Development and Validation of the Spanish Numeracy Understanding in Medicine Instrument

Elizabeth A. Jacobs, MD AMPP[1], Cindy M. Walker, PhD[2], Tamara Miller, PhD[2], Kathlyn E. Fletcher, MD, MA[3], Pamela S. Ganschow, MD[4], Diana Imbert, BS[5], Maria O'Connell, BA[4], Joan M. Neuner, MD, MPH[3], and Marilyn M. Schapira, MD, MPH[5,6]

[1]Department of Medicine, School of Medicine and Public Health, University of Wisconsin, Madison, WI, USA; [2]Department of Educational Psychology, University of Wisconsin-Milwaukee, Milwaukee, WI, USA; [3]Department of Medicine, Medical College of Wisconsin and the Zablocki VA Medical Center, Milwaukee, WI, USA; [4]Department of Medicine, School of Medicine, Rush University, Chicago, IL, USA; [5]Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; [6]The Center for Health Equity Research Program, Philadelphia VA Medical Center, Philadelphia, PA, USA.

BACKGROUND: The Spanish-speaking population in the U.S. is large and growing and is known to have lower health literacy than the English-speaking population. Less is known about the health numeracy of this population due to a lack of health numeracy measures in Spanish.

OBJECTIVE: we aimed to develop and validate a short and easy to use measure of health numeracy for Spanish-speaking adults: the Spanish Numeracy Understanding in Medicine Instrument (Spanish-NUMi).

DESIGN: Items were generated based on qualitative studies in English- and Spanish-speaking adults and translated into Spanish using a group translation and consensus process. Candidate items for the Spanish NUMi were selected from an eight-item validated English Short NUMi. Differential Item Functioning (DIF) was conducted to evaluate equivalence between English and Spanish items. Cronbach's alpha was computed as a measure of reliability and a Pearson's correlation was used to evaluate the association between test scores and the Spanish Test of Functional Health Literacy (S-TOFHLA) and education level.

PARTICIPANTS: Two-hundred and thirty-two Spanish-speaking Chicago residents were included in the study.

KEY RESULTS: The study population was diverse in age, gender, and level of education and 70 % reported Mexico as their country of origin. Two items of the English eight-item Short NUMi demonstrated DIF and were dropped. The resulting six-item test had a Cronbach's alpha of 0.72, a range of difficulty using classical test statistics (percent correct: 0.48 to 0.86), and adequate discrimination (item-total score correlation: 0.34–0.49). Scores were positively correlated with print literacy as measured by the S- TOFHLA ($r = 0.67$; $p < 0.001$) and varied as predicted across grade level; mean scores for up to eighth grade, ninth through twelfth grade, and some college experience or more, respectively, were 2.48 (SD ± 1.64), 4.15 (SD ± 1.45), and 4.82 (SD ± 0.37).

CONCLUSIONS: The Spanish NUMi is a reliable and valid measure of important numerical concepts used in communicating health information.

INTRODUCTION

Health numeracy is the term used to describe a person's ability to use quantitative information to communicate with their providers, participate in medical decision making, and take care of their health. Health numeracy is one domain of the general construct of health literacy,[1] with theoretical frameworks identifying a range of skills encompassed by the construct including basic number sense, use of tables and graphs, and concepts of probability and statistics.[2–5] Previous studies reporting an association between low health literacy and poorer health outcomes have used measures that incorporate both print and numeric domains.[6] However, an emerging literature supports the association of skills in the health numeracy domain to health-related outcomes including accuracy of risk perceptions,[7,8] ability to interpret of risk information,[9–12] identification of high quality care options,[13] decreased susceptibility to framing biases,[14] reduced hospitalizations for asthma,[15] improved anticoagulation control[16] and management of diabetes[17] and increased medication management for HIV infection.

Numeric concepts may differ across cultures and be challenging to translate across languages. Understanding a patient's level of health numeracy may help a provider to communicate more effectively and provide a higher level of patient-centered care.[18] Many limited English proficient (LEP) individuals in the US have inadequate or marginal general health literacy,[19–22] defined as "the degree to which individuals have the capacity to obtain, process, and

understand basic health information needed to make appropriate health decisions."[23] Low health literacy has been associated with disparities across a range of conditions and settings for LEP patients.[24–28] This is a large and growing problem in the U.S; the proportion of people with limited English proficiency increased by 80 % between 1990 and 2010[29] and the vast majority of this population is Spanish speaking.

