Med Decis Making OnlineFirst, published on July 19, 2007 as doi:10.1177/0272989X07304449

Measuring Numeracy without a Math Test: Development of the Subjective Numeracy Scale

Angela Fagerlin, PhD, Brian J. Zikmund-Fisher, PhD, Peter A. Ubel, MD, Aleksandra Jankovic, MS, Holly A. Derry, MPH, Dylan M. Smith, PhD

Background. *Basic numeracy skills are necessary before patients can understand the risks of medical treatments. Previous research has used objective measures, similar to mathematics tests, to evaluate numeracy.* ***Objectives.*** *To design a subjective measure (i.e., self-assessment) of quantitative ability that distinguishes low- and high-numerate individuals yet is less aversive, quicker to administer, and more useable for telephone and Internet surveys than existing numeracy measures.* ***Research Design.*** *Paper-and-pencil questionnaires.* ***Subjects.*** *The general public ($N = 703$) surveyed at 2 hospitals.* ***Measures.*** *Forty-nine subjective numeracy questions were compared to measures of objective numeracy.* ***Results.*** *An 8-item measure, the Subjective Numeracy Scale (SNS), was developed through several rounds of testing. Four items measure people's beliefs about their skill in performing various mathematical operations, and 4 measure people's preferences regarding the presentation of numerical information. The SNS was significantly correlated with Lipkus and others' objective numeracy scale (correlations: 0.63–0.68) yet was completed in less time (24 s/item v. 31 s/item, $P < 0.05$) and was perceived as less stressful (1.62 v. 2.69, $P < 0.01$) and less frustrating (1.92 v. 2.88, $P < 0.01$). Fifty percent of participants who completed the SNS volunteered to participate in another study, whereas only 8% of those who completed the Lipkus and others scale similarly volunteered (odds ratio = 11.00, 95% confidence interval = 2.14–56.65).* ***Conclusions.*** *The SNS correlates well with mathematical test measures of objective numeracy but can be administered in less time and with less burden. In addition, it is much more likely to leave participants willing to participate in additional research and shows much lower rates of missing or incomplete data.* ***Key words:*** *numeracy; risk communication; decision making; literacy; measurement.* ***(Med Decis Making XXXX; XX:xx–xx)***

To make informed decisions, patients need to be aware of and understand the risks and benefits of their treatment options. However, many patients have substantial difficulty understanding such risks and benefits because of poor literacy and numeracy. According to the most recent National Adult Literacy Survey,[1] 21% of Americans scored in the lowest 2 levels of prose literacy (which corresponds to being able to make only low-level inferences that are easily found in a short passage). Equally alarming is the fact that 22% of Americans scored in the lowest 2 levels for quantitative literacy, a performance level that corresponds to having the ability to solve only single-operation arithmetic problems (e.g., what is the difference in cost between 2 items?). Not surprising, this latter figure varies with level of education; 51% of high school graduates, 16% of college graduates, and 11% of people with graduate training scored in the 2 lowest levels.

DOI: 10.1177/0272989X07304449

Although numerous studies have explored the impact of literacy on health behaviors and outcomes,[2–5] relatively little attention has been directed toward understanding the role of numeracy, which has been defined as aptitude with probabilities, fractions, and ratios. This is unfortunate because adequate numeracy skills are a necessary component of any complete understanding of the risks of health behaviors and medical treatments. Difficulty interpreting risk information could lead patients to make treatment decisions that are incompatible with their own preferences, resulting in a decline in patient care and satisfaction. In fact, research has shown that whereas most women overestimate the benefits of mammography, a greater proportion of women with lower numeracy (compared to women with higher numeracy) overestimated the benefits.[6] Furthermore, people with low numeracy have difficulty answering time tradeoff and standard gamble elicitations.[7] Other research has shown that low numeracy can impede comprehension and elicitation of health statistics.[6,8–10] These findings show that numeracy plays a critical role in people's medical decisions.

Measuring numeracy in a research setting can provide valuable information regarding how well an intervention works in various populations. Also, understanding how numeracy affects health behaviors is valuable in itself. For instance, if research were to find that low-numeracy patients choose riskier treatments, future research could test whether differences in behavior resulted from misunderstanding the statistical risks and benefits of the treatment or other differences. In addition, measuring numeracy may help explain one potential cause for racial disparities found in health care. In fact, the use of numeracy in research could be analogous to the manner in which literacy is currently being used in research.

