistics with relatively small samples.<sup>43,44</sup> Interested participants are first screened for cognitive status using the Short Portable Mental Status Questionnaire<sup>45</sup> and paragraphs from the Wechsler Memory Scale,<sup>46</sup> using cutoff scores previously developed in a study of computer use in elderly participants (Czaja, unpublished data, 2012). Participants are also screened for vision and auditory abilities, using a visual screener and auditory comprehension of material presented over headphones calibrated with a handheld decibel meter (Digital Sound Level Meter, model 407730; Extech Instruments, Waltham, MA, USA). Spanish-speaking participants are being recruited from several different national backgrounds, including the countries of Central and South America as well as the USA and Mexico. The language of assessment is determined using the procedure developed in phase 1, using the language preference subscale of the Marin Acculturation Scale<sup>41</sup> supplemented with additional testing when participants indicate significant use of both languages.

In addition to the new health literacy items, participants complete a battery of existing health literacy measures (TOFHLA in both Spanish and English; REALM or Short Assessment of Health Literacy for Spanish-speaking Adults (SAHLSA), and the self-report questions developed by Chew et al<sup>48</sup>) to assess their literacy- and numeracy-related academic skills and basic cognitive abilities. They also provide information on health status, health-related quality of life, and health service utilization. Participants complete assessments in two sessions (individually-administered cognitive and health literacy measures in one, and questionnaires and the health literacy measures administered by touch screen computer in the other) with the order of administration of each session randomly counterbalanced to account for order effects. Because of the length of the assessment sessions, participants can complete both either in a single day (during which they take at least a one-hour break for lunch) or on 2 days. Measures have been selected to allow the evaluation of the relation of the new measure to existing assessments of health literacy; basic cognitive skills; relevant academic skills, such as reading and math skills; and health status variables.

The REALM assesses health literacy based on an individual's ability to read orthographically irregular words in English. As Spanish has few orthographically irregular words, the SAHLSA is administered to Spanish speakers as the closest equivalent. Participants also respond to the self-report health literacy screening questions developed by Chew et al<sup>48</sup> and evaluated in English and Spanish by Sarkar et al.<sup>49</sup>

## Sample

Participants are recruited via flyers, presentations at community organizations, and by recruitment from previous studies. Sampling focuses on recruiting groups of Spanish- and English-speaking participants in the age ranges 18–30, 31–40, 41–50, 51–60, 61–70, 71–80, and 81 years and older. Recruitment is targeted to various socioeconomic, occupational, educational backgrounds (eg, ranging from grade school to doctoral-level graduate education), and in the case of Spanish-speaking participants, to a range of national origins (Central and South America as well as Spain, Mexico, and the USA). Participants are compensated \$80.00 for each completed session. Participants who complete the study in a single day are provided with lunch, and funds are available to reimburse participants for their use of public transportation. All participants have completed both sessions, and none has dropped out of the study between sessions.

## Data analyses

The initial analyses assessed each item's relation to overall ability (item discrimination, evaluated as an item-total correlation greater than 0.20) and the extent to which each was equivalent in Spanish and English. Inspection of nonparametric item response curves, chi-square testing for DIF, and ETS classification, in which items are rated depending on the degree and clinical significance of the DIF,<sup>39</sup> were performed. Further analyses included exploratory and confirmatory factor analyses using Mplus (Muthén and Muthén, Los Angeles, CA, USA) and focused on choosing items that were clearly related to content-defined scales (ie, an item that required an arithmetic computation was related to the numeracy scale). These analyses first established that the new measure reflected more than one factor through results of exploratory factor analyses. The judgment of the number of factors to use was based on inspection of the scree plot of eigenvalues. The multiple factor model suggested by the exploratory analyses was then tested in confirmatory models the adequacy of which were evaluated using standard fit indices. Confirmatory models then evaluated the multifactor model with its separate scales reflecting general health literacy (HL scale), numeracy (NUM scale), listen-



ing comprehension (LIS scale), and conceptual knowledge of health-related facts (FACT scale). The equivalence of the factor model for both Spanish and English speakers was assessed in separate confirmatory factor analyses, for each language group as well as in the combined sample.

## Development of scales

Scales of the measure were developed through a combined process of exploratory and confirmatory factor analysis, and rational scale construction. Factor models were evaluated in both language groups separately, and items that did not load significantly on any factor were eliminated. Four scales were developed to reflect general health literacy; numeracy; listening comprehension, based on responses to video-related items; and conceptual knowledge, based on responses to questions that require only knowledge of specific health-related facts (eg, "Hemoglobin A1C measures which of the following?"). The fit of this four-factor model was evaluated separately for both Spanish- and English-speaking participants as well as in the combined sample.