Less is known about health numeracy among LEP patients regarding numerical concepts important in health-related self-care and decision making.[2–5,30] While there are health numeracy measures available, they vary in their area of focus and whether they assess knowledge, skills, and/or preference for use of numbers or the full range of skills an adult needs to understand and use numbers in decision making and medication administration.[11,15,16,31–37] Relatively few health numeracy specific measures that have been translated into Spanish. We identified three composite health literacy measures that include measures of health numeracy and have also been translated and validated in Spanish; the Parental Health Literacy Activities Test (PHLAT-Spanish),[38] the Newest Vital Sign,[36] and the Spanish Test of Functional Health Literacy in Adults (STOFHLA).[39] In addition, we identified one specific health numeracy measure, the Diabetes-Specific Numeracy Measure, that has been validated in Spanish.[40] These measures have strengths, but also limitations that include that they are not self-administered (PHLAT-Spanish and NVS) or were not developed to measure the full range of numeracy skills important to communication, decision making, and participation in one's health care. The PHLAT-Spanish is also specifically designed to assess health literacy and numeracy among parents of young children rather than adults in general. Finally, none of these measures were developed using modern measurement approaches to scale development such as item response theory.[41] Modern approaches have several advantages over classical test theory, including the ability to evaluate item bias that may exist between groups of respondents, facilitating the development and validation of versions translated to other languages.[41]

Our objective was to develop and validate a short and easy to use measure of health numeracy for Spanish-speaking adults in the United States using item response theory psychometric methods.

METHODS

Overview

The Spanish Numeracy Understanding In Medicine Instrument (NUMi) was developed using response theory (IRT). Details of the development and validation of the 20-item and eight-item versions of the English NUMi have been previously published.[42,43] We chose to base our Spanish NUMi on the shorter version of the English NUMi (eight-item) to create a scale that has as low a respondent burden as possible while maintaining strong psychometric properties. In brief, the NUMi was based on an empirically derived theoretical framework of health numeracy including the domains of number sense, tables and graphs, probability, and statistics.[2,4] The use of IRT psychometric methods offers several advantages for scale development. IRT methods allow for the assessment of measurement bias through use of differential item functioning (DIF) analysis, a valuable approach for the development and evaluation of cross-cultural measurement tools.[44,45] We used IRT methods to generate and calibrate a large item bank (n = 110) to assess the health numeracy construct.[42] Items were calibrated using data from the English respondents using BILOG software.[46] Qualitative studies [5,30] were conducted using both English and Spanish-speaking adults to develop items that would work in both Spanish-speaking and English-speaking and Hispanic and Non-Hispanic populations. Data used for psychometric analyses was initially obtained from English speakers, with 29 % of the sample identifying as being Hispanic.

The selection of items involved a systematic process that considered the psychometric properties of each item and the range of item content. Data from respondents who took the English version of the test were used as a reference group and those who took the Spanish version of the test were used as a focal group in the DIF analysis. The assumption of unidimensionality and model fit were evaluated as described below. Scale reliability was evaluated with a Cronbach alpha statistic and convergent validation was evaluated through a comparison of test scores with level of education and general health literacy as measured by the Spanish TOFHLA.

Spanish Translation

We translated all items originally created in English into Spanish using the group translation method. In this process, multiple qualified individuals translated the items individually, compared their translations, and, using group consensus, modified and agreed upon a final translation that would be most accessible and understandable to Spanish-speaking patients from a variety of backgrounds. Six translators participated in this process. One Spanish translator then reviewed the agreed-upon final translation of the Spanish items and compared them to the English items; their equivalency in both Spanish and English was then evaluated by this translator and the bilingual members of the study team. Several modifications were made to the Spanish items in this step to make them understandable and culturally relevant and as equivalent as possible to the already validated English items. We then conducted 16 cognitive interviews with Spanish-speaking participants from a variety of educational levels to make sure they were measuring the same constructs of numeracy in Spanish as they were in English.

Study Population

The resulting Spanish item bank was administered to 232 Spanish speakers. Participants were recruited using a

purposive convenience sampling. We recruited participants from Milwaukee (A Community Center) and Chicago (three clinical sites, three schools, three community centers, and one church). Prior to enrollment, potential participants were interviewed in Spanish by a native Spanish-speaking and bilingual research assistant to ensure that the participant's Spanish fluency was high enough to allow them to be able to comfortably respond to the Spanish items. In addition to ascertaining basic demographic characteristics in this interview, we asked them if they could read and write in Spanish, and provided an example of a test question in Spanish. Potential participants who could respond adequately to the questions in Spanish were invited to participate in this study. Tests were administered in a classroom setting. All participants had the option of having the test items read to them in a private setting. Additional inclusion criteria included self-identifying Spanish as a primary language and being 21 years of age or older. Exclusion criteria included visual acuity less than 20/50 with corrective lenses. Test packets included the numeracy items, the Spanish S-TOHFLA and socio-demographic data including age, gender, level of education, and country or region of origin.

