DOI: 10.1111/j.1475-6773.2006.00532.x

# Development of an Easy-to-Use Spanish Health Literacy Test

*Shoou-Yih D. Lee, Deborah E. Bender, Rafael E. Ruiz, and Young Ik Cho*

**Objective.** The study was intended to develop and validate a health literacy test, termed the *Short Assessment of Health Literacy for Spanish-speaking Adults* (SAHLSA), for the Spanish-speaking population.

**Study Design.** The design of SAHLSA was based on the *Rapid Estimate of Adult Literacy in Medicine* (REALM), known as the most easily administered tool for assessing health literacy in English. In addition to the word recognition test in REALM, SAHLSA incorporates a comprehension test using multiple-choice questions designed by an expert panel.

**Data Collection.** Validation of SAHLSA involved testing and comparing the tool with other health literacy instruments in a sample of 201 Spanish-speaking and 202 English-speaking subjects recruited from the Ambulatory Care Center at UNC Health Care.

**Principal Findings.** With only the word recognition test, REALM could not differentiate the level of health literacy in Spanish. The SAHLSA significantly improved the differentiation. Item response theory analysis was performed to calibrate the SAHLSA and reduce the instrument to 50 items. The resulting instrument, SAHLSA-50, was correlated with the *Test of Functional Health Literacy in Adults*, another health literacy instrument, at $r = 0.65$. The SAHLSA-50 score was significantly and positively associated with the physical health status of Spanish-speaking subjects ($p < .05$), holding constant age and years of education. The instrument displayed good internal reliability (Cronbach's $\alpha = 0.92$) and test–retest reliability (Pearson's $r = 0.86$).

**Conclusions.** The new instrument, SAHLSA-50, has good reliability and validity. It could be used in the clinical or community setting to screen for low health literacy among Spanish speakers.

**Key Words.** Health literacy, test instrument, Spanish speakers, SAHLSA

Health literacy means "the degree to which individuals have the capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions" (Selden et al. 2000). The past several

years have witnessed a growing societal concern with low health literacy and its health toll (Lee, Arozullah, and Cho 2004). *Healthy People 2010 Objectives: Draft for the Public Comment* notes that health literacy is "increasingly vital to help people critically evaluate health information" (U.S. Department of Health and Human Services 1998). Indeed, empirical research has linked low health literacy to low health status (Baker et al. 1997, 1998; Williams et al. 1998; Gazmararian et al. 1999) and poorer knowledge about disease prevention and treatment (Kalichman et al. 2000). Preliminary analyses have also suggested that low health literacy may threaten care quality and cause unnecessary hospital costs (Baker et al. 1997; Marwick 1997; Roter, Rudd, and Comings 1998; Rudd, Moeykens, and Colton 2000). A study of Medicaid participants revealed that those who read at the lowest grade levels (grades 0–2) had average annual health care costs of $12,974 compared with $2,969 for the overall sample studied (Weiss et al. 1994).

The adverse health effects of low health literacy may be more pronounced in the Spanish-speaking Latino population, the fastest growing ethnic group in the United States. A census report in 2002 indicated that Latinos surpassed African Americans as the largest minority group in the country for the first time (U.S. Census Bureau 2002). By the mid-2000s, Latinos are expected to comprise more than 20 percent of the nation's population (U.S. Census Bureau 2000). Latinos tend to trail the rest of the U.S. population in terms of educational attainment (del Pinal 1996). The 2000 U.S. Census showed that the proportion of Latinos aged 25 years and older who had less than a fifth grade education was about 14 times greater than that of non-Latino whites. Such a low literacy level, coupled with the language barrier, may increase the vulnerability of the Latino population to multiple health threats and limit the access of the population to health insurance and basic, needed health care (Ruiz 1993, 1995; Schur and Albers 1996; Ammons 1997; Weinick and Krauss 2000).

A recent study found that Latinos seen at Latino mental health clinics reported more difficulties managing medications than comparison groups,

---

Address correspondence to Shoou-Yih D. Lee, Ph.D., Department of Health Policy and Administration, School of Public Health, University of North Carolina at Chapel Hill, 1101 McGavran-Greenberg Hall (CB# 7411), Chapel Hill, NC 27599-7411. Deborah E. Bender, Ph.D., M.P.H. and Rafael E. Ruiz, Sc.M., are with the Department of Health Policy and Administration, School of Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC. Young Ik Cho, Ph.D., is with the Survey Research Laboratory (MC 336), University of Illinois at Chicago, Chicago, IL.

indicating that the problem with medications was not adequately addressed by receiving services in a culturally sensitive clinic (Diaz et al. 2001). The authors suggested the problem was likely because of the persistent effects of language barriers and limited literacy ability to function in the health care arena. Thus, the adverse effects of low health literacy observed in the English-speaking population may be more salient among Spanish-speaking Latinos, even when they are offered culturally sensitive care.

## Assessment of Health Literacy in Spanish

Without accurate assessment, efforts to understand and ameliorate the health problems associated with low health literacy among Spanish-speaking Latinos are unlikely to succeed. Identifying individuals with low health literacy is difficult because information such as age, educational attainment (i.e., years of schooling), and self-reported literacy skills do not reliably reflect the individuals' health literacy level (Davis et al. 1993, 1996; Nurss et al. 1997; Bass et al. 2002).[1] Thus, health professionals working with the Spanish-speaking population have long expressed the need for an easy-to-use test of health literacy (Davis et al. 2004).