Most previous studies that have examined the issue of numeracy have employed an objective measure, in the form of a list of mathematical problems that tested people's understanding of frequency, probability, and percentages. Schwartz and colleagues first assessed numerical ability using a set of 3 questions that tested people's ability to translate percentages into fractions and understanding of how often a fair coin would come up heads or a fair die would land on an even number.[6] More recently, Lipkus and colleagues evaluated a set of 11 numeracy questions, composed of the 3 questions previously developed by Schwartz and others and an additional 8 questions that focus more directly on the ability to understand disease risk (see Table 1).[11]

Received 28 November 2006 from the VA Health Services Research & Development Center for Practice Management and Outcomes Research, VA Ann Arbor Healthcare System, Ann Arbor, Michigan (AF, BJZ-F, PAU, DMS); the Center for Behavioral and Decision Sciences in Medicine, Ann Arbor, Michigan (AF, BJZ-F, PAU, AJ, HAD, DMS); the Division of General Internal Medicine, University of Michigan, Ann Arbor, Michigan (AF, BJZ-F, PAU, DMS); and the Department of Psychology, University of Michigan, Ann Arbor, Michigan (PAU). Financial support for this study was provided by grants from the National Institutes for Health (R01 CA87595-01A1 and 1 P50 CA101451-01). Drs Fagerlin and Smith are supported by MREP early career awards from the US Department of Veterans Affairs, Dr Zikmund-Fisher was supported by an HSR&D postdoctoral fellowship from the US Department of Veterans Affairs, and Dr Ubel is the recipient of a Presidential Early Career Award for Scientists and Engineers. The funding agreements ensured the authors' independence in designing the study, interpreting the data, and publishing the report. Portions of this research were presented at the annual meeting of the Society for Medical Decision Making, Chicago, Illinois, 22 October 2003.

Address correspondence to Angela Fagerlin, 300 North Ingalls Building, Room 7C27, Ann Arbor, MI 48109-0429; e-mail: fagerlin@med.umich.edu.

Table 1 Lipkus and Others' Objective Numeracy Scale

Item
Imagine that we have a fair, 6-sided die (for example, from a board game or a casino craps table). Imagine we now roll it 1000 times. Out of 1000 rolls, how many times do you think the die would come up even (numbers 2, 4, or 6)?*[2]
In the Big Bucks Lottery, the chances of winning a $10.00 prize is 1%. What is your best guess about how many people would win a $10.00 prize if 1000 people each buy a single ticket to Big Bucks?*[2]
In the Acme Publishing Sweepstakes, the chance of winning a car is 1 in 1000. What percentage of tickets to Acme Publishing Sweepstakes win a car?*[2]
Which of the following numbers represents the biggest risk of getting a disease?[2]
______ 1 in 100
______ 1 in 1000
______ 1 in 10
Which of the following numbers represents the biggest risk of getting a disease?[2]
______ 1%
______ 10%
______ 5%
If person A's risk of getting a disease is 1% in 10 years, and person B's risk is double that of A's, what is B's risk?[1r,2]
If person A's chance of getting a disease is 1 in 100 in 10 years, and person B's risk is double that of A's, what is B's risk?[1r,2]
If the chance of getting a disease is 10%, how many people out of 100 would be expected to get the disease?[2]
If the chance of getting a disease is 10%, how many people out of 1000 would be expected to get the disease?[1,2]
If the chance of getting a disease is 20 out of 100, this would be the same as having a ______ % chance of getting the disease.[1,2]
The chance of getting a viral infection is 0.0005. Out of 10,000 people, about how many of them are expected to get infected?[1,2]

Note: Items marked with an asterisk (*) indicate those originally developed by Schwartz and others.[6] 1/1r are items used in study 1. 1r items were revised (number 1 was changed to 3). Items marked with a "2" indicate items used in study 2 and study 3. No items were revised.

Taken together, these objective numeracy questions assess people's ability to perform mathematical tasks on risk magnitudes using percentages and proportions as well as their ability to convert percentages to proportions, proportions to percentages, and probabilities to proportions.

Most authors have found significant deficits in numerical ability even in highly educated samples. For example, in one study in which 88% of the sample reported having a college education or higher, only 51% could report accurately the likelihood of a fair, 6-sided die coming up even.[11] Similarly, 20% were unable to report which represented the biggest risk: 1 in 10, 1 in 100, or 1 in 1000.