## Reliability and validity

Cronbach's alpha and correlations of the new measure's scales with the TOFHLA, REALM, SAHLSA, and self-report questions were calculated in order to evaluate the new measure's reliability and concurrent validity. Known-groups validity was assessed by assigning participants to groups based on their total TOFHLA score and evaluating the ability of the general HL scale to differentiate among them.

## Participant satisfaction

After completing the new measure, participants completed a questionnaire based on the Technology Acceptance Model<sup>50</sup> that asked for their ratings of the measure's usefulness, ease of use, enjoyment, and whether they would use it again.

All study procedures were completed under a protocol approved by the Institutional Review Board of Nova Southeastern University. All participants provided written informed consent before engaging in the main study activities; verbal assent was obtained for the completion of initial screening.

## Results

### Phase I

The demographic data for phase 1 are presented in the top portion of Table 2. In phase 1, only data on participant age, race, ethnicity, and education were collected, as the primary purpose of this phase was to screen a larger number of items in order to have a substantial number for use in the validation study

**Table 2** Descriptive statistics for phase 1 and 2 samples

	Phase 1	
	Spanish	English
Gender M/F	22/50	64/50
Hispanic	72	12
African American		51
Afro Caribbean		12
Asian/Pacific Islander		
White	72	51
<b>Continuous variables: means (standard deviations)</b>		
Age	47.0 (14.6)	47.5 (12.7)
Education	14.4 (2.6)	13.5 (2.0)
	Phase 2	
	Spanish	English
Gender M/F	41/52	52/53
Hispanic	93	3
African American		37
Afro Caribbean		14
Asian/Pacific Islander		2
White	93	52
<b>Age group (years)</b>	<b>N</b>	<b>N</b>
<b>Recruitment by age group</b>		
18–30	6	15
31–40	8	15
41–50	22	18
51–60	26	22
61–70	13	13
71–80	12	13
Greater than 80	2	3
<b>Continuous variables: means (standard deviations)</b>		
Age	52.4 (14.7)	50.2 (16.4)
Education	12.7 (2.8)	13.5 (2.0)
TOFHLA reading <sup>a</sup>	42.6 (8.3)	46.0 (4.4)
TOFHLA numeracy <sup>a</sup>	43.7 (6.2)	47.9 (2.8)
REALM	N/A <sup>a</sup>	62.6 (6.6)
SAHLSA	45.8 (3.6)	N/A <sup>a</sup>
Hospital <sup>b</sup>	0.66 (0.89)	0.35 (0.76)
Forms <sup>b</sup>	1.82 (1.20)	2.4 (0.94)
Info <sup>b</sup>	0.65 (0.95)	0.55 (0.92)

**Notes:** <sup>a</sup>Reading measures were only administered to participants in phase 2. The REALM was only administered to English speakers, and the SAHLSA was only administered to Spanish speakers; <sup>b</sup>self-report screening questions (Chew et al):<sup>47</sup> Hospital = the participant needs help reading hospital materials: 0 = never to 4 = always; forms = confident in filling out medical forms: 0 = not at all to 3 = quite a bit; info = difficulty in understanding written medical information: 0 = never to 3 = always.

**Abbreviations:** REALM, Rapid Estimate of Adult Literacy in Medicine; SAHLSA, Short Assessment of Health Literacy for Spanish-speaking Adults; TOFHLA, Test of Functional Health Literacy in Adults; N/A, not applicable.

in phase 2. The 208 items administered in this phase were evaluated to assess participant difficulties, discriminations, and both age- and language-related DIF. Of the 208 items, 113 items were deleted due to low discrimination (ie, low relation to overall health literacy), presence of language-related DIF, or a combination of low difficulty and redundant content. When several items tapped similar content or skill and had similar levels of difficulty, items were chosen according to the more nearly unique content or for greater difficulty.

## Phase 2

The demographic data for the samples (93 Spanish- and 105 English-speaking participants) are presented in Table 2. All Spanish-speaking participants described their racial background as white, consistent with findings from the US Census,<sup>51</sup> while the English-speaking participants were African American, Afro Caribbean, white, and Asian. Consistent with our sampling strategy, the average age of our participants was approximately 50, and the average level of education was equivalent to having completed high school. Our actual recruitment, by target age groups, is also included in this table.

## Scale development

Initial exploratory factor analyses suggested the presence of more than one factor underlying our data. Scales (HL, NUM, LIS, or FACT) were created based on results of the preliminary factor analysis, and item content and format, eliminating items with low relations (item loadings) to any of the scales. Confirmatory analyses indicated that some items from the NUM, LIS, and FACT scales were also related to general health literacy, and these were included in the HL as well as the other scales. The fit of the resulting factor models was assessed in confirmatory analyses that evaluated progressively more complex models (encompassing first HL, then adding NUM, then FACT, and then LIS). Fit for the final four-factor model in the combined sample was good, with a nonsignificant chi-square value ( $\chi^2$  [df = 2,676] = 2759.22) ( $P = 0.13$ ), a root mean square error of approximation (RMSEA) value of 0.01, and a confirmatory fit index value of 0.97.