Two instruments—the *Test of Functional Health Literacy in Adults* (TOFHLA) and the *Instrument for Diagnosis of Reading* (a.k.a. the *Instrumento Para Diagnosticar Lecturas* or IDL)—have been translated into Spanish and used in medical research settings (Davis et al. 1998). Both are comprehension tests (Table 1). The Spanish version of TOFHLA, called TOFHLA-S, has good internal consistency, reliability, and content validity. However, like the English version, it takes at least 20 minutes to administer (Parker et al. 1995). The IDL is a Spanish reading assessment instrument that tests comprehension of written text. Similar to TOFHLA-S, administration of IDL requires 20–30 minutes or longer. The lengthy administration time of these two Spanish assessment tools limits their application as screening instruments for low health literacy in clinical and community settings. Abbreviated versions of the TOFHLA have recently been developed (Baker et al. 1999; Nurss et al. 2001). But users of the TOFHLA are concerned that the test design may be intimidating and not suitable for people with limited literacy skills.[2]

An alternative method of assessing health literacy is word recognition tests, which require examinees to read aloud from a list of individual words. The tests are based on a strong correlation between reading ability and comprehension—i.e., if the examinees have trouble pronouncing words, a

Table 1: Summary of Health Literacy Assessment Tools*

| *Instrument Name* | *Test Type* | *Design* | *Spanish Version* |
|---|---|---|---|
| *Test of Functional Health Literacy in Adults* (TOFHLA) | Cloze-type comprehension test† | It consists of a reading comprehension section, containing 50 items that measure a patient's ability to read and fill in the missing word on selected passages about upper GI series, a Medicaid application, and a procedure consent form. It also contains 17 numeracy items that assess a patient's ability to understand numbers such as those on prescription bottle labels, blood glucose results, and schedule times on appointment slips. The test is timed and the examinee must stop after a set time period for the separate reading and numeracy sections. Short versions of the test have been developed | Available |
| *Instrument for Diagnosis of Reading* | Comprehension test | Procedurally, the examinee reads text from a particular grade level, the text is then taken away, and the examinee answers multiple-choice, literal, and inferential questions. There is no time limit for reading text or answering questions | Available; *Instrumento Para Diagnosticar Lecturas* (IDL) |
| *Rapid Estimate of Adult Literacy in Medicine* (REALM) | Word recognition test | It is composed of a list of common medical words and layman's terms for human anatomy and illness. The examinee is instructed to read aloud from the list of medical words and his/her literacy level is rated by the number of words that are correctly pronounced. Dictionary pronunciation is the scoring standard | Unavailable; translation unsuccessful due to the close phoneme-grapheme correspondence in Spanish |

*For a more comprehensive and detailed discussion of assessment tools, see Davis et al. (1998).

†A Cloze test requires the examinee to fill in words that have been systematically deleted from a sample of text. The assumption is that more literate examinees will understand better the context of the text and be more capable of filling in the missing words in the text.

beginning-level reading skill, they are likely to have difficulty with comprehension, a higher order skill (Davis et al. 1998).

A widely used word recognition test of health literacy is the *Rapid Estimate of Adult Literacy in Medicine* (REALM) (Davis et al. 1991; Murphy et al. 1993). The REALM emphasizes common medical words for human anatomy and illness (Appendix 1). All words are chosen from written materials given to patients in primary care settings and are presented in ascending order of difficulty so as not to discourage the examinee. This feature makes it particularly useful for screening individuals with low reading levels. The original REALM, which contained 125 words, has been shortened to 66 words (Davis et al. 1993). The shortened version correlates well with three other tests and has high test–retest reliability. It is easy to use, taking about 1–3 minutes to administer and score, and has been used in public health and clinical settings (Davis et al. 1993, 1994; Murphy et al. 1993).

An effort to translate the REALM into Spanish was unsuccessful (Nurss et al. 1995). The study tested a Spanish REALM in a sample of 52 Spanish-speaking patients that had an average of 6.5 years of schooling. Of the patients tested, 6 percent could read none of the 66 words, 17 percent scored from 52 to 59, and the remaining 77 percent scored from 60 to 66. (No one scored between 1 and 51.) Thus, the test resulted in dichotomous groups—those who could not pronounce any of the words and those who could pronounce most of the words—and failed to differentiate the health literacy skills in the group of Spanish-speaking patients.

According to the researchers, the result was due to the phonetic structure of the Spanish language. In comparison with English, Spanish has regular phoneme–grapheme correspondence, meaning that one sound is usually represented by one letter and vice versa. Therefore, it is relatively easy to pronounce words in Spanish so long as one can recognize letters, and a low-level reader can usually score high on a word recognition test. This feature of the Spanish language violates the design basis of the REALM that there exists a high correspondence between reading ability and comprehension.

In this paper, we report how we overcame the problem of phoneme–grapheme correspondence in Spanish and developed an easy-to-use health literacy test for Spanish speakers. Specifically, we incorporated simple questions in a REALM-like instrument to test the examinee's comprehension of commonly used medical terms. The new test, named the *Short Assessment of Health Literacy for Spanish-speaking Adults* or SAHLSA, requires examinees to read aloud a list of medical terms (same as the test in the REALM) and associate each term to another word similar in meaning to demonstrate

comprehension. The following sections describe the development of the SAHLSA, the methods used to validate the instrument, results of the validation, and recommendations for improvement and use of the instrument.

## METHODS

### *Research Setting*

Subjects for the study were recruited at the Ambulatory Care Center of UNC Health Care. The center has approximately 72,000 outpatient visits per year in a range of clinics, including dermatology, internal medicine, obstetrics and gynecology, orthopedics, pediatrics, and day surgery. About 5 percent of the visits are made by Spanish-speaking patients. The center was deemed an appropriate setting for the study for three reasons. First, as the outpatient center for North Carolina's referral hospital system, it treats a diverse group of patients. Second, it offers services that address a broad spectrum of health care problems, similar to the range of health care situations represented in the REALM. Third, respondents at a health care setting are likely to be interested in the purpose of the study and be responsive to the interview.

To be eligible for participation in the study, the subjects had to meet the following criteria: (1) be fluent in either English or Spanish; (2) aged 18 years or older but less than 80 years old; (3) without obvious signs of cognitive impairment; (4) without vision or hearing problems; and (5) showing no signs of drug or alcohol intoxication. The research protocol was approved by the Institutional Review Board at the School of Public Health, the University of North Carolina at Chapel Hill.

### *Instrument Development*

The design of SAHLSA was based on the 66-term REALM (Davis et al. 1993). To test comprehension, we incorporated in the instrument simple questions that assumed the standard format of multiple-choice tests (Haladyna 1999). The format, consisting of a stem in the form of a question and choices in the form of an answer to the question, is as follows.