Testing the Assumption of Unidimensionality

A key assumption of IRT is that the items are representing a single construct, a concept defined as unidimensionality. The unidimensionality assumption was tested using the full item bank obtained from English speakers using Stout's Test of Essential Dimensionality (DIMTEST).[47] Unidimensionality was tested across the four domains that comprised our heath numeracy construct: number sense, tables and graphs, probability, and statistics. Two hypotheses were tested: 1) the null hypothesis of unidimensionality between items in the statistics domain and remaining items and 2) the null hypothesis of unidimensionality between items in the probability and statistics domains combined compared to remaining items. Neither hypothesis yielded significant findings (t = 0.00, $p = 0.500$; and t = -0.37, $p = 0.64$, respectively). The analysis was then conducted on a 20-item version of the test with five items from each domain with similar results, supporting the assumption of unidimensionality.

Assessment of Model Fit

We used a two-parameter IRT model for item analysis.[42] Log likelihood ratio tests were conducted to compare model fit for the one-parameter and two-paremter IRT models for our data. The chi-square statistic was large with a rejection of the null hypothesis of no difference between the one- and two-parameter models at a significance level of $p < 0.001$, indicating that the two-parameter model resulted in a statistically significant improvement in fit compared to the one-parameter model. We did not evaluate the three-parameter model (including a guessing parameter), as we directed our

respondents to leave items blank rather than guess if they did not know the answer to a question.

Differential Item Functioning

Differential item functioning (DIF) refers to the potential that there are unequal probabilities of providing a correct response due to factors other than the primary construct being measured by test items. The presence of DIF would require the re-calibration of item parameters in the Spanish version of the test. We evaluated all candidate items and included only those that did not demonstrate DIF in the Spanish NUMi. DIF was evaluated using Simultaneous Item Bias Testing (SIBTEST).[45,48,49] SIBTEST compares the performance on candidate studied items in the focal group (Spanish-speaking respondents) to performance on those items in the reference group (English-speaking respondents), after matching respondents by ability. The difference in item performance is represented by a Beta Index with a level of 0.10 set as a significant degree of DIF. The null hypothesis of a non-zero difference is typically rejected if the *p* value is less than 0.05, with adjustments for multiple comparisons. A two-step analysis was conducted. In step 1, each Spanish item was compared to performance on the corresponding item in English, using the remaining items as the matching subtest. Items flagged for DIF were then removed from the matching or conditioning subtest. In step 2, the DIF analyses were repeated on all items using this purified matching subtest.[44,45,48,49] For this analysis, the reference group consisted of English-speaking examinees (n = 200) and the focal group consisted of Spanish-speaking examinees (n = 232).

Reliability and Validity Testing

We used Cronbach's alpha to evaluate the internal consistency of the Spanish NUMi and evaluated its convergent validity by comparing scores on the Spanish NUMi to the Spanish Test of Functional Health Literacy in Adults (S-TOFHLA)[39] and level of education. The association of the Spanish NUMi and S-TOFHLA was evaluated using a Spearman correlation test. An independent sample *t* test was performed, assuming unequal variances, comparing the mean Spanish NUMi scores of the high, medium, and low education groups. We divided our sample into these three groups based on the natural distribution of our sample into those who had middle school level of education or less, 9–12 years of education, and those who had some college or more. These groups are also likely to have different numeracy levels. We also evaluated the correlation of scores on the Spanish NUMi with the number correct of the 18 out of 20 translated items (those that did not demonstrate DIF) from the original NUMi.

We expected the Spanish- NUMi to be positively associated with higher health literacy as measured by the S-TOFHLA (with print and numeracy components) higher levels of education, and scores from the translated longer original NUMi measure.

RESULTS

Study Population

The study population was diverse in age, gender, and level of education (Table 1). The majority (70 %) of participants reported that Mexico was their country of origin. Ninety-three percent of participants (93 %) cited Spanish as their primary language. Forty-one percent (41 %) had no more than an eighth grade level education. Twenty-nine percent (29 %) had inadequate health literacy as measured by the Spanish Test of Functional Health Literacy in Adults (S-TOFHLA).

Translation Issues

A particular challenge was the translation of concepts pertaining to risk, uncertainty, and statistical significance. For example, the English version of items that asked about the risk of side effects used the phrases "chance of a serious side effect" and "how many would be expected" to have a side effect to assess understanding of the probability of an occurrence. The Spanish version translated this concept to "probabilidades de experimenta un effecto secundario" and "anticipa que experimentaran." However, the concepts of a probabilistic expectation were not easy to translate. The term riesgo (risk) was not interpreted by participants as a probabilistic concept as much as a term to identify exposures that put one at risk for adverse outcomes. Items that attempted to assess false positive test results and the meaning of a *p* value were also difficult to translate, perhaps due to difficulty in translating concepts of chance, risk, and statistical significance between English and Spanish.