Factor scores for each factor in the combined group were calculated and used in analyses of correlations between the new measure's scales and other measures of health literacy. After scale development, multiple indicators and multiple causes (MIMIC)<sup>52</sup> models evaluated whether any items continued to show language-related DIF; no remaining items showed significant DIF. The final scales are listed in Table 3, along with example item objectives and measures of scale reliability for each language group as well as for the entire sample. The HL and NUM scales have acceptable reliabilities, but the FACT and LIS scale have borderline to low reliabilities. This suggests that further work is needed in developing these scales, although we note that some items from each scale contribute to the HL scale, which has an acceptable reliability. We suggest that these scales be regarded as experimental, pending further development.

The scale intercorrelations and correlations with other measures of health literacy are presented in Table 4. The new measure's scales are significantly intercorrelated, reflecting the common composition of HL with the other scales (some items were common to HL and the other scales) but not correlations among other scales (NUM, FACT, and LIS), which were not influenced by common items (ie, no items loaded on pairs of these factors). Thirteen items from the FACT scale and eleven items from LIS are included in the HL scale. This strategy was employed in order to allow us to develop a broad general health literacy scale based on the items that were most closely related to overall health literacy, as suggested by factor analyses.

The new measure's scales are also related to other measures of health literacy. The HL scale correlated significantly

**Table 3** Scale descriptions, Cronbach's alpha, and examples

Scale	Examples
General health literacy (HL): the ability to read and complete mental operations on health care information, including identify relevant information in prose, documents, and figures (39 items). Cronbach's $\alpha$ for Spanish speakers = 0.81; for English = 0.84; for entire sample = 0.84.	Prose: after reading instructions for laboratory test preparation, correctly identify appointment time. Document: correctly identify fields in an insurance form; use an electronic device on a web page to calculate body mass index.
Numeracy (NUM): the application of quantitative skills, including arithmetic operations and appraisal of relations, among numeric concepts, such as ratios and percentages (24 items). Cronbach's $\alpha$ for Spanish speakers = 0.83; for English = 0.82; for entire sample = 0.84.	Quantitative: correctly identify meaning of terms related to probability; correctly identify number of grams of fat consumed in a meal based on values in a table.
Conceptual knowledge (experimental scale; FACT): demonstrate understanding of specific concepts related to health care (15 items). Cronbach's $\alpha$ for Spanish speakers = 0.58; for English = 0.72; for entire sample = 0.67.	Correctly identify the organ treated by a medical specialist, such as a cardiologist.
Listening comprehension (experimental scale; LIS): the ability to acquire and remember information presented orally (13 items). Cronbach's $\alpha$ for Spanish speakers = 0.56; for English = 0.60; for entire sample = 0.58.	After viewing a video of clinician giving information about participation in a clinical research study, correctly identify treatment alternatives.

**Table 4** Scale intercorrelations and correlations of the new measure with other measures of health literacy

	HL	NUM <sup>a</sup>	FACT <sup>a</sup>	LIS <sup>a</sup>	TOFHLA reading	TOFHLA numeracy	REALM <sup>b</sup>	SAHLSA <sup>b</sup>	Hospital <sup>c</sup>	Forms <sup>c</sup>	Info <sup>c</sup>
HL	1.00	0.99**	0.67**	0.80**	0.62**	0.34**	0.46**	0.48**	-0.15*	0.38**	-0.24**
NUM		1.00	0.67**	0.77**	0.62**	0.35**	0.48**	0.46**	-0.16*	0.38**	-0.24**
FACT			1.00	0.81**	0.42**	0.26**	0.44**	0.56**	-0.11	0.28**	-0.19*
LIS				1.00	0.53**	0.20**	0.34**	0.62**	-0.05	0.32**	-0.17*
TOFHLA reading					1.00	0.29**	0.69**	0.57**	-0.13	0.40**	-0.25**
TOFHLA numeracy						1.00	0.24*	0.17	-0.14	0.16*	-0.12
REALM <sup>b</sup>							1.00	n/a <sup>b</sup>	-0.28**	0.23*	-0.23*
SAHLSA <sup>b</sup>								1.00	0.10	0.16	-0.25*
Hospital <sup>c</sup>									1.00	-0.29**	0.34**
Forms <sup>c</sup>										1.00	-0.31**
Info <sup>c</sup>											1.00

**Notes:** \*Correlation is significant at the 0.05 level (two-tailed); \*\*correlation is significant at the 0.01 level (two-tailed); <sup>a</sup>some items on NUM, FACT, and LIS were also included in the HL scale, resulting in higher scale intercorrelations; <sup>b</sup>the REALM was only administered to English speakers, and the SAHLSA was only administered to Spanish speakers; <sup>c</sup>hospital = need help reading hospital materials; Forms = confident in filling out medical forms; Info = difficulty in understanding written medical information.