- Stem: question
- Key: correct choice
- Distractor: plausible but incorrect choice

Specifically, two common and simple words were chosen to match each of the REALM medical terms (the stem) ("don't know" was also included as an option). One of the words (the key) was meaningfully associated with the REALM medical term and the other (the distractor) was not. Thus, the test was akin to "defining" in educational achievement testing, which measures understanding or comprehension based on correct identification of a paraphrased version of an original concept, fact, principle, or procedure as presented during instruction (Haladyna 1999). Because the purpose of the association questions was to verify the comprehension of the given medical terms, examinees were instructed not to guess. The difficulty of the two added words was kept minimal so that any examinee with a low level of education could understand them.

The instrument was developed by an expert panel through a Delphi process (Veney and Kaluzny 1998). The panel consisted of five experts who were fluent in both English and Spanish and had extensive experience working with Spanish speakers in educational, medical, and public health settings. The Delphi process involved two steps. The first was translation of the 66 REALM medical terms into Spanish. The translation took into account both the dictionary definition and the commonality of usage in daily conversations. The second step was selection of the key and distractor for each REALM medical term according to three criteria: (1) the key and distractor should be no more complicated or difficult than the medical term; (2) the relation of the key to the medical term should be a subset, an example, a larger class, a synonym, or a function; and (3) the distractor should be an incorrect but plausible choice in terms of its association with the medical term and it should be comparable in length and complexity to the key. Discrepancies that arose during the Delphi process were discussed and resolved among panel members during three half-day meetings held between December 2002 and March 2003.

The Delphi process produced both the English and Spanish drafts of the new instrument. The drafts were pretested with 10 English-speaking and 10 Spanish-speaking subjects. As a result of the pretest, the administration of the new instrument was modified by using laminated $4'' \times 5''$ flash cards in the subsequent field test. Each card had a REALM medical term printed in boldface on the top and the two association words—i.e., the key and the distractor—at the bottom. The order of the two association words was determined randomly. Respondents were shown the cards and asked to read aloud the word in boldface. The interviewer then read the two association words and queried the respondent which one was meaningfully associated with the word in boldface. The respondent's answer (including "don't know") was recorded

on a separate sheet of paper and the answer was deemed correct only when the respondent correctly pronounced the word and made the right association.

### *Field Test and Verification of the Association Questions*

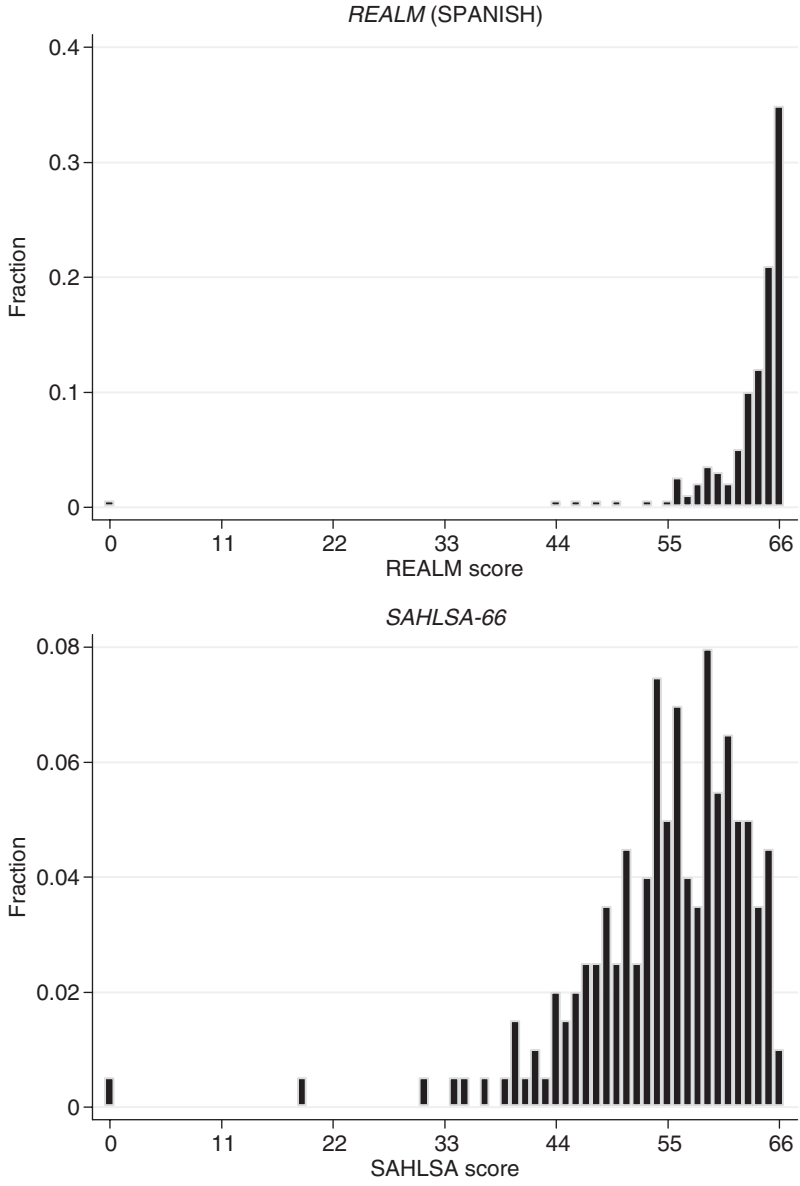
The field test was conducted with 202 English-speaking and 201 Spanish-speaking respondents. The two groups of respondents had similar gender composition, with female respondents representing around 56 percent of the total sample. On average, Spanish-speaking respondents tended to be younger (34.2 versus 43.7 years) and have fewer years of schooling (10.1 versus 13.0 years) than English-speaking respondents. The interview was conducted by trained bilingual interviewers using a questionnaire that included the 66 SAHLSA test items and questions regarding the respondents' demographic attributes (i.e., years of schooling, gender, age, and marital status) and physical health status (based on the physical health component of the SF-12 battery). Also included in the questionnaire was the Spanish TOFHLA, used as a comparison in the validation of SAHLSA. To assess test–retest reliability, a subsample of 40 Spanish-speaking respondents was randomly selected to retake the SAHLSA 2 weeks after the first interview.