In addition to the studies using mathematics test-like questions, a second line of research has considered the numerical skills needed to be an effective patient. The Test of Functional Health Literacy in Adults (TOFHLA) assesses patients' ability to read and understand health-related materials and contains both literacy and numeracy measures.[12,13] Low scores on the TOFHLA are associated with lower disease knowledge in patients with chronic diseases,[14,15] lower use of preventive care,[16] and higher rates of hospital admission and emergency department use.[2,4,5] However, the numeracy component of the TOFHLA is limited in that the questions reflect such simple qualitative procedures (e.g., the items test people's ability to understand directions for taking medicine, monitoring blood glucose, keeping clinic appointments) that it is difficult to generalize to their ability to perform more complex numerical calculations, such as those needed to understand risk communications or quantitatively contrast different treatment outcomes.

Although objective numeracy measures have been successful in demonstrating differences in risk perception and utility measurement, there are several significant problems with their use. First, as might be expected, participants in research studies are not especially receptive to taking an aptitude test. In fact, in our own studies, research assistants have received negative feedback from participants, many of whom have been displeased about answering these types of questions. Second, the rather aversive nature of objective numeracy measures could lead to lower completion rates as well as higher attrition rates for longitudinal studies. Third, conducting objective numeracy tests over the Internet or the telephone could be problematic. When completing the objective numeracy tests over the Internet, it is possible that respondents will use calculators or ask others to help them solve the problems. Difficulty with answering objective numeracy questions over the phone could be caused by the cognitive load required for respondents to retain all the components of the question in their memory while also trying to perform the calculations.

These concerns led us to develop the Subjective Numeracy Scale (SNS). Our primary objective in developing the SNS was to identify questions that would be quicker and more palatable for survey participants to answer yet still would be powerful predictors of people's ability to perform numerically intensive tasks. This article describes the development of the measure as well as a test of participants' experiences during survey completion.

Overview of Research

This article presents 3 studies, 2 of which describe the developmental process used to create the SNS and 1 of which tested participants' differential reactions to the objective and subjective numeracy scales. Study 1 describes the process of developing 42 items and narrowing the scale down to a preliminary set of 6 items. Study 2 builds on the results of study 1 by testing the 6 items that had emerged and 7 additional questions that we hypothesized would improve the correlation between objective and subjective numeracy. Study 2 results in the final, 8-item SNS. Study 3 compares participants' reactions to completing the objective numeracy questionnaire and our 8-item SNS, as well as their willingness to participate in future research. All studies received Institutional Review Board approval.

STUDY 1

Overview

We began the development of a subjective numeracy measure by creating and testing 42 items. Participants ($N = 364$) completed both the 42-item SNS and an objective numeracy scale, providing us with the ability to examine the correlations between the 2 scales. By examining correlations between each subjective item and the objective measure of mathematical ability, we were able to significantly shorten the questionnaire and also to determine what other types of questions needed to be included in the next round of testing.

Methods

Respondents

Participants were recruited from waiting areas in a university hospital and a Veterans Affairs hospital. After research assistants briefly described the study, those who agreed to participate received a questionnaire. Respondents at 1 site were given pens as compensation for their participation (an Institutional Review Board requirement).

Of the 364 participants who completed a questionnaire, 46% were female, 91% identified themselves as Caucasians, and the average age was 50 years ($sd = 16.4$, range = 19–86). There was substantial diversity in the educational attainment, with 21% having a high school education or less, 44% having completed at least some trade school or college (but receiving less than a bachelor's degree), and 33% having obtained at least a bachelor's degree.

Questionnaires

In addition to demographics questions (e.g., age, gender, race, education), participants completed 2 sets of numeracy questions.

Subjective numeracy questions. We had little prior research on the structure of numeracy to guide us in the development of items; we also had no a priori reasons to hypothesize that some areas of perception would correlate better with performance than others. For both of these reasons, we chose a broad, exploratory approach to item development, one that did not limit our focus only to those domains related to current objective numeracy scales.

Items were generated using a multistep procedure. First, we conducted an item-generation focus group with 15 people, during which participants were asked to generate factors that they believed would be related to numeric skills. Participants first free associated about what kinds of items would be needed for a SNS. Next, participants were given 4 constructs that we hypothesized might be related to people's ability to interpret numerical information: 1) experience with numerical tasks (experience), 2) perceived cognitive ability characterized by how well and how quickly respondents believe they can perform numerical tasks (ability), 3) comfort and interest in performing numerical tasks (interest), and 4) preference for the presentation of numerical information (preference). Participants were asked to develop items specifically related to these constructs. Close to 100 items were generated during this process. From all these generated items, we (the authors) chose the 42 items that, by consensus, were most clearly worded and relevant **(see http://mdm.sagepub.com/cgi/content/full/Volume/Issue/Page#/DC1 for table)**. Incidentally, all items could be classified under 1 of the 4 constructs listed above.