**Abbreviations:** REALM, Rapid Estimate of Adult Literacy in Medicine; SAHLSA, Short Assessment of Health Literacy for Spanish-speaking Adults; TOFHLA, Test of Functional Health Literacy in Adults; NUM, FLIGHT/VIDAS numeracy; FACT, FLIGHT/VIDAS conceptual knowledge; LIS, FLIGHT/VIDAS listening comprehension; HL, FLIGHT/VIDAS general health literacy.

( $P < 0.01$ ) with the TOFHLA reading and numeracy scales, the REALM, the SAHLSA, and the self-report items. Each of the remaining scales was correlated significantly with the TOFHLA, REALM, and SAHLSA. That the FACT and LIS scales are significantly correlated with other measures (TOFHLA, REALM, and the SAHLSA) suggests that the constructs they measure (health-related knowledge and listening comprehension) are related to health literacy in spite of these scales' low reliability. The correlations of the new measure's scales with existing measures of health literacy are similar to the correlations found between the existing measures, suggesting that the new measure has substantial concurrent validity.

### Known-groups validity

In order to further assess the usefulness of the HL scale, we evaluated its ability to differentiate among participants assigned to groups based on their scores on the TOFHLA. We grouped participants in three groups: those with total TOFHLA scores less than 91 (of 100), those with total scores between 91 and 95, and those with scores greater than 95. Analysis of variance (ANOVA) models were assessed separately for English- and Spanish-speaking participants.

Of the English-speaking participants, 19 were in the group with scores less than 91, 29 were in the group with scores between 91 and 95, and 53 were in the group with scores greater than 95. The ANOVA model for English speakers showed that the HL scale scores differed significantly across the groups ( $F [2,101] = 23.74$ ) ( $P < 0.001$ ). Post hoc analyses showed that all groups differed significantly from each other (all  $P < 0.01$ ).

For Spanish-speaking participants, 53 participants were in the group with scores less than 91, 22 were in the group

with scores from 91 to 95, and 17 were in the group with scores greater than 95. The model for this group also showed that the HL scale scores differed significantly across groups ( $F [2,92] = 24.72$ ) ( $P < 0.001$ ). Post-hoc analyses again showed that all groups differed significantly from each other (all  $P < 0.01$ ).

### Participant satisfaction

The average of the participant ratings of the measure's usefulness, on a scale from 0 to 6, was 5.03 (standard deviation [SD] = 1.25). Their mean rating of how easy it was to use was 5.37 (SD = 0.95). Participant ratings of enjoyment in using the measure were also positive (5.00 [SD = 1.03]) as were their ratings on items asking if they would use the measure again (5.08 [SD = 0.73]). Participant ratings of usefulness, ease of use, enjoyment, or intent to use in the future were not related to the language in which they completed the measure (in  $t$ -tests comparing language groups, all  $P > 0.10$ ).

### Discussion

The purpose of this study has been to develop and validate a computer-administered measure of health literacy. In this study, health literacy is defined more broadly than in other measures, to encompass the domains of content and skills outlined in the 2004 IOM report,<sup>1</sup> while reflecting the literacy formats of other measures.<sup>39,53</sup> Our approach has focused on creating a measure that assesses as broad a range of health literacy skills as possible within the constraints of computer administration and scoring. The resulting measure is thus able to evaluate participants' comprehension of written language, understanding of health-related documents, and ability to use quantitative skills in performing health-related tasks and

making probability-based judgments. By including video vignettes of health-related situations, such as an encounter with a provider giving instructions for a new medication, it also assesses listening comprehension. The new HL scale includes items related to finding health information on the Internet, making it one of the first measures to evaluate these skills. Through the development process described in this paper, it has been possible to create a multidimensional assessment instrument that has significant relations to other measures, suggesting that it has concurrent validity. Participants' reactions to using the measure have been uniformly positive, showing that it may be acceptable for more general use.

Our central purpose in developing FLIGHT/VIDAS has been to develop a measure of health literacy that would address the criticisms of existing measures, as summarized in several reviews.<sup>16,23</sup> Core issues identified by researchers have been the range of content assessed by measures; their psychometric characteristics, such as reliability and validity; their relation to actual health behaviors; the lack of measures that are equivalent in both Spanish and English; and lack of basis in an actual theory or model of health literacy, with the apparent result that different measures of "health literacy" may actually assess different things.<sup>27</sup> In the FLIGHT/VIDAS project, we have addressed each of these issues by developing and evaluating a model of health literacy; by employing rigorous methods and testing in item and scale development, while creating items with a broad range of content and response formats, including video simulations of actual health encounters; by developing the measure simultaneously in both Spanish and English; and by including a range of measures for use in validating the new measure.