The main purpose of including English-speaking respondents was to verify the design of the association words in the SAHLSA using the correlation between the REALM score and the SAHLSA association test score. The analysis indicated a high correlation ($r = 0.76$), suggesting the design of association questions was adequate. In addition, plots of the test scores in the Spanish-speaking subjects were generated (Figure 1). A plot of the scores based on Spanish-speaking repondents' pronunciation of REALM medical terms revealed a highly skewed distribution with no subject scoring between 1 and 42, confirming that the REALM could not satisfactorily differentiate the health literacy level among Spanish speakers (Nurss et al. 1995). Incorporating the association questions in the new instrument significantly increased the variation of the test scores.

### *Psychometric Assessment of the SAHLSA*

As the purpose of the study was to develop a health literacy test for the Spanish-speaking population, validation of the SAHLSA was performed with data collected from the Spanish-speaking sample. First, following Hambleton and Rovinelli (1986), the assumption of unidimensionality was examined based on the eigenvalue plot (i.e., the scree plot) of the interitem correlation matrix (using tetrachoric correlations). The plot was generated using *STATA 8*

Figure 1: Comparison of the 66-Item Rapid Estimate of Adult Literacy in Medicine (REALM) (above) and Short Assessment of Health Literacy for Spanish-Speaking Adults-66 (bottom) Scores in 201 Spanish-Speaking Subjects.



software to determine whether a dominant first factor was present among the items included in the SAHLSA. Confirmatory factor analysis was also performed using the *M-Plus* computing program to verify unidimensionality (Muthen and Muthen 1998).

Second, item response theory (IRT) was used to calibrate the SAHLSA. IRT is a model-based and item-oriented approach to psychometric test development (Zickar 1998; Embretson and Reise 2000; Ellis and Mead 2002). It assumes that an examinee's response to an item on a test is related to a latent trait which the test is presumed to measure and that the relationship can be represented by a mathematical function (usually an s-shaped, logistic function) known as an item characteristic curve (ICC). Stated differently, IRT uses information from both the examinees and the item to determine the likelihood that a person with a given level of ability (referred to as $\theta$) responds correctly to a given item. Because the mathematical model used to derive item parameters is based on the estimated latent trait ($\theta$), rather than the examinee's total score, the item and test information generated by IRT is parameter invariant—that is, information obtained from one sample using IRT, assuming the sample is sufficiently large (200 or more for a unidimensional model with dichotomously scored items) but not necessarily representative of the target population, will be equivalent to that obtained from another sample, irrespective of the average ability level of the examinees in the two samples.

The ICCs of the dichotomously scored items in SAHLSA could be estimated using the three-, two-, and one-parameter logistic models (3PLM, 2PLM, and 1PLM, respectively). The 3PLM is written as:

$$P_i(\theta) = c_i + (1 - c_i)\frac{1}{[1 + \exp\{-Da_i(\theta - b_i)\}]}$$

where $P_i(\theta)$ is the probability an examinee with ability $\theta$ (in this case, health literacy) answers item $i$ correctly; $a_i$ is the discrimination parameter indicating the degree to which small differences in ability are associated with different probabilities of correctly answering item $i$; $b_i$ is the difficulty parameter corresponding to the ability level associated with a 0.50 probability of answering item $i$ correctly; and $c_i$ is the guessing parameter or the probability that an examinee who is infinitely low on the ability answers item $i$ correctly. The 2PLM assumes no guessing and estimates item difficulty and discrimination. The 1PLM estimates item difficulty only and assumes that the discrimination parameter is equal across items. When fit to the data and accurate parameter estimates are needed, 2PLM and 3PLM are favored (Embretson and Reise 2000). Thus, the relative fit of the two models and the parameters would be

estimated using the *MULTILOG* program, suitable for dichotomous and polytomous item analysis (Thissen 1991).

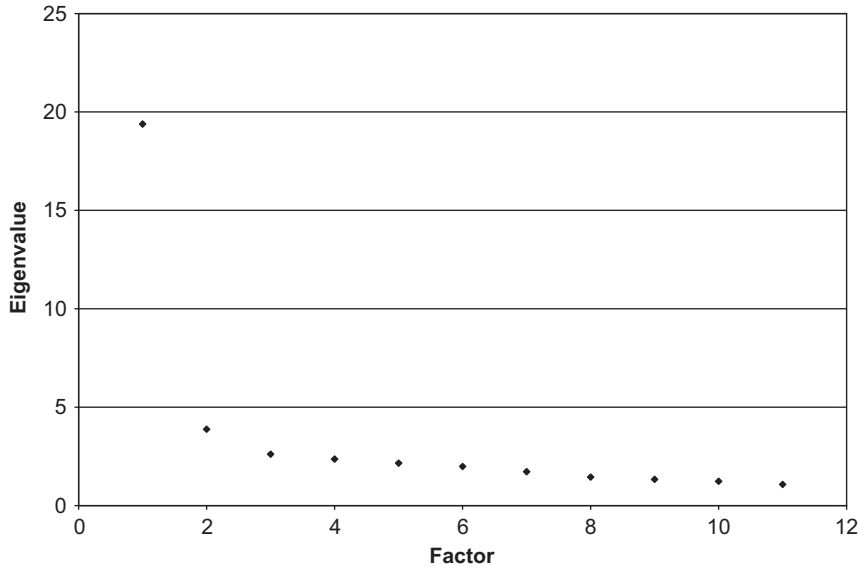
The validity (convergent validity and predictive validity) and reliability (internal reliability and test–retest reliability) of SAHLSA were also examined using *STATA 8*. Convergent validity measured the degree to which the SAHLSA was correlated with the Spanish TOFHLA using Pearson's $r$ correlation. Predictive validity assessed whether the SAHLSA could significantly predict health status among Spanish speakers, holding constant age and years of education. Internal reliability, based on the Cronbach's $\alpha$, indicated the extent to which the SAHLSA yielded similar results among its different components. Test–retest reliability, or the consistency of test results at different time points, was assessed using the Pearson's $r$ correlation between the first and second SAHLSA scores in the subsample of 40 Spanish-speaking subjects.

