Published in final edited form as:

J Health Commun. 2014 October ; 19(0 2): 240–253. doi:10.1080/10810730.2014.933916.

Development and Validation of the Numeracy Understanding in Medicine Instrument Short Form

Marilyn M. Schapira,

Department of Medicine, Perelman School of Medicine, University of Pennsylvania and the Center for Health Equity Research Program, Philadelphia VA Medical Center, Philadelphia, PA

Cindy M. Walker,

Department of Educational Psychology, University of Wisconsin-Milwaukee, Milwaukee, WI

Tamara Miller,

Department of Educational Psychology, University of Wisconsin-Milwaukee, Milwaukee, WI

Kathlyn A Fletcher,

Department of Medicine, Medical College of Wisconsin and the Zablocki VA Medical Center, Milwaukee, WI

Pamela G. Ganschow,

Department of Medicine, School of Medicine, Rush University, Chicago, IL

Elizabeth A Jacobs,

Department of Medicine, School of Medicine, University of Wisconsin, Madison, WI

Diana Imbert,

Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

Maria O'Connell, and

Department of Medicine, School of Medicine, Rush University, Chicago, IL

Joan M. Neuner

Department of Medicine, Medical College of Wisconsin, Milwaukee, WI

Abstract

Background—Health numeracy can be defined as the ability to understand and use numeric information and quantitative concepts in the context of health. We previously reported the development of the Numeracy Understanding in Medicine Instrument (NUMi); a 20-item test developed using item response theory. We now report the development and validation of a short form of the NUMi.

Methods—Item statistics were used to identify a subset of 8-items representing a range of difficulty and content areas. Internal reliability was evaluated with Cronbach's alpha. Divergent and convergent validity was assessed by comparing scores of the S-NUMI with existing measures of education, print and numeric health literacy, mathematic achievement, cognitive reasoning, and the original NUMi.

Results—The 8-item scale had adequate reliability (Cronbach's alpha: 0.72) and was strongly correlated to the 20-item NUMi (0.92). The S-NUMi scores were strongly correlated with the Lipkus numeracy test (0.62), Wide Range of Achievement Test-Mathematics (WRAT-M) (0.72), and Wonderlic cognitive reasoning test (0.76). Moderate correlation was found with education level (0.58) and print literacy as measured by the TOFHLA (0.49).

Conclusion—The short Numeracy Understanding in Medicine Instrument is a reliable and valid measure of health numeracy feasible for use in clinical and research settings.

Introduction

Numeracy can be defined as one's ability to understand and use numeric information conveyed in numbers, tables and graphs, probability, and statistics in the context of caring for your health and making health care decisions[1]. The hierarchical nature of health numeracy skills has been well described and includes a range of skills, from basic arithmetic to a more conceptual understanding of concepts such as risk, probability, and statistics[2-5]. The ability to measure the robust construct of health numeracy is important in the design of patient health education, self-management, and shared decision making interventions. The length and cognitive demands of health numeracy instruments that captures that range of skills, however, may be a barrier to their use in clinical or research settings.

The measurement of patient health numeracy is relevant to clinical practice and research for several reasons. First, a measure of patient health numeracy at the point of care would facilitate the alignment of patient health numeracy and the numeracy demand of education and decision support materials[6]. Second, screening for health numeracy could guide the communication strategies used in the doctor-patient visit [7, 8]. When alerted to a patient with a low health numeracy level, the clinician might take additional time when conveying numeric information and/or use a talk back approach that is directed at the numeric information conveyed[9, 10]. Third, if a patient is identified as having a low health numeracy level, interventions could be designed to overcome patient numeracy skills that are specific to a given clinical context or clinical goal [11]. The use of universal precautions that incorporate best practices for the clear presentation of health information is a sound approach[12]. However, we posit that individual level health numeracy assessments, when feasible, can inform additional tailored communication strategies and research related to those strategies. The use of health numeracy assessments to tailor communication strategies must consider the potential interaction between stress, learning and test performance [13, 14].