Differential Item Functioning

Ninety-six items from our original item bank were evaluated for DIF with 15 demonstrating DIF, including two items selected for the English NUMi Eight-Item Test (Table 2). Items evaluating understanding of false positive tests and the meaning of a statistically significant finding were dropped from the Spanish version due to the finding of DIF between the English and Spanish versions, resulting in a six-item Spanish NUMi. The English version of these items and those dropped due to DIF are presented in Table 2.

Psychometric Analysis of Spanish NUMi

The six-item Spanish NUMi (available in the online appendix) demonstrated adequate reliability with a Cronbach's alpha of 0.72. Items demonstrated a range of difficulty using classical test statistics (percent correct: 0.49 to 0.86) and adequate discrimination (item-total score correlation: 0.34–0.56). Similarly, IRT parameters indicate a range in the difficulty parameter. Typically, IRT difficulty parameters range from -3.0 (easiest) to 3.0 (hardest) and discrimination parameters range from 0 (poor discrimination) to 3.0 (excellent discrimination). The easiest items on the test addressed interpretation of a nutrition label and determining if a glucose level was within goal range (difficulty parameters of -1.31 and -1.27, respectively). The hardest item on the test addressed interpretation of small risks and had a difficulty parameter of 0.199. The discrimination parameters ranged from a low of 0.62 to a high of 1.40 indicated acceptable discrimination (Table 3). The Item Characteristic Curves (ICCs) and Item Information Curves (IICs) demonstrate the range of difficulty and discrimination across the items (Figs. 1, and 2). The majority of items had negative IRT difficulty parameters, indicating they were relatively easy for examinees with high health numeracy. The Test Information Function of the final Spanish NUMi indicates that the test is most discriminating at a lower health numeracy level (Fig. 3).

Unidimensionality Assumption

The unidimensionality assumption for the item bank was supported by the Stout's Test of Essential Dimensionality (DIMTEST).[42] In order to provide supportive evidence for

Table 1. Characteristics of the Study Population

Patient Characteristics	N (%) Total: 232
GENDER	
Male	88 (38)
Female	144 (62)
AGE	
< 45 years	129 (56)
45–59 years	63 (27)
60–74 years	37 (16)
≥ 75 years	3 (1)
Missing	
RACE	
White	13 (6)
Black and/or African American	1 (0.4)
American Indian and Alaska Native	1 (0.4)
Asian	0 (0)
Native Hawaiian and/Pacific Islander	1 (0.4)
ETHNICITY	
Hispanic/Latino: Yes	223 (96)
Hispanic/Latino: No	4 (2)
Missing	5 (2)
S-TOFHLA Score	
0–16 (Inadequate health literacy)	68 (29)
17–22 (Marginal health literacy)	27 (12)
23–36 (Adequate health literacy)	136 (59)
Missing	1 (0.4)
EDUCATION	
Up to 5th grade	43 (19)
6–8th grade	50 (22)
9–12th grade	78 (34)
3 years of college	20 (9)
4 or more years of college	41 (18)
Spanish Primary Language	
Yes	215 (93)
No	13 (6)
Missing	4 (2)
Country/Region of Origin	
Mexico	163 (70)
USA	35 (15)
South America*	20 (7)
Central America†	4 (2)
Puerto Rico	1 (0.4)
Cuba	1 (0.4)
Missing	1 (0.4)

*Argentina, Chile, Columbia, Ecuador, Peru, Venezuela
† El Salvador, Honduras

Table 2. English Version of Candidate Items for Spanish NUMi

Numeracy Understanding in Medicine Short Form Candidate Items	Item Response Options *Correct Response		Item Number
James has diabetes. His goal is to have his blood sugar between 80 mg/dL and 150 mg/dL in the morning. Which of the following blood sugar readings is within his goal?	55 mg/dL 140 mg/dL* 165 mg/dL 180 mg/dL		1
Nathan has a pain rating of 5 on a pain scale of 1 (no pain) to 10 (worst possible pain). One day later Nathan still has pain but not as much. Now, what pain rating might Nathan give?	3* 5 7 9		2
Frank has a test done to look for blockages in the arteries of his heart. The doctor said that a person with a higher percent (%) blockage has a high chance of having a heart attack. Which percent (%) blockage has the highest chance of a heart attack?	33 % 50 % 75 % 99 %*		3
Natasha started a new medicine that may cause the side effects listed below. Which side effect is Natasha least likely to have?	Side Effects Dizziness Nausea Stomach Pain Allergic Reaction*	Chance of Occurring 1 in 5 people 1 in 10 people 1 in 100 people 1 in 200 people	4
James starts taking a new blood pressure medicine. The chance of a serious side effect is 0.5 %. If 1000 people take this medicine, about how many are expected to have a serious side effect?	1 person 5 people* 50 people 500 people		5
A nutrition label is shown below. How many calories did Mary eat if she had 2 cups of food?	140 calories 280 calories 560 calories* 680 calories		6
The PSA (prostate specific antigen) is a blood test that looks for prostate cancer. The test has false alarms so about 30 % of men who have an abnormal test turn out not to have prostate cancer. John has an abnormal test. What is the chance that John has prostate cancer?	0 % 30 % 70 %* 100 %		7 (Not in Spanish NUMi)
A study found that a new diabetes medicine led to control of blood sugar in 8 % more patients than the old medicine. This difference was statistically significant ($p = 0.05$). The likelihood that this finding was due to chance alone is best described as less than:	1 in 5 1 in 10 1 in 15 1 in 20*		8 (Not in Spanish NUMi)