## RESULTS

In testing unidimensionality, the scree plot showed a clear dominance of the first factor. The eigenvalue for the first factor of the SAHLSA items was over five times larger than that of the second largest, and the second largest eigenvalue was similar to the smaller ones, suggesting the items were indicators of a common, latent factor (Figure 2). Results of confirmatory factor analysis also indicated generally good fit of the single-factor model (i.e., unidimensionality). The $\chi^2$ value was 64.2, with a $p$-value of .07; the relative $\chi^2$/df equaled 1.31; and the comparative fit index (CFI) and the root mean square error of approximation (RMSEA) were 0.844 and 0.04, respectively.[3]

The $-2$ log likelihood of the 2PLM (6201.2) and 3PLM (6139.5) were estimated to assess relative fit of the models. The difference ($\chi^2 = 61.7$) was not significant at 66 degrees of freedom, suggesting that the 3PLM did not provide a better fit for the data and that the guessing parameter could be assumed to be 0 (i.e., there was no guessing in the SAHLSA testing). Using the 2PLM, item parameters $a$ (discrimination) and $b$ (difficulty) were estimated and the results are presented in Table 2. Most items appeared to have good discrimination ($a > 1$) and low difficulty ($b < 0$). To achieve satisfactory discrimination and to test a wide range of health literacy ability, we limited the value of parameter $a$ to be between 0.6 and 3 and parameter $b$ to be between $-3$ and $+3$.[4] Based on these criteria, 14 items were eliminated. Two more items—"nausea" and "emergencia"—were also eliminated on the basis of their having similar

Figure 2: Eigenvalue Plot for the Items of Short Assessment of Health Literacy for Spanish-speaking Adults in 201 Spanish-speaking Subjects.




parameters as others ("depresión" and "osteoporosis") and therefore providing redundant information in the test. The resulting instrument contained 50 items, shown in Appendix 2 with the associated key and distractor words and with the items rank-ordered by the estimated difficulty parameter *b*.

The SAHLSA-50 score was significantly correlated with the TOFHLA score (Pearson's $r = 0.65$) in the sample of Spanish-speaking subjects. As expected, the SAHLSA-50 score was also significantly and positively associated with the physical health status of Spanish-speaking subjects ($p < .05$), holding constant age and years of education (Table 3). The instrument displayed good internal reliability (Cronbach's $\alpha = 0.92$) and test–retest reliability (Pearson's $r = 0.86$).

We were unable to convert SAHLSA-50 scores to grade-equivalent reading levels because an appropriate comparison instrument, such as the IDL, was unavailable in the study to make the conversion. As an alternative, a cutoff point for inadequate health literacy ($\leq 37$) was determined by plotting the distribution of SAHLSA-50 scores against the distributions of educational attainment and TOFHLA scores. Based on this criterion, 50 (24.8 percent) of the Spanish speakers in our sample had inadequate health literacy.

Table 2: IRT Fit Statistics and 2PLM Parameter Estimates for SAHLSA Items

| *Item** *Spanish* | *(English)* | *a* | *b* |
|---|---|---|---|
| grasa | (fat) | 2.01 | – 2.35 |
| gripe | (flu) | 1.71 | – 2.48 |
| pastilla | (pill) | 1.35 | – 1.95 |
| **dosis** | **(dose)** | **1.71** | **– 3.00** |
| ojo | (eye) | 1.97 | – 1.85 |
| estrés | (stress) | 1.05 | – 1.80 |
| papanicolao | (pap smear) | 1.84 | – 1.63 |
| **nervios** | **(nerves)** | **12.96** | **– 3.21** |
| **microbios** | **(germs)** | **4.94** | **– 1.72** |
| comidas | (meals) | 2.34 | – 2.39 |
| **enfermedad** | **(disease)** | **1.74** | **– 3.20** |
| **cáncer** | **(cancer)** | **0.80** | **– 3.04** |
| cafeína | (caffeine) | 1.70 | – 2.25 |
| ataque | (attack) | 1.33 | – 1.49 |
| riñón | (kidney) | 1.48 | – 0.83 |
| **hormonas** | **(hormones)** | **3.42** | **– 2.34** |
| herpes | (herpes) | 1.03 | – 1.07 |
| convulsiones | (seizure) | 1.41 | 0.13 |
| intestinos | (bowel) | 1.53 | – 0.57 |
| asma | (asthma) | 2.80 | – 2.34 |
| rectal | (rectal) | 2.22 | – 1.89 |
| incesto | (incest) | 1.57 | – 2.10 |
| **fatiga** | **(fatigue)** | **1.33** | **– 3.55** |
| **pélvico** | **(pelvic)** | **0.47** | **– 2.40** |
| ictericia | (jaundice) | 1.71 | – 1.65 |
| **infección** | **(infection)** | **1.85** | **– 3.08** |
| **ejercicio** | **(exercise)** | **0.95** | **– 3.26** |
| comportamiento | (behavior) | 0.68 | – 1.41 |
| **receta** | **(prescription)** | **1.29** | **– 3.42** |
| avisar | (notify) | 1.45 | – 2.48 |
| vesícula biliar | (gallbladder) | 0.71 | 0.74 |
| calorías | (calories) | 1.58 | – 0.78 |
| depresión | (depression) | 1.93 | – 2.16 |
| aborto espontáneo | (miscarriage) | 2.08 | – 1.75 |
| embarazo | (pregnancy) | 1.27 | – 2.13 |
| artritis | (arthritis) | 1.45 | 0.12 |
| nutrición | (nutrition) | 1.35 | – 1.36 |
| menopausia | (menopause) | 2.73 | – 1.49 |
| apéndice | (appendix) | 2.02 | – 1.43 |
| abnormal | (abnormal) | 1.86 | – 1.81 |
| sífilis | (syphilis) | 1.76 | – 1.27 |
| hemorroides | (hemorrhoids) | 1.41 | – 1.13 |

