Measuring Graph Literacy without a Test: A Brief Subjective Assessment

Rocio Garcia-Retamero, PhD, Edward T. Cokely, PhD, Saima Ghazal, PhD, Alexander Joeris, MD

Background**. Visual aids tend to help diverse and vulnerable individuals understand risk communications, as long as these individuals have a basic understanding of graphs (i.e., graph literacy). Tests of objective graph literacy (OGL) can effectively identify individuals with limited skills, highlighting vulnerabilities and facilitating custom-tailored risk communication. However, the administration of these tests can be time-consuming and may evoke negative emotional reactions (e.g., anxiety). **Objectives**. To evaluate a brief and easy-to-use assessment of subjective graph literacy (SGL) (i.e., self-reported ability to process and use graphically presented information) and to estimate the robustness and validity of the SGL scale and compare it with the leading OGL scale in diverse samples from different cultures. **Participants**. Demographically diverse residents (n = 470) of the United States, young adults (n = 172) and patients (n = 175) from Spain, and surgeons (n = 175) from 48 countries. **Design**. A focus group and 4 studies for instrument development and initial validation (study 1), reliability and convergent and discriminant validity evaluation (study 2), and predictive validity estimation (studies 3 and 4). **Measures**. Psychometric properties of the scale. **Results**. In about 1 minute, the SGL scale provides a reliable, robust, and valid assessment of skills and risk communication preferences and evokes fewer negative emotional reactions than the OGL scale. **Conclusions**. The SGL scale can be suitable for use in clinical research and may be useful as a communication aid in clinical practice. Theoretical mechanisms involved in SGL, emerging applications, limitations, and open questions are discussed. **Key words:** graph literacy; medical decision making; numeracy; risk communication; risk literacy; visual aids. **(Med Decis Making 2016;36:854–867)

Received 15 July 2014 from Department of Experimental Psychology, University of Granada, Spain (RGR); National Institute for Risk and Resilience & Department of Psychology, University of Oklahoma, Norman, OK, USA (ETC); Max Planck Institute for Human Development, Berlin, Germany (RGR, ETC); Department of Psychology, University of the Punjab, Lahore, Pakistan (SG); and AO Clinical Investigation and Documentation, Zurich, Switzerland (AJ). Financial support for this research was provided by the Ministerio de Economía y Competitividad (Spain) (PSI2011-22954 and PSI2014-51842-R) and the National Science Foundation (United States) (SES-1253263). The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report. Revision accepted for publication 18 April 2016.

Address correspondence to Rocio Garcia-Retamero, PhD, Facultad de Psicología, Universidad de Granada, Campus Universitario de Cartuja s/n, 18071 Granada, Spain; telephone: +34 958 246240 and Center for Adaptive Behavior and Cognition, Max Planck Institute for Human Development, Lentzeallee 94, D-14195, Berlin; e-mail: rretamer@ugr.es.

DOI: 10.1177/0272989X16655334

Visual aids can confer benefits when communicating health-relevant risk information to diverse and vulnerable individuals.[1–4] Visual aids are especially effective when they are transparent—that is, when their elements are well defined and they accurately and clearly represent the relevant information, making part-to-whole relationships in the data visually available.[5] Nevertheless, even the simplest visual aids are not beneficial for everyone.[6–8] Garcia-Retamero and Galesic[9] found that visual aids tend to be most beneficial when people have at least a moderate level of objective graph literacy (OGL)—that is, when they have the skills that allow them to extract data and meaning from graphical representations of quantitative information.[10] Garcia-Retamero and Galesic's estimate[9] was based on results from large probabilistic national samples in the United States and Germany, wherein visual aids increased accuracy from less than 20% to nearly 80% among people who were moderately graph literate, even when they had very low levels of objective numeracy (ON)—when they have the ability to use numbers to inform decision making.[11–15]

However, people with low graph literacy generally failed to benefit from visual aids.[16] Other related studies[17–19] have documented similar findings in diverse samples of participants, suggesting a need for more personalization of interventions in medical risk communication (e.g., custom tailoring).[6]

Most research on the relations between visual aids and graph literacy has assessed OGL with the scale developed by Galesic and Garcia-Retamero.[10] This scale consists of 13 items and covers several frequently used graph types. The scale includes items dealing with the communication of risks, treatment efficiency, and prevalence of diseases, and it measures 3 abilities of graph comprehension: 1) the ability to read the data, that is, to find specific information in the graph (e.g., the ability to read off the height of a bar within a bar chart); 2) the ability to read between the data, that is, to find relationships in the data as shown on the graph (e.g., the ability to read off the difference between 2 bars); and 3) the ability to read beyond the data, or make inferences and predictions from the data (e.g., the ability to project a future trend from a line chart).