Some time ago, Baker<sup>14</sup> noted that it would be helpful to have a measure of conceptual knowledge of health and illness. In this project, we have attempted to develop just such a measure in the FACT scale that assesses health-related conceptual knowledge. Analyses suggest that this initial attempt has been partially successful. The FACT scale's reliability is lower than may be desirable, but it is significantly correlated with other measures of health literacy. Because of its clear content validity, it may be useful for research in which the role of cognitive skills and disease-related knowledge in health literacy are examined. Chin et al<sup>54</sup> for example, used this strategy. Understanding the relation of specific disease knowledge to health literacy may facilitate the development of measures in which general skills, such as reading or numeracy, are integrated with disease-related knowledge. Researchers have successfully employed this

strategy in developing disease-specific health literacy measures,<sup>55</sup> and it may also be useful in guiding the development of interventions.

The model of health literacy underlying FLIGHT/VIDAS explicitly includes conceptual knowledge as an aspect of health literacy. While it may be possible to separate conceptual knowledge from other health literacy skills, our decision to include health-related conceptual knowledge is grounded in research showing that basic knowledge interacts with reading or listening skills in producing competent performances, whether in basic reading among school children<sup>56,57</sup> or in complex activities, such as sight reading piano music<sup>58</sup> and playing chess.<sup>59</sup> Other research has supported this view in studying health literacy.<sup>54</sup> Implicit in this view of health literacy is the belief that it reflects the ability to carry out complex activities that may include using existing information and skills to solve new problems.<sup>60</sup> Assessing core conceptual knowledge related to health care, from this point of view, is essential to understanding the ways in which health literacy affects the ability to obtain and use health information.

This development process is similar to that taken by others in creating new measures of health literacy. Using IRT techniques and sampling Spanish and English speakers, Hahn et al,<sup>24</sup> Yost et al,<sup>26</sup> and Lee et al<sup>25</sup> have developed measures of health literacy that can be used in both languages and which have clearly defined psychometric properties. The differences in our approach compared with the Health LiTT measure described by Hahn et al include the response formats used (our measure does not use the cloze procedure, as it may have a differential impact on older persons), our focus on the equivalence of items in both languages (the Health LiTT prose passages are different and nonequivalent in the Spanish and English versions), and the use of a sample drawn from the community rather than from clinics. For the new scale, we developed a method to assess participants' language dominance (the language in which they clearly performed better) and assessed participants in that language so that those who completed the Spanish language version of the measure are clearly characterized as to linguistic competence and level of acculturation. As reported by Aguirre et al<sup>28</sup> and Zun et al,<sup>29</sup> Hispanics who complete assessments in English may be at a disadvantage to non-Hispanics whose native language is English, even when they appear to be able to speak and understand English. As there are wide variations in the cultural and linguistic backgrounds of individuals from Spanish-speaking countries who live in the US,<sup>61</sup> the assessment of the Spanish-language items with individuals from a number of different

Spanish-speaking countries may be an important aspect of making the measure more generally useful.

While the Health LiTT may be more useful in assessing whether patients have deficiencies in health literacy (for example, the authors indicate that they focused mainly on items in the range of sixth- to eleventh-grade-equivalent difficulties [Yost et al<sup>26</sup>]), our measure may be more relevant to understanding health literacy in the general population, especially in normal elderly (because the new measure does not use the cloze procedure) and in younger and healthier individuals (because they have been included in our development and validation samples).

The instrument developed by Lee et al<sup>25</sup> may be useful for persons interested in assessing patients' health literacy as their ability to recognize health-related words and link those words to their meanings. This measure is brief and may be used for screening, but it assesses a smaller range of health literacy skills (notably, it does not include numeracy) and is neither computer administered nor scored. Each of these measures, including the one reported upon here, have strengths and weaknesses. We suggest that the advantages of the new measure are that it assesses a broad range of skills, does not use the cloze procedure, and includes items that are equivalent in both Spanish and English.

The limitations of our approach include the indirect assessment of oral expression in the measure (which is only measured by asking participants to imagine their responses to problematic encounters with providers and does not constitute a separate scale); the current modest sample size; and the need to administer the measure by means of computers. While computers are available in many clinical and research settings, they may not be available in many other settings in which health literacy might be assessed. It may be useful to develop a measure that can be administered on more convenient portable devices, such as the iPad<sup>®</sup> (Apple Inc, Cupertino, CA, USA), and our long-term development plans include developing such a version. Another important limitation of the FACT and LIS scales are their low reliability in one or both language groups. Because of their clear content validity and relation to other measures of health literacy, these scales may be useful for research purposes. They should be regarded as experimental and potentially useful pending further development.