*Continued*

Table 2: *Continued*

| *Item** *Spanish* | *(English)* | *a* | *b* |
|---|---|---|---|
| **náusea** | **(nausea)** | **1.94** | **– 2.15** |
| indicado | (directed) | 1.05 | – 1.50 |
| alérgico | (allergic) | 1.07 | – 1.02 |
| menstrual | (menstrual) | 1.66 | – 2.63 |
| testículo | (testicle) | 1.03 | – 1.90 |
| colitis | (colitis) | 1.22 | – 0.54 |
| **emergencia** | **(emergency)** | **1.33** | **– 2.27** |
| medicamento | (medication) | 1.21 | – 0.75 |
| empleo | (occupation) | 0.95 | – 2.71 |
| **sexualmente** | **(sexually)** | **0.97** | **– 3.85** |
| alcoholismo | (alcoholism) | 2.36 | – 2.38 |
| irritación | (irritation) | 1.33 | – 1.81 |
| estreñimiento | (constipation) | 2.22 | – 2.15 |
| inflamatorio | (inflammatory) | 1.83 | – 1.27 |
| diabetes | (diabetes) | 1.03 | – 1.36 |
| hepatitis | (hepatitis) | 0.60 | 0.75 |
| **antibióticos** | **(antibiotics)** | **– 0.01** | **– 140.49** |
| **diagnóstico** | **(diagnosis)** | **0.44** | **– 4.05** |
| potasio | (potassium) | 1.28 | – 0.57 |
| anemia | (anemia) | 1.61 | – 0.66 |
| obesidad | (obesity) | 0.82 | – 0.01 |
| osteoporosis | (osteoporosis) | 1.35 | – 2.25 |
| impétigo | (impetigo) | 2.13 | – 1.55 |
| próstata | (prostate) | 0.96 | – 2.98 |

*Items in boldface were deleted based on the IRT results.

IRT, item response theory; 2PLM, two-parameter logistic model; SAHLSA, *Short Assessment of Health Literacy for Spanish-speaking Adults.*

Table 3: Ordinary Least Square Regression Results for Assessing the Predictive Validity of the SAHLSA-50

| *Variables* | *β* | *Standard Error* |
|---|---|---|
| SAHLSA-50 score | 0.17* | 0.08 |
| Age | – 0.11* | 0.04 |
| Years of education | – 0.03 | 0.13 |
| Constant | 37.86*** | 3.85 |

$N = 198^{\dagger}$; $F(3, 194) = 3.90$; adjusted $R^2 = 0.06$

†Three observations contained missing values on years of education and age.

*$p < .05$;

***$p < .001$.

SAHLSA, *Short Assessment of Health Literacy for Spanish-speaking Adults.*

Despite the care taken by the expert panel to use standard medical lexicon in the SAHLSA design, a concern could be raised about the application of SAHLSA-50 to Spanish-speakers from different countries where different idiomatic expressions are used. To examine this concern, we calculated the SAHLSA-50 (and TOFHLA) scores for subjects who were born in Mexico ($N = 141$), Central America ($N = 22$), South America ($N = 16$), and other countries ($N = 22$).[5] No significant difference was found among groups, except for those born in South American countries. Results showed that subjects from South American countries had a higher mean SAHLSA-50 score, which may be explained by the finding that they also had a higher level of educational as well as a higher mean TOFHLA score.

## DISCUSSION

This paper reports the development of SAHLSA-50, designed to provide an easy-to-administer test of health literacy in the Spanish-speaking population. Unlike the REALM, which is a word recognition test and is available only in English, the new instrument includes questions that evaluate the subject's comprehension of medical terms commonly used in clinical and public health settings. Results show that the instrument has good validity and reliability. Guessing does not appear to be a concern if clear instruction is given before the test. Administration of the instrument is relatively easy; it took a short time (approximately 3–6 minutes) and required minimal training (Appendix 3). A rather high cutoff point is found for adequate health literacy, suggesting that the SAHLSA-50 is particularly useful for identifying Spanish speakers with low health literacy.

Several limitations are worth noting. It is important to point out that the United States is host to people from different Latin American countries who use different idiomatic expressions. Further testing of the instrument's application in different Latino subpopulations may be needed. Similar to most instrument development studies, our study did not include a random, representative sample of Latinos in the community. The hospital-based participants recruited for the study may be more receptive to a health literacy test. What kind of difficulties may arise in applying the SAHLSA-50 to a community-based sample remains to be evaluated. Further testing would also be needed in order to convert the SAHLSA-50 score to grade-equivalent reading levels, should the information be necessary for designing appropriate educational materials for Spanish speakers.

It should be emphasized that SAHLSA-50 is appropriate as a screening tool, particularly for identifying Spanish-speakers with inadequate health literacy. Its usage as an assessment tool may be limited by the low differentiation power among individuals with a score greater than 37. The test takes 3–6 minutes to administer. In clinics overwhelmed with demand of patient care, it may be difficult to add the test onto current clinical activities.

Despite these limitations, the instrument has several practical applications. First, it can be used to screen for individual Spanish speakers' health literacy level in public health and clinical settings. The identification of patients with low health literacy can alert health care providers to the possibility that these patients may have difficulty with printed educational materials or highly scientific explanations of complex medical conditions. It has been shown that those who score very poorly on a health literacy test also have trouble with oral provider–patient communication (Doak, Doak, and Root 1996). Thus efforts can be taken to develop educational multimedia (e.g., videotapes, audiotapes) that are both culturally sensitive and linguistically keyed to the population. Health care providers and health educators should also be trained to alter their mode of communication with low health literacy Spanish-speaking patients according to assessed health literacy levels (Davis et al. 1998). Increased awareness among health care practitioners of the special health and personal needs of low health literacy Spanish-speaking patients may help reduce the level of linguistic complexity used in provider–patient communications, thus preventing serious medical errors because of misunderstanding. This, in turn, has the potential to improve the quality of care and reduce the health care costs of service to the Spanish-speaking population (Flouty and Meyer 2000).

Second, the instrument could be used to assess the level of health literacy in the community. Assessing the aggregate extent of inadequate health literacy among Spanish-speaking Latino residents will provide local governments, community groups, hospital systems, or other health provider groups an understanding of their Latino patient populations' unique health needs (Brandes 1996; Davis et al. 1998). The information could be used to guide the design of appropriate health educational materials (written and/or multimedia) or for devising community intervention programs that are comparable with the health literacy level of the local Latino population.

Research indicates that future development of stable bilingualism in the United States is unlikely (Sole 1990). Given the rate of growth of Latino residents and the tendency of Latinos to reside in Spanish-speaking communities

that use Spanish as the primary language, it is clear that Spanish will remain the most important minority language in the United States for decades to come. The development of Spanish tests, as represented in the study, is the first and necessary step toward understanding the health needs of, and providing high quality of care for, the Latino population.