A range of general health numeracy measures have been developed [15-25]. Measures vary in their area of focus with some assessing skills in probability and risk communication [15, 16], some in the interpretation of medical evidence[26, 27], and some in basic numeric skills needed to accomplish day to day tasks in health such as taking medication or interpreting a nutrition label [22, 25]. The Numeracy Understanding in Medicine Instrument was developed as a robust construct of health numeracy including number sense, tables & graphs, probability and statistics.

Existing numeracy scales also differ with respect to the use of objective or subjective assessments. Objective measures assess knowledge or skills through items that have correct and incorrect responses. Subjective measures, in contrast, assess perceived aptitude and preferences for use of numeric information. The Subjective Numeracy Scale (SNS) is an 8-item measure that assesses beliefs about skill in the use of fractions and percentages and preferences regarding the use of numbers, tables, graphs, and statistics in daily life. It was positively correlated with objective measures of health numeracy but found to be completed in less time and less stressful to respondents [18]. The degree of correlation between subjective and objective health numeracy measures varies across measures [18] [27]. The goal of the current study was to develop a short objective health numeracy measure that presents a low respondent burden while assessing a broad range of health numeracy skills.

The NUMi used item response theory as a psychometric scaling method in order to take advantage of the strength of item level analysis to create an efficient measure of the complex construct of health numeracy [17]. Item response theory uses mathematical modeling to describe the relationship between a respondent's score on the construct being measured and the probability of responding to an item correctly [28]. In a two parameter model, such as used in this study, each item is characterized with respect to a difficulty and discrimination parameter. The item level analysis in IRT has several advantages in comparison to classical test theory (CTT) methods. Item selection and test formation can be adapted for various groups or objectives based upon the level of precision required and the respondent burden acceptable in a given context. The use of IRT methods also allows for determination of differential item functioning (DIF) across groups. A limitation of the 20-item NUMi, however, is the respondent burden due to the length and cognitive demands of the items. The purpose of the current study is to develop and validate a shortened version of the NUMi to increase feasibility of use while retaining the validity of the original instrument [29, 30].

Methods

Development of the Original NUMi

The development and validation methods used for the 20-item NUMi have been previously reported [1]. In summary, the NUMi was developed using item response theory scaling methods. Items were generated based upon qualitative studies [5, 31] and tested in two forms among a large (n=1000) and diverse population of community based adults and primary care patients (Table 1). Respondents were recruited with a range of methods including advertisements in community newspapers, and flyers in community centers, local colleges, and primary care clinical settings. Responses to test items were analyzed using a 2-parameter logistic (2-PL) IRT model to determine the difficulty and discrimination parameters for each item. Goodness of fit statistics found the 2-parameter model to fit our data better than a 1 parameter model in the validation study of the 20 item NUMi [1]. Based on an item level analysis and the distribution of content across domains, a 20-item test was formed and validated against existing measures of health print literacy, health numeracy, mathematics achievement, cognitive ability, and education. The 20-item test was found to demonstrate unidimensionality using the Stout's test of essential unidimensionality (DIMTEST). There was no differential item functioning (DIF) between Hispanics compared

to Non-Hispanics or Blacks compared to Hispanics and Non-Hispanic whites as determined by the test of simultaneous item bias (SIBTEST)[32]. The psychometric data for the current study was a sample of the psychometric data used for the original development of the 20-item NUMi.

Selection of Items for the Short NUMi

Items from the original NUMi were sorted by their difficulty parameters and placed in 4 levels. Within levels of difficulty, selection was based on discrimination parameters and item content. Items that are more discriminating are better able to differentiate between examinees at different ability levels. Items with the highest level of discrimination in each difficulty level were chosen. Consideration to balance of content of items was also a factor in item selection. A series of candidate short forms were selected with 5, 6, and 8 items, respectively. Each candidate short form included at least 1 item from each level of difficulty. Reliability and validity statistics were then calculating using the parameters determined from the original testing of the NUMi item bank and consistent with accepted methods of instrument refinement using IRT [1, 33].

Reliability and Validity Testing

Internal reliability was evaluated using Cronbach's alpha and the correlation of scores with the original 20-item NUM was determined for the three sets of items. The shortest set that provided adequate psychometric properties was determined and further evaluated for convergent and discriminant validity. Scores of the short NUMi were compared to levels of education, print literacy as measured by the Test of Functional Health Literacy in Adults (TOFHLA)[25], Lipkus expanded numeracy scale (Lipkus) [16], Wide Range Achievement Test in Mathematics (WRAT-M)[34], and Wonderlic cognitive reasoning test (Wonderlic) [35].