Objective numeracy questions. Given the large number of subjective numeracy items each participant was asked to complete, we chose to reduce respondent burden by using only a 5-question objective numeracy measure, adapted from Lipkus and others,[11] as our comparison standard. These 5 questions asked respondents to convert frequencies into percentages and percentages into frequencies, to compute the risk if it is doubled, and to convert odds into frequencies (see items labeled 1 and 1r in Table 1).

Data Analysis

We first conducted an exploratory factor analysis (maximum likelihood estimation, with Promax rotation) to examine whether our 42 items loaded onto 4 separate factors corresponding to the categories we used to develop the items (described above). We then created linear composites of the items representing each of the 4 categories and computed the Pearson bivariate correlations between each composite and the total objective numeracy score. After determining which constructs were most strongly related, we chose several individual items (based on the largest individual correlations with objective numeracy) to create a 6-item SNS. We tested the overall correlation of our new shortened scale with the total objective numeracy score and then concluded by using Cronbach's α to test for internal reliability.

In this and all subsequent studies, we calculated an objective numeracy score by counting the number of questions answered correctly. Following the methodology of Lipkus and others, answers left blank were counted as incorrect (as long as the respondent answered any questions at all). For the SNS, we used an imputed SNS score, equal to the average rating across all SNS items answered, for all participants who had missing data yet completed more than half of the SNS question items.*

*The method of coding missing answers on an objective numeracy scale as incorrect is debatable because we cannot know whether the item was left blank because a respondent could not complete it or rather chose not to for some other reason. Therefore, we considered a 2nd approach for handling missing responses and compared the results of the 2 methods. Specifically, we used a multiple imputation procedure using a Markov Chain Monte Carlo (MCMC) method.

Using data from study 1, we compared our key analysis—the correlation between the subjective and objective numeracy scores—using these 2 different techniques and found only small differences. To elaborate, we set the MCMC imputation procedure to replace all missing values on each of the 5 objective numeracy scale items and repeated the process 5 times (using SAS version 8 PROC MI and MIANALYZE[17]). This produced 5 new imputed data sets; we then recomputed the objective numeracy scale composite in each set. The correlations using these 5 imputations were all significant and varied from $r = 0.41$ to 0.44 and were close to the correlation obtained using the "code missing as wrong" imputation procedure ($r = 0.47$). In study 2, using the full Lipkus and others scale, these results were even more consistent. The imputation procedure created objective numeracy scale variables that correlated with subjective numeracy between $r = 0.49$ and $r = 0.51$, compared to the "missing as wrong" value of $r = 0.53$.

Results

A factor analysis confirmed that, with the exception of a minority of questions, the items we created fell into 3 of the 4 hypothesized factors: experience, interest, and ability (the 2 preference items did not hang together, and some of the experience questions did not load on any factor). A factor analysis was not conducted on the objective numeracy scale, as the scale's authors have already conducted and reported the results of a factor analysis of this scale.[11] The results of that factor analysis indicated that the scale has only 1 dimension.

We next looked at correlations between each of the 3 obtained factors and the total objective numeracy score (i.e., how many of the 5 items were answered correctly). We separately analyzed the individual preference items that did not load on 1 of these factors. An examination of these correlations revealed that the strongest correlations were with the ability factor as well as the 2 preference items. Indeed, 6 of the 7 highest individual item correlations were items that were classified as either ability or preference questions (see www.cbdsm.org/files/downloads/nums1apdx.pdf). Based on these evaluations, these 6 items were retained as a preliminary SNS (pSNS). Thus, no experience or interest items were retained. Cronbach's test of reliability for the pSNS showed a reliability of $\alpha = 0.84$. The objective numeracy questions showed an internal reliability of $\alpha = 0.66$ (Kuder-Richardson 20). The correlation between the sum of the final 6 items of the pSNS and the 5-item objective numeracy composite score was 0.47 ($P < 0.01$). We applied a formula to correct for the attenuation in this correlation due to the internal reliability of each scale.[18] The corrected correlation after taking into account the unreliability of the 2 scales was $r = 0.63$. The correlation between the 5-item objective scale and the ability subscale was $r = 0.39$, $P < 0.01$ (the corrected correlation $r = 0.52$), whereas the correlation between the objective numeracy scale and the preference subscale was $r = 0.42$, $P < 0.01$ (the corrected correlation was $r = 0.56$).