the assumption of unidimensionality, we conducted exploratory factor analysis of the eight items considered for inclusion in the Spanish NUM to provide additional supportive evidence for unidimensionality. The analysis yielded only one factor that demonstrated an eigenvalue > 1.0. Evaluation of the residual matrix of the factor analysis also supports the independence of items (Online Fig. A).

Reliability and Validity

A strong correlation as demonstrated by a high Pearson's correlation coefficient was found between the Spanish NUMi and print literacy as measured by the S-TOFHLA ($0.67; p < 0.001$), education ($0.67; p < 0.001$), and the mean score that the respondents had on the longer NUMi form ($0.87; p < 0.001$). We further divided the Spanish-speaking sample into three education levels: up to an eigth grade level education, ninth to twelfth grade education, or some college or more. As anticipated, those in the low education group scored significantly lower 2.48 ± 1.64; m, SD, n = 93) than the examinees in the medium (4.15 ± 1.45, m, SD, n = 78) and high education groups (4.82 ± 1.37, m SD, n =61) (t (135) = -7.109, $p < .001$).

For the entire sample of examinees, the mean score on the six-item Spanish NUMi was 3.66 with a standard deviation of

Table 3. Description and Item Level Analysis of the Spanish NUMi Candidate Items

Item #	Item Description	CTT Difficulty	CTT Discrimination	IRT Difficulty	IRT Discrimination
1	Serum glucose goal for patient with diabetes	0.85	0.34	−1.28	1.37
2	Using a pain scale	0.86	0.50	−1.09	1.40
3	Interpreting percentage format for risk information	0.76	0.44	−0.71	1.32
4	Interpreting frequency format for risk information	0.76	0.50	−0.85	0.87
5	Interpreting small risk estimates with percent and frequency formats	0.49	0.56	0.20	0.62
6	Interpreting nutrition label	0.82	0.43	−1.31	0.73
7	Prostate screening test	0.46	0.48	0.002	0.81
8	Statistical significance	0.27	0.53	1.19	0.85

CTT Difficulty: Classical test theory difficulty calculated as % correct; CTT Discrimination calculated as item-total correlation. IRT Difficulty and Discrimination: Item response theory parameters determined by two-parameter model. Items included on each test. Items 7 and 8 were not included in the Spanish NUMi. The calibration of these items was done using the software BILOG.[46]
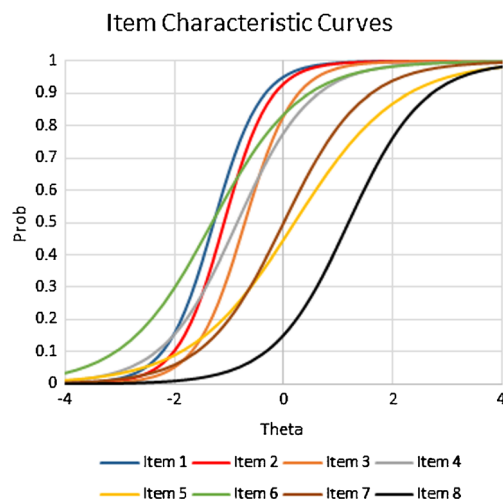
Figure 1. Item Characteristic curves. This figure displays the item information curves for the 8 candidate items for the Spanish NUMi. Items 1–6 were including in the Spanish NUMi. The item content is presented in English in Table 2. The IRT difficulty and discrimination parameters that were used to determine these curves are presented in Table 3. The curves that have their inflection point at a higher Theta value indicate harder items.