## ACKNOWLEDGMENTS

The study was supported by a grant from the Agency for Healthcare Research and Quality (R03-HS13233). Preliminary results of the study were presented at the 2004 AcademyHealth Annual Meeting in San Diego. We thank Dr. Maria Rosa Watson and two anonymous reviewers of the journal for their constructive comments on the manuscript. Michelle Langer assisted with the item response theory analysis, for which we are grateful.

*Disclosures*: None of the authors has any financial interest related to the article.

*Disclaimers*: The views expressed in the article are those of the authors and do not necessarily reflect the views of the Agency for Healthcare Research and Quality, the University of North Carolina at Chapel Hill, or the University of Illinois at Chicago.

## NOTES

1. Use of educational levels or years of schooling tends to overestimate the reading ability on average by 3- to 5-grade levels (Davis et al. 1993, 1996; Nurss et al. 1997).
2. Based on our communication with users of the TOFHLA and our personal experience with the test.
3. Each available index of fit for confirmatory factor analysis addresses a slightly different issue, and no index of fit is considered to be perfect. The generally acceptable values for the fit indexes reported in the paper are: (1) the $\chi^2$ should have a nonsignificant $p$-value, (2) the relative $\chi^2$ should be less than 3, (3) the CFI $>0.9$, and (4) the RMSEA $<0.05$ (Munro 2005).
4. No clear criteria have been established to determine good item parameters. The ranges of the parameters were determined on the basis of the research purpose and in discussion with the Quantitative Psychology Program, led by Professor David Thissen at the University of North Carolina at Chapel Hill.
5. The "other" category included missings.

## REFERENCES

Ammons, L. 1997. "Demographic Profile of Health-Care Coverage in America in 1993." *Journal of the National Medical Association* 89 (11): 737–44.

Baker, D. W., R. M. Parker, M. V. Williams, and W. S. Clark. 1998. "Health Literacy and the Risk of Hospital Admission." *Journal of General Internal Medicine* 13 (12): 791–8.

Baker, D. W., R. M. Parker, M. V. Williams, W. S. Clark, and J. Nurss. 1997. "The Relationship of Patient Reading Ability to Self-Reported Health and Use of Health Services." *American Journal of Public Health* 87 (6): 1027–30.

Baker, D. W., M. V. Williams, R. M. Parker, J. A. Gazmararian, and J. Nurss. 1999. "Development of a Brief Test to Measure Functional Health Literacy." *Patient Education and Counseling* 38 (1): 33–42.

Bass, P. F. III., J. F. Wilson, C. H. Griffith, and D. R. Barnett. 2002. "Residents' Ability to Identify Patients with Poor Literacy Skills." *Academic Medicine* 77 (10): 1039–41.

Brandes, W. L. 1996. *Literacy, Health and the Law: An Exploration of the Law and the Plight of Marginal Readers within the Health Care System: Advocating for Patients and Providers.* Philadelphia: Health Promotion Council of Southeastern Pennsylvania Inc.

Davis, T. C., C. Arnold, H. J. Berkel, I. Nandy, R. H. Jackson, and J. Glass. 1996. "Knowledge and Attitude on Screening Mammography among Low-Literate, Low-Income Women." *Cancer* 78 (9): 1912–20.

Davis, T. C., M. A. Crouch, S. W. Long, R. H. Jackson, P. Bates, R. B. George, and L. E. Bairnsfather. 1991. "Rapid Assessment of Literacy Levels of Adult Primary Care Patients." *Family Medicine* 23 (6): 433–5.

Davis, T. C., R. H. Jackson, R. B. George, S. W. Long, D. Talley, P. W. Murphy, E. J. Mayeaux, and T. Truong. 1993. "Reading Ability in Patients in Substance Misuse Treatment Centers." *International Journal of Addiction* 28 (6): 571–82.

Davis, T. C., E. M. Keenen, J. Gazmararian, and M. Williams. 2004. "Literacy Testing in Health Care Research." In *Understanding Health Literacy: Implications for Medicine and Public Health*, edited by J. Schwartzberg, J. B. VanGeest, and C. Wang, pp. 157–79. Chicago: American Medical Association Press.

Davis, T. C., S. W. Long, R. H. Jackson, E. J. Mayeaux, R. B. George, P. W. Murphy, and M. A. Crouch. 1993. "Rapid Estimate of Adult Literacy in Medicine: A Shortened Screening Instrument." *Family Medicine* 25 (6): 391–5.

Davis, T.C, E. J. Mayeaus, R. Fredrickson, J. Bocchini, R. H. Jackson, and P. W. Murphy. 1994. "Reading Ability of Parents Compared with Reading Level of Pediatric Patient Education Materials." *Pediatrics* 93 (3): 460–8.

Davis, T. C., R. Michielutte, E. N. Askov, M. V. Williams, and B. D. Weiss. 1998. "Practical Assessment of Adult Literacy in Health Care." *Health Education and Behavior* 25 (5): 613–24.

del Pinal, J. H. 1996. "Hispanic Americans in the United States: Young, Dynamic and Diverse." *Statistical Bulletin—Metropolitan Insurance Companies* 77 (4): 2–13.

Diaz, E., H. Prigerson, R. Desai, and R. Rosenheck. 2001. "Perceived Needs and Service Use of Spanish Speaking Monolingual Patients Followed at a Hispanic Clinic." *Community Mental Health Journal* 37 (4): 335–46.

Doak, C. C., L. G. Doak, and J. H. Root. 1996. *Teaching Patients with Low-Literacy Skills.* 2d edition. Philadelphia: JB Lippincott.

Ellis, B. B., and A. D. Mead. 2002. "Item Analysis: Theory and Practice Using Classical and Modern Test Theory." In *Handbook of Research Methods in Industrial and Organizational Psychology*, edited by S. G. Rogelberg, pp. 324–43. Malden, MA: Blackwell.

Embretson, S. E., and S. P. Reise. 2000. *Item Response Theory for Psychologists.* Hillsdale, NJ: Erlbaum.

Flouty, G., and P. Meyer. 2000. *Health Literacy News.* New York: Pfizer Inc.

Gazmararian, J. A., D. W. Baker, M. V. Williams, R. M. Parker, T. L. Scott, D. C. Green, S. N. Fehrenbach, J. Ren, and J. P. Koplan. 1999. "Health Literacy among Medicare Enrollees in a Managed Care Organization." *Journal of the American Medical Association* 281 (6): 545–51.