Results

Study Population and Short Form Selection

The study population was diverse in race, ethnicity, age, gender, and level of education (Table 1). The 8-item test demonstrating adequate reliability with a Cronbach's alpha of 0.72. Cronbach's alpha is equivalent to Kuder-Richardson Coefficient (KR-20) when the data is binary[36]. The 8-item test was strongly correlated with the original NUMi (0.92). Of the three short test sets evaluated, only the 8-item test demonstrated adequate internal reliability (Table 2).

Short Form Content

The content of the 8-item test was representative of the domains included in the theoretical construct. The 8-item test consists of items from the four of the content domains of number sense, table & graphs, probability, and statistics. Four items on the final shortened version of the NUMi are from the number sense domain, one item is from the Tables & Graphs domain, two items are from the probability domain, and one item is from the statistics domain (Table 3).

Item Analysis and Short Form Validity

Items demonstrated a range of difficulty using classical test statistics (0.27 to 0.86) and adequate discrimination (0.34-0.49). Similarly, IRT parameters indicate a range in the difficulty parameter. The easiest item had a difficulty parameter of -1.28 (determining if a glucose level was within goal range) and the hardest item had a difficulty parameter of 1.19 (using percentages to understand small risks). Difficulty parameters typically range from -3.0 (hardest items) to +3.0 (easiest items). The discrimination parameters of the 8-item form ranged from a low of 0.623 to a high of 1.396 (Table 3). Items with discrimination between the values of 0.80 and 2.50 have been thought to represent effective discriminatory power[37]. Our item bank had fewer high discriminating items in the upper two levels of difficulty, necessitating the inclusion of one item with a discrimination level of 0.623 for the fourth highest level of difficulty. The items ranged from most discriminating (Item #2, interpretation of a scale, discrimination parameter of 1.396) to least discriminating (Item #5, interpreting the risks of side effects, discrimination parameter of 0.623).

A comparison of test scores between the S-NUMi and other constructs related to education, ability, and health literacy provide support for the validity of the S-NUMi (Table 4). A strong correlation was found between the S-NUMi and education level (0.60), mathematical achievement as noted by the WRAT-M (0.72), Lipkus (0.62), and Wonderlic reasoning test (0.76). In contrast, there was a more moderate correlation with print literacy as measured by the TOFHLA (0.49). The 8-item test had a mean of 61.9% correct (SD 26.7) with a normative distribution of scores (Figure 1).

Scoring the S-NUMi

The S-NUMi is a continuous variable. Therefore, it is recommended that scores be used as a continuous measure with a potential range of 0 to 8. For some purposes it may be helpful to use cut-off values that indicate how an individual is performing compared to normative data. We therefore, suggest a second scoring strategy based upon test performance in our study population. Further evaluation of these cut-off levels in future studies is needed to validate the cut-off levels we propose. The mean of the S-NUMi in our study population is 4.95 or approximately 5 and the standard deviation is 2.13, or approximately 2. We propose that total scores on the S-NUMi that are less than or equivalent to 3, approximately one standard deviation below the mean, are interpreted as low numeracy; total scores that are greater than or equivalent to 7, approximately one standard deviation above the mean, are interpreted as high numeracy; and scores that are 4, 5, or 6 are considered as average numeracy. Using these proposed guidelines with our current study would result in 121 respondents (25%) classified as low numeracy, 228 respondents (48%) classified as average numeracy, and 131 (27%) classified as high numeracy.

Discussion

We report the development and validation of a short version of the Numeracy Understanding in Medicine Instrument (S-NUMi). We find that the short version retains high reliability and validity in comparison with the original 20-item test. The NUMi is based on a robust construct of health numeracy that includes number sense, tables and graphs, probability, and

statistics [1]. The development of the short form takes advantage of IRT scaling methods to make parsimonious decision regarding items to include in the shortened version. The original NUMi was limited in its application due to the length and cognitive demands of the test. The S-NUMi is a good choice for clinicians and investigators who want to have a comprehensive measure of objective health numeracy, while limiting the respondent burden.