1.80. The distribution of scores demonstrates is slightly negatively skewed in this sample of examinees (Online Fig. B).

Short Form Content

Despite the fact that the short form consists of only six items, these six items reflect a range of domains. Four items on the final shortened version of the NUMi are from the number sense domain, one item is from the tables and graphs domain, and one item is from the probability domain (Table 2).
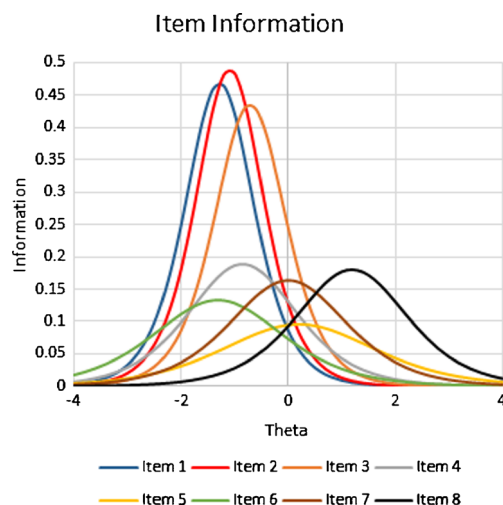

Figure 2. Item information curves. This figure displays the test information function for each candidate item of the Spanish NUMi. Items 1–6 were including in the Spanish NUMi. The item content is presented in English in Table 2. The IRT difficulty and discrimination parameters that were used to determine these curves are presented in Table 3. The curves that have the greatest height indicate items that have the highest discrimination. The Theta value at which the curve peaks indicates the difficulty level at which the item is most discriminating.
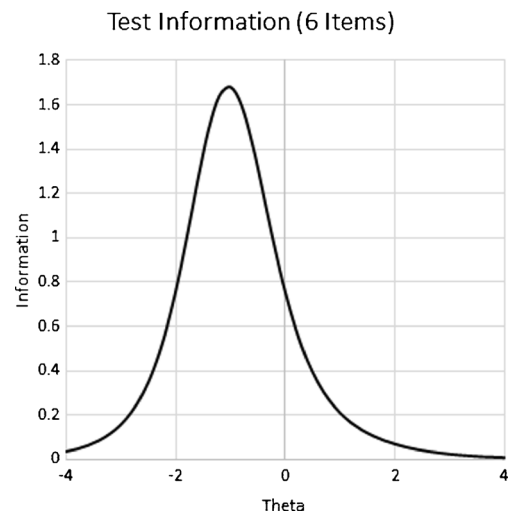

Figure 3. Test Information Function of Spanish NUMi. This figure indicates the level of the numeracy trait, theta, at which the six-item Spanish NUMi test will be most discriminating.

DISCUSSION

We have documented that the Spanish NUMi is a valid measure of the range of important numerical concepts that are used in understanding and communicating health information. While there are other Spanish measures of health numeracy, they have important limitations that the Spanish NUMI does not have. The Spanish NUMi is an important addition to the field, as we know that low health numeracy and literacy are more common in Spanish-speaking populations in the United States, placing this population at greater risk of health disparities. In addition, this tool is short, easy to use, and focuses on health numeracy and the range of numerical skills that individuals need to understand to effectively and actively participate in their health care.

The rigorous development of the NUMi for use in Spanish and English should not be overlooked. We found that some terms and concepts were difficult to translate linguistically and culturally, highlighting the importance of excellent translation and a need for future research to focus on how to overcome these issues.

Measures of health numeracy may be useful in guiding communication strategies in health that are tailored to individuals or targeted to populations.[18] For example, the Spanish NUMi could be given to patients at initial registration for care or while waiting for a visit and information about the patient's numeracy given to providers so they can adjust their communication accordingly. If patients are found to have low health numeracy, providers could use fewer numbers and/or spend more time explaining numerical health information or instructions. Additionally, this additional information on a patient's level of health numeracy could cue the provider to use diagrams, tables, pictographs, and frequencies, rather than percentages, when trying to communicate risks, benefits and other treatment information to a patient with low health numeracy.[31]

Our study was not without limitations. Our sample was predominantly Mexican Spanish speakers from the Midwest.

While this might limit generalizability to other Spanish-speaking populations, we do not think this is likely because we used a group translation method in which Spanish speakers from Mexico and Central and South America were involved. Second, the study met the psychometric requirements of IRT (e.g., a large item bank that was originally calibrated using a large sample). However, our validation is based on one sample. Use of the Spanish NUMi among broader populations is needed to support the predictive validity of the Spanish NUMi with respect to communication, decision making, and health outcomes. Further work is also needed to support meaningful categories of scores that could define low, medium, and how numeracy groups.

Despite these limitations, we were able to document that the Spanish NUMi is a reliable and valid measure of the range of important numerical concepts that are used in understanding and communicating health information. The Spanish NUMi is an easy to use, brief measure of literacy that could be used in Spanish speakers in a large range of settings to assess the health numeracy, and guide communication and care to be more effective and accessible for low-literate Spanish-speaking patients.