The OGL scale was initially validated in several samples from different cultures.[8–10,17,18,20–26] Subsequent research indicates that the scale is generally a robust predictor with good psychometric properties. To illustrate, individuals with high OGL often extract more complex knowledge when viewing graphs than individuals with low OGL (e.g., it is more likely that they process important information in titles, labels, and scales).[22,24] Compared with those who have low OGL, individuals with high OGL also show lower reliance on irrelevant spatial features when interpreting graphs.[23] More graph-literate individuals also benefit more from interventions designed to improve risk understanding or to reduce errors and biases in graph comprehension (e.g., benefit from error bars).[21,27]

Despite its success, like all scales, the OGL scale has some limitations. One of the more pressing issues is that it typically requires 9 to 17 minutes to complete,[10,16,25] which may be too long for some uses (e.g., in clinical practice). Recent research shows that brief measures can be designed to work as well as long assessments. A prominent example is the Berlin Numeracy Test[11,28]—a 2-minute test of ON[29,30] that is also one of the strongest predictors of risk literacy and general decision-making skills in educated participants from industrialized countries (see RiskLiteracy.org). Unfortunately, some people find objective tests aversive and anxiety provoking no matter how brief.[31] This concern is common among people from disadvantaged backgrounds who may struggle to grasp numerical concepts that are essential for understanding health-relevant information.[6]

Some recent research suggests that measuring subjective perceptions of aptitudes can be less stressful and intimidating,[32] offering fair predictive power and good reliability. One such instrument that has proven valuable for clinical and basic research is the subjective numeracy (SN) scale developed by Fagerlin and colleagues.[2,33–36] The scale demonstrates good reliability,[33,37] is significantly correlated with scales measuring ON,[35] and predicts objective performance in problems involving numerical risks.[34,36,38] Inspired by these successes, here we present the development and refinement of a subjective graph literacy (SGL) scale that predicts OGL and risk communication behavior, while reducing assessment of time and anxiety.

OVERVIEW OF THE RESEARCH

This report includes data from a focus group and 4 studies conducted with diverse samples from different cultures. The focus group and study 1 were designed to develop assessment items and test basic attributes of the new scale. Participants were a group of experts and a large sample of residents of the United States, respectively. Study 2 evaluated test reliability and convergent and discriminant validity in a sample of young adults in Spain. Studies 3 and 4 investigated predictive validity in large samples of patients and physicians from 48 countries.

Development and Validation of the Subjective Graph Literacy Scale

Following Fagerlin and others,[33] we used an exploratory approach to item development. We conducted an item generation focus group with 6 experts who generated items representing 3 constructs: 1) perceptions of basic and advanced skills of graph comprehension (i.e., how well and how quickly people believe they can perform tasks that measure the abilities of reading the data, between the data, and beyond the data), 2) experience with tasks involving graphs (e.g., perceptions of easiness and usefulness of tasks involving graphs in daily life), and 3) comfort and interest in performing tasks involving graphs. Experts generated 45 items.