The missing data per question ranged from 2.69% to 17.0% for the objective numeracy measure and from 0.3% to 0.8% for the SNS.

For subjective numeracy, the "missing as wrong" approach is not applicable because there is no wrong answer. We could have employed the complex multiple imputation approach. However, missing rates observed were very low, typically less than 1% per item. Thus, it is highly unlikely that a more sophisticated imputation procedure would alter our results.

STUDY 2

Overview

Based on the results of study 1, we had identified 6 items that were strongly correlated with objective numeracy and had high internal reliability. However, in our analysis of study 1, we noted that the 2 preference items that were highly correlated with objective numeracy were the only 2 preference items that we had included in the initial 42-item question set. We were concerned that had we included other preference related items, these 2 items may not have emerged as the best predictors. This concern led us to conduct study 2, which tested the 6-item pSNS from study 1 as well as an additional 7 questions that tapped into the preference construct.

Methods

Respondents

Respondents were recruited from a Veterans Affairs (VA) hospital waiting room. Recruitment was restricted to this site because of the greater educational and ethnic diversity of patients compared to the university hospital (as found in study 1 but not reported). Furthermore, the VA sample had lower numeracy skills than the university participants (4.06 v. 4.59). The recruitment method was the same as for study 1, and participants were given pens after completing the survey.

Of the 287 respondents who completed a questionnaire, 44% were women, 72% identified themselves as Caucasians, and the average age was 58 years ($sd = 13.5$, range = 19–85). Again, our sample was educationally diverse, with 38% having a high school education or less, 52% having completed at least some trade school or college, and 10% having at least a bachelor's degree.

Questionnaires

Subjective numeracy questions. We included the 6 items with the highest correlations with objective numeracy as determined in study 1 and 7 new items developed to reflect the preference for numerical information construct. All items are listed in Table 2.

Objective numeracy questions. We used the full and unrevised objective numeracy questionnaire from Lipkus and others (see items labeled "2" in Table 1).

Table 2 Subjective Numeracy Items Tested in Study 2

	$\bar{x}$	s	Correlation with Objective Numeracy
Cognitive abilities (1 = *not at all good*, 6 = *extremely good*)			
How good are you at working with fractions?	3.67	1.51	0.41
How good are you at working with percentages?	3.92	1.47	0.43
How good are you at calculating a 15% tip?	4.20	1.54	0.41
How good are you at figuring out how much a shirt will cost if it is 25% off?	4.58	1.40	0.34
Preference for display of numeric information			
When reading the newspaper, how helpful do you find tables and graphs that are parts of a story?* (1 = *not at all*, 6 = *extremely*)	3.83	1.43	0.26
When people tell you the chance of something happening, do you prefer that they use *words* ("it rarely happens") or *numbers* ("there's a 1% chance")?* (1 = *always prefer words*, 6 = *always prefer numbers*)	3.53	1.82	0.33
When you hear a weather forecast, do you prefer predictions using *percentages* (e.g., "there will be a 20% chance of rain today") or predictions using only *words* (e.g., "there is a small chance of rain today")? (1 = *always prefer percentages*, 6 = *always prefer words*; reverse coded)	3.06	1.90	0.27
How *often* do you find numerical information to be useful? (1 = *never*, 6 = *very often*)	4.16	1.50	0.30
When reading about the likelihood of something happening, how helpful is it to see the exact percentage (e.g., 45% chance)?[a]	4.05	1.46	0.26
How much do you like statistics?[a]	3.34	1.42	0.25
How often do you use percentages in conversations (e.g., "I am 75% done with packing.")?[a]	2.92	1.61	0.16
When you ask someone what time it is, do you prefer that they tell you the exact time ("it's 10:04") or the approximate time ("it's a little after 10 o'clock")?[a]	2.98	1.90	0.14
How often do you express an opinion using numbers (e.g., "on a scale from 1–10, I give it a 7")?[a]	3.45	1.55	0.09

Note: The first 8 items in this table are those that compose the final Subjective Numeracy Scale (SNS). The 2 original preferences items from study 1 are marked with an asterisk (*).
a. Items that were tested in study 2 but were not retained in the final SNS.