Corresponding Author: Elizabeth A. Jacobs, MD AMPP; Department of Medicine, School of Medicine and Public HealthUniversity of Wisconsin, Madison, WI 53705, USA (e-mail: eajacobs@medicine.wisc.edu).

Compliance with Ethical Standards:

Conflict of Interest: The authors have no conflicts of interest to report.

REFERENCES

1. **French MG**. Health literacy and numeracy workshop summary. Washington D.C. 2014. Report No.
2. **Ancker JS, Kaufman D.** Rethinking health numeracy: a multidisciplinary literature review. J Am Med Inform Assoc. 2007;14:713–21.
3. **Golbeck AL, Ahlers-Schmidt CR, Paschal AM, Dismuke SE.** A definition and operational framework for health numeracy. Am J Prev Med. 2005;29:375–6.
4. **Lipkus IM, Peters E.** Understanding the role of numeracy in health: proposed theoretical framework and practical insights. Health Educ Behav. 2009;36:1065–81.
5. **Schapira MM, Fletcher KE, Gilligan MA, et al.** A framework for health numeracy: how patients use quantitative skills in health care. J Health Commun. 2008;13:501–17.
6. **Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K.** Low health literacy and health outcomes: an updated systematic review. Ann Intern Med. 2011;155:97–107.
7. **Davids SL, Schapira MM, McAuliffe TL, Nattinger AB.** Predictors of pessimistic breast cancer risk perceptions in a primary care population. J Gen Intern Med. 2004;19:310–5.
8. **Gurmankin AD, Baron J, Armstrong K.** Intended message versus message received in hypothetical physician risk communications: exploring the gap. Risk Anal. 2004;24:1337–47.
9. **Garcia-Retamero R, Galesic M.** Who profits from visual aids: overcoming challenges in people's understanding of risks [corrected]. Soc Sci Med. 2010;70:1019–25.
10. **Gardner PH, McMillan B, Raynor DK, Woolf E, Knapp P.** The effect of numeracy on the comprehension of information about medicines in users of a patient information website. Patient Educ Couns. 2011;83:398–403.
11. **Schwartz LM, Woloshin S, Black WC, Welch HG.** The role of numeracy in understanding the benefit of screening mammography. Ann Intern Med. 1997;127:966–72.
12. **Weinfurt KP, Castel LD, Li Y, et al.** The correlation between patient characteristics and expectations of benefit from Phase I clinical trials. Cancer. 2003;98:166–75.
13. **Hibbard JH, Peters E, Dixon A, Tusler M.** Consumer competencies and the use of comparative quality information: it isn't just about literacy. Med Care Res Rev. 2007;64:379–94.
14. **Peters E, Levin IP.** Dissecting the risky-choice framing effect: Numeracy as an individual-difference factor in weighting risky and riskless options. Judgm Decis Mak. 2008;3:435–48.
15. **Apter AJ, Cheng J, Small D, et al.** Asthma numeracy skill and health literacy. J Asthma. 2006;43:705–10.
16. **Estrada CA, Martin-Hryniewicz M, Peek BT, Collins C, Byrd JC.** Literacy and numeracy skills and anticoagulation control. Am J Med Sci. 2004;328:88–93.
17. **Cavanaugh K, Huizinga MM, Wallston KA, et al.** Association of numeracy and diabetes control. Ann Intern Med. 2008;148:737–46.
18. **Hamm RM, Bard DE, Hsieh E, Stein HF.** Contingent or universal approaches to patient deficiencies in health numeracy. Med Decis Mak. 2007;27:635–7.
19. **Leyva M, Sharif I, Ozuah PO.** Health literacy among Spanish-speaking Latino parents with limited English proficiency. Ambul Pediatr. 2005;5:56–9.
20. **Singleton K, Krause EMS.** Understanding Cultural and Linguistic Barriers to Health Literacy. Ky Nurse. 2010;58:4–9. **Reprinted from the Online Journal of Issues in Nursing, 14(3), Manuscript 4, Sept. 30, 2009.**
21. **Smith PC, Brice JH, Lee J.** The relationship between functional health literacy and adherence to emergency department discharge instructions among Spanish-speaking patients. J Natl Med Assoc. 2012;104:521–7.
22. **Williams MV, Parker RM, Baker DW, et al.** Inadequate functional health literacy among patients at two public hospitals. JAMA. 1995;274:1677–82.
23. **Literacy H.** A perscription to end confusion. Committee on Health Literacy. Institute of Medicine. Washington DC: The National Academy Press; 2004.
24. **Lindau ST, Tomori C, Lyons T, Langseth L, Bennett CL, Garcia P.** The association of health literacy with cervical cancer prevention knowledge and health behaviors in a multiethnic cohort of women. Am J Obstet Gynecol. 2002;186:938–43.
25. **Mazor KM, Roblin DW, Williams AE**, et al. Health literacy and cancer prevention: it's not what you say it's what they hear. Available at: http://works.bepress.com/sarah_cutrona/27. Accessed April 20, 2016
26. **Oldach BR, Katz ML.** Health literacy and cancer screening: a systematic review. Patient Educ Couns. 2014;94:149–57.
27. **Roman L, Meghea C, Ford S, et al.** Individual, provider, and system risk factors for breast and cervical cancer screening among underserved Black, Latina, and Arab women. J Womens Health (Larchmt). 2014;23:57–64.
28. **Wang J, Moehring J, Stuhr S, Krug M.** Barriers to colorectal cancer screening in Hispanics in the United States: an integrative review. Appl Nurs Res. 2013;26:218–24.
29. **Pandya C, McHugh M, Batalova J**. Limited english proficient individuals in the united states: number, share, growth, and linguistic diversity. Washington DC 2011. Report No.
30. **Schapira MM, Fletcher KE, Ganschow PS, et al.** The meaning of numbers in health: exploring health numeracy in a Mexican-American population. J Gen Intern Med. 2011;26:705–11.
31. **Fagerlin A, Zikmund-Fisher BJ, Ubel PA, Jankovic A, Derry HA, Smith DM.** Measuring numeracy without a math test: development of the subjective numeracy scale. Med Decis Mak. 2007;27:672–80.
32. **Huizinga MM, Elasy TA, Wallston KA, et al.** Development and validation of the diabetes numeracy test (DNT). BMC Health Serv Res. 2008;8:96.
33. **Lipkus IM, Samsa G, Rimer BK.** General performance on a numeracy scale among highly educated samples. Med Decis Mak. 2001;21:37–44.
34. **Osborn CY, Cavanaugh K, Wallston KA, Rothman RL.** Self-efficacy links health literacy and numeracy to glycemic control. J Health Commun. 2010;15(Suppl 2):146–58.
35. **Schwartz LM, Woloshin S, Welch HG.** Can patients interpret health information? An assessment of the medical data interpretation test. Med Decis Mak. 2005;25:290–300.