Table 1 Items of the 10-Item and the 5-Item SGL Scales and Correlation between Item Scores with Total Scores in the OGL and SGL Scales

Items	OGL Scale	10-Item SGL Scale	5-Item SGL Scale
1. How good are you at working with bar charts? (1 = *not at all good;* 6 = *extremely good*)	.30*	.78*	.87*
2. How good are you at working with line plots? (1 = *not at all good;* 6 = *extremely good*)	.31*	.76*	.86*
3. How good are you at working with pies? (1 = *not at all good;* 6 = *extremely good*)	.32*	.76*	.85*
4. How good are you at inferring the size of a bar in a bar chart? (1 = *not at all good;* 6 = *extremely good*)	.28*	.79*	.87*
5. How good are you at determining the difference between 2 bars in a bar chart? (1 = *not at all good;* 6 = *extremely good*)	.33*	.77*	.83*
6. How good are you at projecting a future trend from a line chart? (1 = *not at all good;* 6 = *extremely good*)	.16*	.75*	.69*
7. Are graphs easier to understand than numbers? (1 = *not at all;* 6 = *much easier*)	.07	.62*	.37*
8. How often do you find graphical information to be useful? (1 = *never;* 6 = *very often*)	.25*	.74*	.52*
9. To what extent do you believe in the saying "a picture is worth one thousand words"? (1 = *not at all;* 6 = *extremely*)	−.06	.48*	.21*
10. When reading books or newspapers, how helpful do you find graphs that are part of a story? (1 = *not at all;* 6 = *extremely*)	.17*	.64*	.38*

Note: Items 1 to 10 were included in the 10-item subjective graph literacy (SGL) scale. Items 1 to 5 were selected to be included in the 5-item SGL scale.
*$P < 0.05$.

Content analysis was performed by 2 experts to select 10 items (see Table 1). Item selection was guided by the following 4 criteria: 1) efficiency: each item had to measure a different aspect of SGL; 2) duration: the scale had to be short; 3) the items had to cover different types of graphs; and 4) the items had to investigate perceptions of both basic and advanced skills of graph comprehension. Items were designed to be answered using 6-point scales. To evaluate the 10-item SGL scale and select the items for the refined final version of the scale, we conducted study 1.

STUDY 1

Participants

The 10-item SGL scale was administered on computers via Amazon's Mechanical Turk, which provides access to demographically diverse samples of participants.[38] Participants were 470 residents of the United States. Of those, 70 participants did not answer all the questions or were classified as distracted respondents (i.e., the amount of time that they spent reading and answering the questions was more than 3 standard deviations from the average). These participants were excluded from data analyses. The final sample included 400 participants. Participants' age, sex, and education were not recorded in study 1 because of a programming error. Demographics from large samples of studies that we have previously conducted via Mechanical Turk ($n > 3000$) include a reasonable balance of sex (42% male) and age (mean, 32 years; range, 18−78 years). Many participants (42.4%) have some college experience. All participants were compensated with $1 and consented to participation through an online consent form at the beginning of the study.

Materials and Procedure

To test convergent validity, participants completed the 10-item SGL scale and the OGL scale developed by Galesic and Garcia-Retamero[10]. To assess predictive (criterion) validity, participants completed a 10-item battery requiring interpretation of graphical information about health risks.[39] The battery included 4 graphs (2 line plots and 2 bar charts) representing realistic health risks taken from published studies in medical journals. Two of the

Table 2 Basic Attributes and Psychometric Properties of the Scales in Study 1

		OGL Scale	SGL Scale	
			10-Item	5-Item
Basic attributes	Length	8 min 46 s	1 min 28 s	52 s
	Mean	10.28	43.58	21.93
	Standard deviation	2.14	8.25	4.87
	Skewness	−1.20	−.45	−.46
Reliability	Cronbach's α	.67	.89	.91
Convergent validity	Correlation with scores in OGL		.29*	.35*
Predictive validity	Mean % (SE) of correct answers			
	First quartile	.48 (.02)	.51 (.02)	.48 (.02)
	Second quartile	.59 (.02)	.57 (.02)	.55 (.02)
	Third quartile	.63 (.02)	.61 (.02)	.61 (.02)
	Fourth quartile	.65 (.02)	.62 (.01)	.62 (.01)
	Standardized β coefficients (SEβ)			
	As single predictor	.36* (.05)	.27* (.05)	.29* (.05)
	With SGL/OGL	.29* (.05)	.18* (.05)	.19* (.05)

Note: OGL, objective graph literacy; SGL, subjective graph literacy.
*$P < 0.05$.

graphs reported changes in blood pressure after an intervention in hypertensive patients.[40] A third graph reported the relationship between body mass index and diabetes.[41,42] The fourth graph reported the results of a diabetes prevention program.[43] The Ethics Committee of the Michigan Technological University approved the study. All participants provided informed consent. There were no time constraints.