The S-NUMi has some similarities to existing objective measures of health numeracy that were developed using classical test theory. The Test of Functional Health Literacy in Adults (TOFHLA) and the short TOFHLA (S-TOFHLA) include numeracy sections of 17 and 4 items, respectively[25, 38] that measure skills such as interpreting a prescription bottle, monitoring blood glucose, and scheduling an appointment. The assessment is interactive, asking for responses to cue cards or labeled prescription bottles and is available in English and Spanish. The Newest Vital Sign (NVS) also provides a composite assessment of print and numeric health literacy through 6 items that ask the respondent to interpret a nutrition label[22]. The TOFHLA, S-TOFHLA, and the NVS require that the tool is administered by an interviewer. The Lipkus expanded numeracy scale is an 11 item test that can be self-administered, with items that assess understanding of risk and probability[39]. Our measure differs from these existing scales in the scope of numeracy skills it seeks to measure and the psychometric methods used in development. The optimal choice of a numeracy measure will vary with the goals and structure of the research and clinical context in which it will be applied. The S-NUMi provides an objective measure of the broad health numeracy construct with a low respondent burden.

Review of the content of items included in the S-NUMi reflects the scope of this numeracy assessment. Each item reflects a numeracy skill and a clinical context. Skills assessed include understanding numeric range, understanding how to use a scale, ordering whole numbers and frequencies with respect to magnitude, interpreting a probability using percentage formats, interpreting a P-value, and reading a chart. The clinical contexts of the questions include diabetes management, pain control, cardiology test interpretation, side effects of medications, false positive prostate cancer screening tests, interpretation of scientific evidence, and the interpretation of a nutrition label. Although not every context and skill in the full item bank could be included, the scope of items is robust with respect to numeric skills and clinical context, providing face validity for the measure. Further, results of previous dimensionality analyses, based on IRT methods, allow us to relax rules of distribution of content domains.

The role of a valid health numeracy assessment is increasingly evident. A growing literature highlights that health numeracy is a distinct construct from general health literacy or print literacy[40]. Health numeracy encompasses skills that are essential for health communication, self-management, and decision making [2-4, 31, 41, 42]. Our basic understanding of disease processes often require an understanding of numeric concepts, such as using number to represent degrees of severity of disease processes or using numbers to set and recognize achievement of individual health goals[5, 31]. In addition to the important uses of basic number ordering and arithmetic operations, the construct of health numeracy includes a conceptual understanding of constructs such as risk and probability. Outcomes in medicine are uncertain and, to varying degrees, unpredictable. It is important to assess a

patient's understanding of these concepts so that information that involves uncertainty (such as prognosis or diagnostic accuracy) can be communicated appropriately for each patient. Finally, as care moves towards a patient centered approach, there is more effort being made to share scientific evidence with patients [43]. To this end, there is a need to know whether basic principles of scientific study design and strength of evidence are understood. Evidence suggests that this aspect of scientific literacy is varied among members of the public [44, 45]. This knowledge could help clinicians and health educators tailor their communication strategies accordingly. Although it is not feasible to assess all aspects of this knowledge in one short measure, the psychometric methods used to develop the NUMi, and now the S-NUMi, strengthen its validity as a robust measure of the health numeracy construct.

Additional research is needed to evaluate the efficacy and outcomes of screening for health numeracy in the clinical setting. A debate continues regarding the use of communication strategies tailored to the level of health numeracy in comparison to a uniform approach for the development of health education materials [10]. Studies report that the effect of risk communication strategies such as the use of a risk ladder, ratio, or pictogram differs across level of health numeracy [7, 46, 47]. Our qualitative work suggests that more conceptual domains of health numeracy including probability and statistics are viewed by both clinicians and patients as important for a full understanding of important concepts in medicine[5, 31]. Therefore, caution must be used in limiting communication strategies to only basic numeric concepts. Use of the S-NUMi as could facilitate research to test the efficacy of communication strategies tailored to level of health numeracy.