The purpose of this study is thus to develop and validate an innovative computer-administered measure of health literacy that assesses a broad range of health literacy skills, within the constraints of a computer-administered measure. It is anticipated that by expanding the range of

skills assessed, the new measure may have wider usefulness and better ability to predict important health-related outcomes. Two new scales in this measure, the FACT and the LIS, may provide information useful in developing a broader understanding of which skills are most important in particular contexts; for example, listening comprehension might be important to patients' ability to benefit from typical health encounters that rely heavily on clinicians' oral communication of information, while written and numeracy skills might be related to patients' success at following written treatment guidelines or becoming involved in health promotion activities. The new measure's inclusion of items related to health information on the Internet may also be a useful addition to health literacy assessment that merits further exploration as technology becomes more and more important in health care. Given the new measure's relation to existing measures, we believe that we have created a useful and valid measure of health literacy skills. Data collection in phase 2 is continuing and will be completed in 2014. We anticipate being able to make the measure available shortly thereafter.

## Acknowledgment

The study described in this paper was supported by a grant (number R01HL096578) to Dr Ownby, from the National Heart, Lung, and Blood Institute.

## Disclosure

The authors report they have no conflicts of interest to disclose.

## References

1. Committee on Health Literacy Board on Neuroscience and Behavioral Health; Nielsen-Bohman L, Panzer AM, Kindig DA, editors. *Health Literacy: A Prescription to End Confusion*. Washington DC, National Academies Press; 2004.
2. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. *Ann Intern Med*. 2011;155(2):97–107.
3. Dewalt DA, Berkman ND, Sheridan S, Lohr KN, Pignone MP. Literacy and health outcomes: a systematic review of the literature. *J Gen Intern Med*. 2004;19(1):1228–1239.
4. Bostock S, Steptoe A. Association between low functional health literacy and mortality in older adults: longitudinal cohort study. *BMJ*. 2012;344:e1602.
5. Sudore RL, Yaffe K, Satterfield S, et al. Limited literacy and mortality in the elderly: the health, aging, and body composition study. *J Gen Intern Med*. 2006;21(8):806–812.
6. Ownby RL, Waldrop-Valverde D, Caballero J, Jacobs RJ. Baseline medication adherence and response to an electronically delivered health literacy intervention targeting adherence. *Neurobehav HIV Med*. 2012;4: 113–121.
7. Sheridan SL, Halpern DJ, Viera AJ, Berkman ND, Donahue KE, Crotty K. Interventions for individuals with low health literacy: a systematic review. *J Health Commun*. 2011;16 Suppl 3:S30–S54.