Standard demographic questions were also included. Also included was a short risk communication or utility elicitation task. These data (reported as Study 2 in our companion article)[19] followed all numeracy questions and are not used in the present analyses.

Data Analysis

We conducted individual Pearson bivariate correlations between all subjective numeracy items and the objective numeracy measure and between the objective measure and the final SNS. Cronbach's α tested the reliability of the final scale.

Results

Correlations between each of the 7 new questions and respondents' objective numeracy scores ranged from 0.09 to 0.30 (n for each of the questions ranged from 283 to 287). The 2 original preference items (from study 1, marked with an asterisk in Table 2) were found to be among the mostly highly correlated items and were retained. In addition, 2 items that were the most highly correlated with objective numeracy (preference for the presentation of a weather forecast and the usefulness of numerical information) were combined with the previous 6 items to make an 8-item final SNS. The final scale was reliable, $\alpha = 0.82$. We again observed a significant correlation with the objective numeracy measure ($\alpha = 0.75$; $r = 0.53$, $P < 0.01$). After correcting for attenuation, the final correlation was $r = 0.68$. The correlation between the Lipkus and others scale and the ability subscale was $r = 0.47$, $P < 0.01$ (the corrected correlation was $r = 0.60$), whereas the correlation between the Lipkus and others scale and the preference subscale was $r = 0.44$, $P < 0.01$ (the corrected correlation was $r = 0.56$).

The missing data per question ranged from 1.9% to 20.7% for the objective numeracy measure and from 0.0% to 1.6% for the SNS.

Descriptive Characteristics of the Final SNS

Participants' scores on the final SNS ranged from 1.00 to 6.00, with a mean of 4.03 and a standard deviation of 1.04. The SNS scale was generally normally distributed, with only a slight negative skew (−0.36).

STUDY 3

Overview

To confirm our anecdotal perceptions that research participants find Lipkus and colleagues' objective numeracy measure to be aversive and to test whether the SNS is more palatable, we conducted a small survey that focused exclusively on the study participants' reaction to both objective and subjective numeracy measures.

Methods

Respondents

Participants were recruited from the cafeteria of a university hospital. The sample was composed of 52 individuals, of which 55% were women, 88% identified themselves as Caucasians, and the average age was 43 years ($sd = 14.8$, range = 18–78). Participants had a range of educational experience, with 8% having a high school education or less, 32% having completed at least some trade school or college, and 60% having at least a bachelor's degree.

Questionnaires

Participants were randomized to receive either the full 11-item Lipkus and others' objective numeracy measure (labeled "2" in Table 1) or the 8-item SNS. All participants rated 4 questions about their emotional reaction to the questionnaire using a 6-point Likert-type scale: 1) I enjoyed answering the questions in this study, 2) The questions in this study were annoying, 3) The questions in this study made me feel stressed, and 4) The questions in this study were frustrating. A final question asked participants their willingness to participate in another study: "We have another survey, similar in length, which you could also participate in. Would you be willing to do this additional study?" (answers were yes and no). Standard demographic questions (as described in study 1) were also included.

Procedures

Participants were instructed to work straight through the survey and return it as soon as they were done. Research assistants used a stopwatch to record the time it took each participant to complete the survey. Research assistants carefully and unobtrusively observed the participants, and stopwatches were paused if the participant stopped working on the questionnaire (e.g., participant engaged in a conversation).

Data Analysis

t Tests were used to compare performance (i.e., time needed to complete the measure) and attitudes between those who completed the objective numeracy measure and those who completed the SNS. A logistic regression analysis compared the participants' willingness to participate in a similar survey (yes/no) by what survey they completed (objective v. subjective).

Results

The results from this study are reported in Table 3. On average, participants receiving the SNS were able to complete the survey about 2.5 min more quickly than those who had to complete the objective numeracy test, a highly significant difference ($t = 3.52$, $P < 0.01$). This translates into an average time per item for subjective numeracy of only 24 s, versus an average of 31 s per item for the objective numeracy scale, $t = 2.41$, $P < 0.05$.