There are some limitations to the S-NUMi. First, objective measures of health numeracy may be more burdensome to obtain than subjective measures even when shorter in length [18]. Objective measures, however, may also be a more accurate reflection of how well patients can accomplish numeric related tasks. The shorter measure presented in this paper should increase feasibility of administration. Second, although the study met the psychometric requirements of IRT (e.g. a large item bank and testing among a large sample which was diverse in race, ethnicity, and education)[30] our validation study is based on a single sample of respondents. Use of the S-NUMi among different populations in future studies is needed to support the predictive validity of the S-NUMi with respect to medical outcomes.

In summary, we present the development and validation of the short form of the Numeracy Understanding in Medicine Instrument. The S-NUMi validation study suggests a robust measure of health numeracy for use in general and clinical populations. Future studies are needed to better understand its performance among diverse populations and clinical settings. We anticipate that the S-NUMi will be a useful tool to assess understanding of numeric concepts in clinical populations and support the alignment of communication strategies with a patient's ability to comprehend numeric Information.

Acknowledgments

Funding: This work was supported by grants from the National Cancer Institute of the National Institutes of Health NC1R01CA115954 and the American Cancer Society 121158-RSG-11-104-01-CPPB.

Appendix: Numeracy Understanding in Medicine Instrument: Short Form

1. 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?
 - a. 55 mg/dL
 - b. 140 mg/dL
 - c. 165 mg/dL
 - d. 180 mg/dL
2. 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?
 - a. 3
 - b. 5
 - c. 7
 - d. 9
3. 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?
 - a. 33%
 - b. 50%
 - c. 75%
 - d. 99%
4. Natasha started a new medicine that may cause the side effects listed below. Which side effect is Natasha least likely to have?

Side Effect	Chance of Occurring
a Dizziness	1 in 5 people
b Nausea	1 in 10 people
c Stomach pain	1 in 100 people
d Allergic reaction	1 in 200 people
5. James starts a new blood pressure medicine. The chance of a serious side effect is 0.5%. If 1000 people take this medicine, about how many would be expected to have a serious side effect?
 - a. 1 person
 - b. 5 people
 - c. 50 people

- d. 500 people
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. 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. 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

Nutrition Facts Serving Size 1 cup (228g) Servings per Container 2	
Amount Per Serving	
Calories 280	Calories from Fat 120
	% Daily Value*
Total Fat 13g	20%
Saturated Fat 5g	25%
Trans Fat 2g	
Cholesterol 2mg	10%
Sodium 660 mg	28%
Total Carbohydrate 31g	10%
Dietary Fiber 3g	0%
Sugars 5g	

Nutrition Facts Serving Size 1 cup (228g) Servings per Container 2	
Amount Per Serving	
Calories 280	Calories from Fat 120
Protein 5g	
Vitamin A 4%	Vitamin C 2%
Calcium 15%	Iron 4%

* Percent Daily Values are based on a 2,000-calorie diet. Your Daily values may be higher or lower depending on your calorie needs