8. Osborn CY, Cavanaugh K, Wallston KA, et al. Health literacy explains racial disparities in diabetes medication adherence. *J Health Commun.* 2011;16 Suppl 3:S268–S278.
9. Waldrop-Valverde D, Osborn CY, Rodriguez A, Rothman RL, Kumar M, Jones DL. Numeracy skills explain racial differences in HIV medication management. *AIDS Behav.* 2010;14(4):799–806.
10. 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(10):537–541.
11. Murphy PW, Davis TC, Long SW, Jackson RH, Decker BC. Rapid estimate of adult literacy in medicine (REALM): A quick reading test for patients. *Journal of Reading.* 1993;37(2):124–130.
12. Weiss BD, Mays MZ, Martz W, et al. Quick assessment of literacy in primary care: the newest vital sign. *Ann Fam Med.* 2005;3(6):514–522.
13. 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(1):33–42.
14. Baker DW. The meaning and the measure of health literacy. *J Gen Intern Med.* 2006;21(8):878–883.
15. Pleasant A, McKinney J. Coming to consensus on health literacy measurement: an online discussion and consensus-gauging process. *Nurs Outlook.* 2011;59(2):95–106. e1.
16. Jordan JE, Osborne RH, Buchbinder R. Critical appraisal of health literacy indices revealed variable underlying constructs, narrow content and psychometric weaknesses. *J Clin Epidemiol.* 2011;64(4):366–379.
17. Ackerman PL, Beier ME, Bowen KR. Explorations of crystallized intelligence: Completion tests, cloze tests, and knowledge. *Learn Individ Differ.* 2000;12(1):105–121.
18. Ackerman PL, Cianciolo AT. Cognitive, perceptual-speed, and psychomotor determinants of individual differences during skill acquisition. *J Exp Psychol Appl.* 2000;6(4):259–290.
19. Embretson SE, Reise SP. *Item Response Theory for Psychologists.* Mahwah, NJ: Lawrence Erlbaum Associates; 2000.
20. Noar SM, Benac CN, Harris MS. Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions. *Psychol Bull.* 2007;133(4):673–693.
21. Ownby RL, Hertzog C, Czaja SJ. Tailored information and automated reminding to improve medication adherence in Spanish- and English-speaking elders treated for memory impairment. *Clin Gerontol.* 2012;35(3):221–238.
22. Jerant A, Sohler N, Fiscella K, Franks B, Franks P. Tailored interactive multimedia computer programs to reduce health disparities: opportunities and challenges. *Patient Educ Couns.* 2011;85(2):323–330.
23. Pleasant A, McKinney J, Rikard RV. Health literacy measurement: a proposed research agenda. *J Health Commun.* 2011;16 Suppl 3: S11–S21.
24. Hahn EA, Choi SW, Griffith JW, Yost KJ, Baker DW. Health literacy assessment using talking touchscreen technology (Health LiTT): a new item response theory-based measure of health literacy. *J Health Commun.* 2011;16 Suppl 3:S150–S162.
25. Lee SY, Stucky BD, Lee JY, Rozier RG, Bender DE. Short Assessment of Health Literacy-Spanish and English: a comparable test of health literacy for Spanish and English speakers. *Health Serv Res.* 2010;45(4):1105–1120.
26. Yost KJ, Webster K, Baker DW, Choi SW, Bode RK, Hahn EA. Bilingual health literacy assessment using the Talking Touchscreen/ la Pantalla Parlanchina: Development and pilot testing. *Patient Educ Couns.* 2009;75(3):295–301.
27. Haun J, Luther S, Dodd V, Donaldson P. Measurement variation across health literacy assessments: implications for assessment selection in research and practice. *J Health Commun.* 2012;17 Suppl 3:S141–S159.
28. Aguirre AC, Ebrahim N, Shea JA. Performance of the English and Spanish S-TOFHLA among publicly insured Medicaid and Medicare patients. *Patient Educ Couns.* 2005;56(3):332–339.
29. Zun LS, Sadoun T, Downey L. English-language competency of self-declared English-speaking Hispanic patients using written tests of health literacy. *J Natl Med Assoc.* 2006;98(6):912–917.
30. McCormack L, Bann C, Squiers L, et al. Measuring health literacy: a pilot study of a new skills-based instrument. *J Health Commun.* 2010;15 Suppl 2:S51–S71.
31. <http://www.rti.org> [homepage on the Internet]. Health literacy skills instrument: *RTI International*, Research Triangle Park, NC. Available from: <http://www.rti.org/page.cfm?objectId=66F893E4-5056-B100-100C834F234F368198>. Accessed May 20, 2013.
32. Bann CM, McCormack LA, Berkman ND, Squiers LB. The Health Literacy Skills Instrument: a 10-item short form. *J Health Commun.* 2012;17 Suppl 3:S191–S202.
33. Ownby RL. Technology and psychogeriatrics: Health and computer literacy. *IPA Bulletin.* 2011;28(3):5–6.
34. Ownby RL, Czaja SJ. Healthcare website design for the elderly: improving usability. *AMIA Annu Symp Proc.* 2013;2003:960.
35. DeVellis RF. Classical test theory. *Med Care.* 2006;44(11 Suppl 3): S50–S59.
36. Gulliksen H. *Theory of Mental Tests.* New York, NY: Wiley & Sons; 1950.
37. Lord FM, Novick MR. *Statistical Theories of Mental Test Scores.* Boston, MA: Addison-Wesley Publishing; 1968.
38. Ownby RL, Waldrop-Valverde D, Jacobs RJ, Acevedo A, Caballero J. Cost effectiveness of a computer-delivered intervention to improve HIV medication adherence. *BMC Med Inform Decis Mak.* 2013; 13:29.
39. Kirsch IS. *The International Adult Literacy Survey (IALS): Understanding What was Measured.* Princeton, NJ: Educational Testing Service; 2001.
40. Acevedo A, Krueger KR, Navarro E, et al. The Spanish translation and adaptation of the Uniform Data Set of the National Institute on Aging Alzheimer's Disease Centers. *Alzheimer Dis Assoc Disord.* 2009;23(2): 102–109.
41. Marin G, Sabogal F, Marin BV, Otero-Sabogal R, Perez-Stable EJ. Development of a short acculturation scale for Hispanics. *Hispanic Journal of Behavioral Sciences.* 1987;9:183–205.
42. Zwirk R. A Review of ETS Differential Item Functioning Assessment Procedures: Flagging Rules, Minimum Sample Size Requirements, and Criterion Refinement. Princeton, NJ: Educational Testing Service; 2012. Available from: <http://outcomes.cancer.gov/conference/irt/orlando.pdf>. Accessed July 10, 2013.
43. Orlando M. Critical issues to address when applying item response theory (IRT) models. Paper presented at: Advances in Health Outcomes Measurement: Exploring the Current State and the Future of Item Response Theory, Item Banks, and Computer-Adaptive Testing; June 23–25, 2004; Bethesda, MD.
44. Wingersky MS, Lord FM. An investigation of methods for reducing sampling error in certain IRT procedures. *Appl Psychol Meas.* 1984;8(3):347–364.
45. Pfeiffer E. A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients. *J Am Geriatr Soc.* 1975;23(10):433–441.
46. Wechsler D. *Manual for the Wechsler Memory Scale-Fourth Edition.* San Antonio, TX: Pearson Education Inc; 2008.
47. Lee SY, Bender DE, Ruiz RE, Cho YI (2006). Development of an easy-to-use Spanish health literacy test. *Health Serv Res.* 41, 1391–1412.
48. Chew LD, Griffin JM, Partin MR, et al. Validation of screening questions for limited health literacy in a large VA outpatient population. *J Gen Intern Med.* 2008;23(5):561–566.
49. Sarkar U, Schillinger D, López A, Sudore R. Validation of self-reported health literacy questions among diverse English and Spanish-speaking populations. *J Gen Intern Med.* 2011;26(3):265–271.
50. Venkatesh V. Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the Technology Acceptance Model. *Information Systems Research.* 2000;11(4):342–365.
51. Ennis SR, Rios-Vargas M, Albert NG. *The Hispanic Population: 2010.* Washington, DC: US Department of Commerce, US Census Bureau; 2011. Available from: <http://www.census.gov/prod/cen2010/briefs/c2010br-04.pdf>. Accessed July 20, 2013.