Haladyna, T. M. 1999. *Developing and Validating Multiple-Choice Test Items.* 2d edition. Mahwah, NJ: Lawrence Erlbaum Associates.

Hambleton, R. K., and R. J. Rovinelli. 1986. "Assessing the Dimensionality of a Set of Test Items." *Applied Psychological Measurement* 10: 287–302.

Kalichman, S. C., E. Benotsch, T. Suarez, S. Catz, J. Miller, and D. Rompa. 2000. "Health Literacy and Health-Related Knowledge among Persons Living with HIV/AIDS." *American Journal of Preventive Medicine* 18 (4): 325–31.

Lee, S.-Y. D., A. Arozullah, and Y. I. Cho. 2004. "Health Literacy, Social Support and Health: A Research Agenda." *Social Science and Medicine* 58 (7): 1309–21.

Marwick, C. 1997. "Patients' Lack of Literacy May Contribute to Billions of Dollars in Higher Hospital Costs." *Journal of the American Medical Association* 278 (12): 971–2.

Munro, B. H. 2005. *Statistical Methods for Health Care Research.* 5th edition. Philadelphia: Lippincott Williams & Wilkins.

Murphy, P. W., T. C. Davis, S. W. Long, R. H. Jackson, and B. C. Decker. 1993. "Rapid Estimate of Adult Literacy in Medicine (REALM): A Quick Reading Test for Patients." *Journal of Reading* 37: 121–30.

Muthen, L. K., and B. O. Muthen. 1998. *Mplus User's Guide.* Los Angeles, CA: Muthen & Muthen.

Nurss, J. R., D. W. Baker, T. C. Davis, R. M. Parker, and M. Williams. 1995. "Difficulties in Functional Health Literacy Screening in Spanish-Speaking Adults." *Journal of Reading* 38: 632–7.

Nurss, J. R., I. M. el-Kebbi, D. L. Gallina, D. C. Ziemer, V. C. Musey, S. Lewis, Q. Liao, and L. S. Phillips. 1997. "Diabetes in Urban African Americans: Functional Health Literacy of Municipal Hospital Outpatients with Diabetes." *Diabetes Education* 23 (5): 563–8.

Nurss, J. R., R. M. Parker, M. V. Williams, and D. W. Baker. 2001. *Test of Functional Health Literacy in Adults.* Snow Camp, NC: Peppercorn Books & Press.

Parker, R. M., D. W. Baker, M. V. Williams, and J. R. Nurss. 1995. "The Test of Functional Health Literacy in Adults: A New Instrument for Measuring Patients' Literacy Skills." *Journal of General Internal Medicine* 10 (10): 537–41.

Roter, D. L., E. R. Rudd, and J. Comings. 1998. "Patient Literacy—A Barrier to Quality of Care." *Journal of General Internal Medicine* 13 (12): 850–1.

Rudd, R. E., B. A. Moeykens, and T. C. Colton. 2000. "Health and Literacy: A Review of Medical and Public Health Literature." *Annual Review of Adult Learning and Literacy* 1: 158–99.

Ruiz, P. 1993. "Access to Health Care for Uninsured Hispanics: Policy Recommendations." *Hospital and Community Psychiatry* 44 (10): 958–62.

———. 1995. "Assessing, Diagnosing and Treating Culturally Diverse Individuals: A Hispanic Perspective." *Psychiatric Quarterly* 66 (4): 329–41.

Schur, C. L., and L. A. Albers. 1996. "Language, Sociodemographics, and Health Care Use of Hispanic Adults." *Journal of Health Care for the Poor and Underserved* 7 (2): 140–58.

Selden, C. R., M. Zorn, S. Ratzan, and R. M. Parker (Eds). 2000. *Health Literacy. Current Bibliographies in Medicine. No 2000–1.* Bethesda, MD: National Library of Medicine [Available at http://www.nlm.gov/pubs/cbm/hliteracy.html].

Sole, Y. R. 1990. "Bilingualism: Stable or Transitional? The Case of Spanish in the United States." *International Journal of the Sociology of Language* 84: 35–80.

Thissen, D. 1991. *MULTILOG User's Guide: Multiple Categorical Item Analysis and Test Scoring Using Item Response Theory.* Chicago: Scientific Software Int.

U.S. Census Bureau. 2000. "The Hispanic Population." Available at http://www.census.gov/population/www/pop-profile/hisppop.html

U.S. Census Bureau. 2002. *American Community Survey 2002.* Available at http://www.sscnet.ucla.edu/issr/da/index/techinfo/M9491.HTM

U.S. Department of Health and Human Services. 1998. *Healthy People 2010 Objectives: Draft for Public Comment.* Washington, DC: Public Health Service.

Veney, J. E., and A. D. Kaluzny. 1998. *Evaluation & Decision Making for Health Services.* 3d edition. Chicago: Health Administration Press.

Weinick, R. M., and N. A. Krauss. 2000. "Racial/Ethnic Differences in Children's Access to Care." *American Journal of Public Health* 90 (11): 1771–4.

Weiss, B. D., J. S. Blanchard, D. L. McGee, G. Hart, W. Warren, M. Burgoon, and K. J. Smith. 1994. "Illiteracy among Medicaid Recipients and Its Relationship to Health Care Costs." *Journal of Health Care for the Poor and Underserved* 5 (2): 99–111.

Williams, M. V., D. W. Baker, R. M. Parker, and J. R. Nurss. 1998. "Relationship of Functional Health Literacy to Patient's Knowledge of Their Chronic Disease: A Study of Patients with Hypertension and Diabetes." *Archives of Internal Medicine* 158 (2): 166–72.

Zickar, M. J. 1998. "Modeling Item-Level Data with Item Response Theory." *Current Directions in Psychological Science* 7: 104–9.

## Supplementary Material

The following supplementary material for this article is available online:

Appendix 1. Rapid Estimate of Adult Literacy in Medicine (REALM).

APPENDIX 2. The 50 items of SAHLSA, rank-ordered according to the parameter *b* of item difficulty (keys and distracters are listed in the same random order as in the field interview).

APPENDIX 3. Instruction for Administering SAHLSA-50.