References

- Schapira MM, et al. The numeracy understanding in medicine instrument: a measure of health numeracy developed using item response theory. *Med Decis Making*. 2012; 32(6):851–65. [PubMed: 22635285]
- Golbeck AL, et al. A definition and operational framework for health numeracy. *Am J Prev Med*. 2005; 29(4):375–6. [PubMed: 16242604]
- Apter AJ, et al. Numeracy and communication with patients: they are counting on us. *J Gen Intern Med*. 2008; 23(12):2117–24. [PubMed: 18830764]
- Ancker JS, Kaufman D. Rethinking health numeracy: a multidisciplinary literature review. *J Am Med Inform Assoc*. 2007; 14(6):713–21. [PubMed: 17712082]
- Schapira MM, et al. A framework for health numeracy: how patients use quantitative skills in health care. *J Health Commun*. 2008; 13(5):501–17. [PubMed: 18661390]
- Joram E, et al. The numeracy demands of health education information: an examination of numerical concepts in written diabetes materials. *Health Commun*. 2012; 27(4):344–55. [PubMed: 21916698]
- Peters E, Hart PS, Fraenkel L. Informing patients: the influence of numeracy, framing, and format of side effect information on risk perceptions. *Med Decis Making*. 2011; 31(3):432–6. [PubMed: 21191122]
- Clement S, et al. Complex interventions to improve the health of people with limited literacy: A systematic review. *Patient Educ Couns*. 2009; 75(3):340–51. [PubMed: 19261426]
- Howard T, Jacobson KL, Kripalani S. Doctor talk: physicians' use of clear verbal communication. *J Health Commun*. 2013; 18(8):991–1001. [PubMed: 23577746]
- Hamm RM, et al. Contingent or universal approaches to patient deficiencies in health numeracy. *Med Decis Making*. 2007; 27(5):635–7. [PubMed: 17921452]
- Moore JO, et al. Designing interventions to overcome poor numeracy and improve medication adherence in chronic illness, including HIV/AIDS. *J Med Toxicol*. 2011; 7(2):133–8. [PubMed: 21455810]
- DeWalt DA, et al. Developing and testing the health literacy universal precautions toolkit. *Nurs Outlook*. 2011; 59(2):85–94. [PubMed: 21402204]
- Meyer T, et al. Acute stress differentially affects spatial configuration learning in high and low cortisol-responding healthy adults. *European Journal of Psychotraumatology*. 2013; 4
- Espin L, et al. Acute pre-learning stress and declarative memory: impact of sex, cortisol response and menstrual cycle phase. *Horm Behav*. 2013; 63(5):759–65. [PubMed: 23587533]
- Schwartz LM, et al. The role of numeracy in understanding the benefit of screening mammography. *Ann Intern Med*. 1997; 127(11):966–72. [PubMed: 9412301]
- Lipkus IM, Samsa G, Rimer BK. General performance on a numeracy scale among highly educated samples. *Med Decis Making*. 2001; 21(1):37–44. [PubMed: 11206945]

17. Schapira MM, et al. The development and validation of the hypertension evaluation of lifestyle and management knowledge scale. *J Clin Hypertens (Greenwich)*. 2012; 14(7):461–6. [PubMed: 22747619]
18. Fagerlin A, et al. Measuring numeracy without a math test: development of the Subjective Numeracy Scale. *Med Decis Making*. 2007; 27(5):672–80. [PubMed: 17641137]
19. Apter AJ, et al. Asthma numeracy skill and health literacy. *J Asthma*. 2006; 43(9):705–10. [PubMed: 17092853]