Ratings of how much the survey annoyed, stressed, and frustrated participants showed significant differences, with higher negative ratings coming from those individuals who had received the Lipkus and others objective numeracy measure (t's ranged from 2.54 to 3.34; all P's < 0.05). Participants' ratings of how

Table 3 Time to Complete and Emotional Response to Numeracy Surveys

Item	Subjective Numeracy Scale ($n = 27$), $\bar{x}$ (s)	Objective Numeracy Scale ($n = 25$), $\bar{x}$ (s)	P Value
Average time to complete (min)	5.03 (1.79)	7.49 (3.17)	< .01
Enjoyment rating	4.27 (1.12)	3.96 (1.34)	> .10
Annoyance rating	2.42 (1.36)	3.42 (1.47)	< .01
Stress rating	1.62 (0.90)	2.69 (1.38)	< .01
Frustration rating	1.92 (1.32)	2.88 (1.33)	< .05
Willing to do another similar survey?	50%	8%	< .05

Note: Analyses are *t* tests for time to complete and all ratings, but a logistic regression was used to evaluate willingness to do an additional survey.

much they enjoyed completing the survey show higher enjoyment by individuals completing the SNS, although this difference was not statistically significant ($P > 0.10$). Finally, in the last question of the survey, we asked participants to indicate whether they would be willing to do another similar survey. Fully half of those receiving the SNS were willing to do so, far more than the 8% receiving the Lipkus and others measure (odds ratio = 11.00, 95% confidence interval = 2.14–56.65).

GENERAL DISCUSSION

The SNS was developed over 3 studies, in which a pool of 49 items (42 items in study 1 and 7 items in study 2) was pared down to a final 8-item scale. The SNS asks participants to assess their ability in performing various mathematical tasks and their preference for the presentation of statistical information. We have shown that the SNS is closely associated ($r = 0.68$) with the well-known Lipkus and others objective measure of quantitative ability, can be administered in significantly less time, is a more pleasant experience for participants (which in turn may result in better completion and attrition rates), and results in less missing data. This suggests that replacing the Lipkus and others objective numeracy questionnaire with the SNS might decrease attrition rates for longitudinal studies and increase the likelihood that people will complete questionnaires. The SNS also holds an advantage over objective numeracy measures when assessed over the telephone or the Internet. Whereas the cognitive load requirements of a phone interview are likely to be high for objective numeracy measures, cognitive load is likely to be low for the SNS, making the scale much easier to administer over the telephone. Furthermore, in computer surveys, researchers do not have to be concerned that some participants' scores will be invalid because they have used calculators or other people to aid in their completion of the task.

There are several limitations of the developmental process. First, we used convenience samples to test the SNS. Our use of self-administered written surveys could have resulted in an exclusion of low-literacy/low-numeracy individuals because of the reading requirement. However, our samples were diverse in terms of educational achievement, and there was a wide range of objective numeracy skills observed.

Another potential limitation is that we may not have included all constructs that could be related to numeracy. In study 1, we generated items that reflected 4 concepts: 1) experience with numerical tasks, 2) perceived cognitive ability characterized by how well and how quickly respondents believe they can perform numerical tasks, 3) comfort and interest in performing numerical tasks, and 4) preference for the presentation of numerical information. It is possible that other constructs might have improved the correlation between the SNS and the Lipkus and others scale and/or the predictive ability of our scale. A related issue was our reliance on a single criterion for item selection: correlation with the Lipkus and others scale (and a modified version). We chose this scale because it is currently the most extensively used in research on numeracy in health settings. Ideally, we could have examined a variety of previously validated measures, covering various aspects of numerical ability in the context of medical decisions; doing so might have resulted in the selection of a different set of final items or perhaps a larger scale with several distinct subscales. However, at this point, there simply is not enough available research on numeracy to adequately address these issues. As noted above, we have made our full set of 42 items available at **http://mdm.sagepub.com**

/cgi/content/full/Volume/Issue/Page#/DC1 for any researchers who may be interested in exploring some of the items not included in our final measure.

A final potential limitation is the fact that our objective numeracy measure changed over the studies. Whereas studies 2 and 3 used the complete (and unrevised) Lipkus and others scale, study 1 used an adaptation of the scale. However, in a separate study that also included numeracy measures (not reported here), respondents received all the items used in all our studies, allowing us to compute the Lipkus and others scales used in study 1 and study 2. Respondents' performance on the 2 versions of the scale was highly correlated ($r = 0.85$).