36. **Weiss BD, Mays MZ, Martz W, et al.** Quick assessment of literacy in primary care: the newest vital sign. Ann Fam Med. 2005;3:514–22.

37. **Zikmund-Fisher BJ, Smith DM, Ubel PA, Fagerlin A.** Validation of the subjective numeracy scale: effects of low numeracy on comprehension of risk communications and utility elicitations. Med Decis Mak. 2007;27:663–71.

38. **Yin HS, Sanders LM, Rothman RL, et al.** Assessment of health literacy and numeracy among Spanish-Speaking parents of young children: validation of the Spanish parental health literacy activities test (PHLAT Spanish). Acad Pediatr. 2012;12:68–74.

39. **Parker RM, Baker DW, Williams MV, Nurss JR.** The test of functional health literacy in adults: a new instrument for measuring patients' literacy skills. J Gen Intern Med. 1995;10:537–41.

40. **White RO 3rd, Osborn CY, Gebretsadik T, Kripalani S, Rothman RL.** Development and validation of a Spanish diabetes-specific numeracy measure: DNT-15 Latino. Diabetes Technol Ther. 2011;13:893–8.

41. **Nguyen TH, Paasche-Orlow MK, Kim MT, Han HR, Chan KS.** Modern measurement approaches to health literacy scale development and refinement: overview, current uses, and next steps. J Health Commun. 2015;20(Suppl 2):112–5.

42. **Schapira MM, Walker CM, Cappaert KJ, et al.** The numeracy understanding in medicine instrument: a measure of health numeracy developed using item response theory. Med Decis Mak. 2012;32:851–65.

43. **Schapira MM, Walker CM, Miller T, et al.** Development and validation of the numeracy understanding in Medicine Instrument short form. J Health Commun. 2014;19(Suppl 2):240–53.

44. **Gierl MJ.** Using dimensionality-based DIF analyses to identify and interpret constructs that elicit group differences. Educ Meas Issue Pract. 2005;24:3–14.

45. **Shealy R, Stout W.** A model-based standardization approach that separates true bias/DIF from group ability differences and detects test bias/DTF as well as item bias/DTF. Psychometrika. 1993;58:159–94.

46. **Zimowski MF, Muraki E, Mislevy R, Cock RD.** BILOG- MG. Chicago, IL: Scientific Software International; 1996.

47. **Stout WF.** A non-parametric approach for assessing latent trait unidimensionality. Psychometrika. 1987;52:589–617.

48. **Stout W.** A nonparametric approach for assessing latent trait unidemensionality. Psychometrika. 1987;52:589–617.

49. **Edelen MO, Reeve BB.** Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. Qual Life Res. 2007;16(Suppl 1):5–18.