20. Osborn CY, et al. Development and validation of the General Health Numeracy Test (GHNT). *Patient Educ Couns*. 2013; 91(3):350–6. [PubMed: 23433635]
21. Cokely ET, et al. Measuring Risk Literacy: The Berlin Numeracy Test. *Judgment and Decision Making*. 2012; 7(1):25–47.
22. Weiss BD, et al. Quick assessment of literacy in primary care: the newest vital sign. *Ann Fam Med*. 2005; 3(6):514–22. [PubMed: 16338915]
23. Huizinga MM, et al. Development and validation of the Diabetes Numeracy Test (DNT). *BMC Health Serv Res*. 2008; 8:96. [PubMed: 18452617]
24. Herndon MB, et al. Older patients perceptions of “unnecessary” tests and referrals: a national survey of Medicare beneficiaries. *J Gen Intern Med*. 2008; 23(10):1547–54. [PubMed: 18592324]
25. Parker RM, et al. The test of functional health literacy in adults: a new instrument for measuring patients' literacy skills. *J Gen Intern Med*. 1995; 10(10):537–41. [PubMed: 8576769]
26. Schwartz LM, Woloshin S, Welch HG. Can patients interpret health information? An assessment of the medical data interpretation test. *Med Decis Making*. 2005; 25(3):290–300. [PubMed: 15951456]
27. Woloshin S, Schwartz LM, Welch HG. Patients and medical statistics. Interest, confidence, and ability. *J Gen Intern Med*. 2005; 20(11):996–1000. [PubMed: 16307623]
28. Bjorner JB, et al. Developing tailored instruments: item banking and computerized adaptive assessment. *Qual Life Res*. 2007; 16 Suppl 1:95–108. [PubMed: 17530450]
29. Hays RD, Lipscomb J. Next steps for use of item response theory in the assessment of health outcomes. *Qual Life Res*. 2007; 16 Suppl 1:195–9. [PubMed: 17351825]
30. Gershon R, et al. Item response theory and health-related quality of life in cancer. *Expert Rev Pharmacoecon Outcomes Res*. 2003; 3(6):783–91. [PubMed: 19807355]
31. Schapira MM, et al. The meaning of numbers in health: exploring health numeracy in a Mexican-American population. *J Gen Intern Med*. 2011; 26(7):705–11. [PubMed: 21336671]
32. WF S. A non-parametric approach for assessing latent trait unidimensionality. *Psychometrika*. 1987; 52:589–617.
33. 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. [PubMed: 17375372]
34. Jastak, S.; W, G. Wide-Range Achievement Test-Revised 3. Wilmington, DE: Jastak Associates; 1993.
35. Matthews TD, Lassiter KS. What does the Wonderlic Personnel Test measure? *Psychological Reports*. 2007; 100(3):707–712. [PubMed: 17688083]
36. Cronbach LJ. My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement*. 2004; 64(3):391–418.
37. dA, RJ. The theory and practice of item response theory. New York, NY: Guilford Press; 2009.
38. Baker DW, et al. Development of a brief test to measure functional health literacy. *Patient Educ Couns*. 1999; 38(1):33–42. [PubMed: 14528569]
39. Lipkus IM, Samsa G, Rimer BK. General performance on a numeracy scale among highly educated samples. *Medical Decision Making*. 2001; 21(1):37–44. [PubMed: 11206945]
40. Golbeck A, et al. Correlating reading comprehension and health numeracy among adults with low literacy. *Patient Educ Couns*. 2011; 84(1):132–4. [PubMed: 20598469]
41. Reyna VF, et al. How numeracy influences risk comprehension and medical decision making. *Psychological Bulletin*. 2009; 135(6):943–73. [PubMed: 19883143]