Future research is clearly needed to determine how best to use the SNS to identify whether patients, or a subset of patients, have the numeracy ability necessary to understand the needed numerical information. Furthermore, although this article has mostly discussed the merits of the SNS in research terms, it may be that a quick and simple numeracy measure (e.g., the measure of Schwartz and others[6]) would ultimately prove useful in a clinical setting. To determine this, however, research is clearly needed to test whether, and how, the SNS could be used to identify which patients can or cannot understand the risks and benefits of treatment. We hope that future research will evaluate the usefulness of the SNS in a clinical setting.

Future research might also explore why the Lipkus and others scale is viewed negatively and whether the scale could be reduced to save time. We cannot say why the objective numeracy questionnaire was rated as more frustrating, stressful, or annoying. It may be the experience of having to take a mathematics test. It could be due to the repetitive nature of the questions. Another possibility is that the questions deal with disease and illness, and that may upset participants.

In our companion article, we report evidence for the SNS's predictive validity.[19] Although there may be research questions and methodologies in which an objective numeracy measure may be more appropriate, we suggest that our measure has many advantages while remaining effective at distinguishing highly numerate individuals from those whose quantitative skills are weaker.

REFERENCES

1. Kirsch IS, Jungeblut A, Jenkins L, Kolstad A. Adult literacy in America: a first look at the results of the National Adult Literacy Survey (NCES 93275). Washington, DC: US Department of Education; 1993.

2. Baker DW, Parker RM, Williams MV, Clark WS. Health literacy and the risk of hospital admission. J Gen Intern Med. 1998; 13(12):791–8.

3. Baker DW, Parker RM, Williams MV, Clark WS, Nurss J. The relationship of patient reading ability to self-reported health and use of health services. Am J Public Health. 1997;87(6):1027–30.

4. Baker DW, Gazmararian JA, Williams MV, et al. Health literacy and use of outpatient physician services by Medicare managed care enrollees. J Gen Intern Med. 2004;19(3):215–20.

5. Baker DW, Gazmararian JA, Williams MV, et al. Functional health literacy and the risk of hospital admission among Medicare managed care enrollees. Am J Public Health. 2002;92(8):1278–83.

6. Schwartz LM, Woloshin S, Black WC, Welch HG. The role of numeracy in understanding the benefit of screening mammography. Ann Intern Med. 1997;127(11):966–72.

7. Woloshin S, Schwartz LM, Moncur M, Gabriel S, Tosteson AN. Assessing values for health: numeracy matters. Med Decis Making. 2001;21(5):382–90.

8. Woloshin S, Schwartz LM, Byram S, Fischhoff B, Welch HG. A new scale for assessing perceptions of chance: a validation study. Med Decis Making. 2000;20(3):298–307.

9. Black WC, Nease RF, Tosteson A. Perceptions of risk and screening effectiveness in women younger than 50 years of age. J Natl Cancer Inst. 1995;87(10):720–731.

10. Schapira MM, Davids SL, McAuliffe TL, Nattinger AB. Agreement between scales in the measurement of breast cancer risk perceptions. Risk Anal. 2004;24(3):665–73.

11. Lipkus IM, Samsa G, Rimer BK. General performance on a numeracy scale among highly educated samples. Med Decis Making. 2001;21(1):37–44.

12. Baker DW, Williams MV, Parker RM, Gazmararian JA, Nurss J. Development of a brief test to measure functional health literacy. Patient Educ Couns. 1999;38(1):33–42.

13. Parker RM, Baker DW, Williams MV, Nurss J. 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.

14. Williams MV, Baker DW, Parker RM, Nurss JR. Relationship of functional health literacy to patients' knowledge of their chronic disease: a study of patients with hypertension and diabetes. Arch Intern Med. 1998;158(2):166–72.

15. Gazmararian JA, Williams MV, Peel J, Baker DW. Health literacy and knowledge of chronic disease. Patient Educ Couns. 2003;51(3):267–75.

16. Scott TL, Gazmararian JA, Williams MV, Baker DW. Health literacy and preventive health care use among Medicare enrollees in a managed care organization. Med Care. 2002;40(5):395–404.

17. SAS Statistical Software, Inc. [computer program]. Version 8.0. Cary, NC: SAS Institute; 2000.

18. Pedhazur EJ. Multiple Regression in Behavioral Research. 3rd ed. Orlando, FL: Harcourt Brace; 1997.

19. Zikmund-Fisher BJ, Smith DM, Ubel PA, Fagerlin A. Validation of the Subjective Numeracy Scale (SNS): effects of low numeracy on comprehension of risk communications and utility elicitations. Med Decis Making. **Forthcoming**.