52. Fleishman JA. Using MIMIC models to assess the influence of differential item functioning. Paper presented at: Advances in Health Outcomes Measurement: Exploring the Current State and the Future of Item Response Theory, Item Banks, and Computer-Adaptive Testing; June 23–25, 2004; Bethesda, MD. Available from: <http://outcomes.cancer.gov/conference/irt/fleishman.pdf>. Accessed on July 10, 2013.
53. White S, Dillow S. *Key Concepts and Features of the 2003 National Assessment of Adult Literacy*. Washington DC: US Department of Education, National Center for Educational Statistics; 2005.
54. Chin J, Morrow DG, Stine-Morrow EA, Conner-Garcia T, Graulich JF, Murray MD. The process-knowledge model of health literacy: evidence from a componential analysis of two commonly used measures. *J Health Commun*. 2011;16 Suppl 3:S222–S241.
55. Ownby RL, Waldrop-Valverde D, Hardigan P, Caballero J, Jacobs R, Acevedo A. Development and validation of a brief computer-administered HIV-Related Health Literacy Scale (HIV-HL). *AIDS Behav*. 2013;17(2):710–718.
56. Beck IL, Perfetti CA, McKeown MG. Effects of long-term vocabulary instruction on lexical access and reading comprehension. *J Educ Psychol*. 1982;74(4):506–521.
57. Nagy WE. *Teaching Vocabulary to Improve Reading Comprehension*. National Council of Teachers of English, Urbana, IL 1988.
58. Meinz EJ, Hambrick DZ. Deliberate practice is necessary but not sufficient to explain individual differences in piano sight-reading skill: the role of working memory capacity. *Psychol Sci*. 2010;21(7):914–919.
59. Hambrick DZ, Meinz EJ. Limits on the predictive power of domain-specific experience and knowledge in skilled performance. *Curr Dir Psychol Sci*. 2011;20(10):275–279.
60. Ownby RL, Waldrop-Valverde D. Health literacy is related to problem solving. Paper presented at: Health Literacy Research Conference; October 19–20, 2009; Washington, DC. Available from : <http://www.bumc.bu.edu/healthliteracyconference/files/2010/04/Abstracts-Presented-at-the-2009-Conference.pdf>. Accessed on July 10, 2013.
61. Motel S, Patten E. The 10 largest Hispanic Origin Groups: Characteristics, Rankings, Top Counties. Washington, DC: Pew Research Center; 2012.

## Patient Related Outcome Measures

### Publish your work in this journal

Patient Related Outcome Measures is an international, peer-reviewed, open access journal focusing on treatment outcomes specifically relevant to patients. All aspects of patient care are addressed within the journal and practitioners from all disciplines are invited to submit their work as well as healthcare researchers and patient support groups.

Submit your manuscript here: <http://www.dovepress.com/patient-related-outcome-measures-journal>

Dovepress

The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.