42. Nelson W, et al. Clinical implications of numeracy: theory and practice. *Ann Behav Med*. 2008; 35(3):261–74. [PubMed: 18677452]
43. Carman KL, et al. Evidence that consumers are skeptical about evidence-based health care. *Health Aff (Millwood)*. 2010; 29(7):1400–6. [PubMed: 20522522]
44. Miller JD. Public understanding of, and attitudes toward, scientific research: what we know and what we need to know. *Public Understanding of Science*. 2004; 13(3):273–294.
45. Miller JD. The measurement of civic scientific literacy. *Public Understanding of Science*. 1998; 7(3):203–223.
46. Keller C, Siegrist M, Visschers V. Effect of risk ladder format on risk perception in high- and low-numerate individuals. *Risk Anal*. 2009; 29(9):1255–64. [PubMed: 19572963]
47. Keller C, Siegrist M. Effect of risk communication formats on risk perception depending on numeracy. *Med Decis Making*. 2009; 29(4):483–90. [PubMed: 19525484]

Figure 1. Distribution of Test Scores for 8-item NUMi

Table 1
Characteristics of the Study Population

NUMi Demographics	Total (n=1000)	Form A (n=520)	Form B (n=480)	Validation Sample (n=99)
	n (%)	n (%)	n (%)	n(%)
GENDER				
Male	399 (39.9)	204 (39.2)	195 (40.6)	34 (34.3)
Female	599 (59.9)	315 (60.6)	284 (59.2)	65 (65.7)
Missing	2 (0.2)	1 (0.2)	1 (0.2)	0
AGE				
< 45 years	536 (53)	277 (53.3)	259 (54)	51 (51.5)
45-59 years	329 (33)	178 (34.2)	151 (31.4)	36 (36.4)
60-74 years	119 (12)	58 (11.2)	61 (12.7)	11 (11.1)
75 years	16 (2)	7 (1.4)	9 (1.9)	1 (1.0)
RACE				
White	448 (44.8)	212 (40.8)	236 (49.2)	63 (63.6)
Black and/or African American	438 (43.8)	252 (48.5)	186 (38.7)	29 (29.3)
American Indian and Alaska Native	14 (1.4)	5 (1.0)	9 (1.9)	1 (1.0)
Asian	38 (3.8)	22 (4.2)	16 (3.3)	4 (4.0)
Native Hawaiian and Other Pacific Islander	3 (0.3)	2 (0.4)	1 (0.2)	0
Multiple Races	17 (1.7)	8 (1.5)	9 (1.9)	0
Missing	42 (4.2)	19 (3.7)	23 (4.8)	2 (2.0)
ETHNICITY				
Hispanic/Latino	290 (29)	137 (26.4)	153 (31.9)	38 (38.4)
Missing	7 (0.7)	3 (0.6)	4 (0.8)	1 (1.0)
TOFHLA Score				
0-16 (Inadequate health literacy)	43 (4)	19 (3.7)	24 (5)	2 (2.0)
17-22 (Marginal health literacy)	43 (4)	25 (4.8)	18 (3.8)	2 (2.0)
23-36 (Adequate health literacy)	914 (92)	476 (91.5)	438 (91.3)	95 (96.0)
EDUCATION				
Up to 12 years	409 (40.9)	215 (41.4)	194 (40.4)	31(31.3)
Some college	274 (27)	149 (28.7)	125 (26)	23(23.2)
Four year college or more	316 (32)	155 (29.8)	161 (33.5)	45 (45.5)
Missing	1 (0.1)	1 (0.2)	0 (--)	0
Wonderlic Cognitive Aptitude Test				
< 10	175 (17.5)	95 (18.3)	80 (16.7)	14 (14.1)
10-19	435 (43.5)	252 (48.4)	183 (38.1)	42 (42.4)
20-29	244 (24)	108 (20.8)	136 (28.3)	22 (22.2)
30-39	102 (10)	46 (8.8)	56 (11.7)	19 (19.2)
40-50	7 (1)	5 (1)	2 (0.4)	2 (2.0)
Missing	37 (4)	14 (2.7)	23 (4.8)	0 (--)

NUMi Demographics	Total (n=1000)	Form A (n=520)	Form B (n=480)	Validation Sample (n=99)
	n (%)	n (%)	n (%)	n(%)
Item Administration				
Read aloud to respondent	46 (5)	16 (3.1)	30 (6.3)	98 (99.0)
Self-administered by respondent	954 (95)	504 (96.9)	450 (93.7)	1 (1.0)

Four items of the S-NUMi were tested in the full study population (n=1000) as they were linked items (Item numbers 4, 6, 7, and 8. Four items were tested only as part of Form B (n=480) (items 1, 2, 3, and 5). One of the validation scales (Lipkus numeracy measure) was tested only in the validation sample (n=99). The remaining validation measures were in the full samples.

Table 2
Short test forms reliability and correlation with 20-item NUMi

Short Test Version	Cronbach's alpha	Correlation with 20-item NUMi
5-item	0.64	0.87
6-item	0.67	0.88
8-item	0.72	0.92

Table 3

Description and Item Level Analysis of the 8-Item NUMi

Item Number	Item Description	CTT Difficulty	CTT Discrimination	IRT Difficulty	IRT Discrimination
1	Range/blood sugar goal in diabetic	0.85	0.340	-1.277	1.367
2	Scale/reporting pain	0.86	0.479	-1.088	1.396
3	Ordering numbers/test results	0.76	0.413	-0.713	1.322
4	Frequency format/side effects	0.76	0.488	-0.653	1.269
5	Small risks/side effects	0.48	0.414	0.199	0.623
6	Calculating probability/screening tests	0.45	0.375	0.003	0.806
7	P-value/interpretation of study results	0.27	0.452	1.187	0.847
8	Reading a table/interpreting a nutrition label	0.82	0.488	-0.871	0.943

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 2-parameter model. Items included on each test: 5-item test: 2, 3, 4, 5, 7; 6-item test: addition of item 1; 8-item test: addition of items 6 and 8.

Table 4
Evidence for Construct Validity of the 8-Item Numi

Measure	n	Correlation	p-value
Education Level	480	0.60	<0.001
S-TOFHLA	480	0.49	<0.001
WRAT-M	83	0.72	<0.001
Lipkus	99	0.62	<0.001
Wonderlic	457	0.76	<0.001
Score on 20-item NUMi	480	0.92	< 0.001

Note: The items in the S-NUMi were in form B and tested among 480 respondents. The WRAT-M and Lipkus were obtained in a smaller validation sample (n=99).