Results and Discussion

We selected items 1 to 5 as the candidate items to be included in the refined final version of the SGL scale. Item selection was guided by the following principle: threshold correlation of item scores with total scores in the OGL scale and the 10-item SGL scale had to be larger than .25 and .75, respectively. We evaluated the 5-item and the 10-item version of the SGL scale on several criteria, including basic attributes, reliability, and validity (see Table 2). The 2 versions of the SGL scale were compared with the OGL scale[10] (see Table 1).

Basic attributes

Participants took less time to complete the 5-item scale than the 10-item scale ($t_{399} = -33.41$, $P = 0.001$) and the OGL scale ($t_{399} = -50.08$, $P = 0.001$), with an average time of 9 and 40 seconds per question for SGL and OGL, respectively. Scores in the SGL scales were approximately normally distributed; scores in the OGL scale showed moderate negative skew.

Reliability

Cronbach's alphas of the scales indicated internal consistency. A principal axis factor analysis of the 5 items of the refined SGL scale showed that all items loaded highly on a single factor explaining 73% of the observed variance, suggesting a 1-factor construct with high confirmatory reliability.

Validity

Scores in the SGL scales were moderately correlated with those in the OGL scale, revealing convergent validity. In addition, the SGL scales predicted accuracy of graph understanding in tasks involving health risks, revealing considerable predictive validity. The 5-item and the 10-item versions of the SGL scale accounted for 47% and 37% of the total variance explained by the full model, respectively; the OGL scale explained 66% of the total variance. Both SGL and OGL predicted accuracy of graph understanding when they were included as single and competing predictors in regression models (see Table 2) (incremental $F_{1,397} = 14.837$, $P = 0.0001$ when SGL was added to the model).

In summary, the SGL scales were found to be psychometrically efficient instruments that assess skill-related beliefs and actual graph comprehension. The final refined 5-item scale shows numerically superior reliability and validity coefficients and takes less time to complete than the 10-item scale (see Table 2).

STUDY 2

Participants

Participants were a sample of 172 undergraduates from the University of Granada in Spain (26% males; mean age, 21 years; range, 19–30 years).

Procedure

Participants completed a paper-and-pencil survey including questions about demographic information. To assess convergent validity, participants completed the 5-item version of the SGL scale and the OGL scale.[10] To assess test-retest reliability, participants completed these scales 2 different times, 2 weeks apart. To test discriminant validity, participants completed the Berlin Numeracy Test,[11] the SN scale,[34] a self-report measure of one's general sense of self efficacy,[44] a measure of need of cognition,[45] and a self-report measure of personality,[46] including self-ratings of subjective experiences to estimate emotional stability, conscientiousness, agreeableness, extraversion, and openness to experience.

The scales were translated into Spanish and back-translated by 2 skilled translators.[6] The Ethics Committee of the University of Granada approved the method, and all participants were compensated with credits and consented to participation. There were no time constraints.

Results and Discussion

Basic attributes

The distribution of the SGL scores was approximately normal; scores in the OGL scale showed moderate negative skew (see Table 3).

Reliability

Cronbach's alphas of the scales indicated internal consistency. The test-retest analysis showed considerable self-report stability.

Validity

The SGL scale was highly correlated with the OGL scale, showing convergent validity. SGL was also moderately correlated with SN, ON, and self-efficacy, suggesting that these constructs share some common ability elements. In contrast, SGL was not correlated with need of cognition or personality traits, providing evidence of discriminant validity. Both SGL and SN predicted OGL when they were included as single predictors in regression models ($\beta = .58$, $t_{170} = 9.29$, $P = 0.0001$ and $\beta = .28$, $t_{170} = 3.83$, $P = 0.0002$ for SGL and SN, respectively). When SGL and SN were included as competing predictors, the effect of SN on OGL became unreliable and nonsignificant ($\beta = .58$, $t_{169} = 8.10$, $P = 0.0001$ and $\beta = -.01$, $t_{169} = -.10$, $P = 0.94$ for SGL and SN, respectively), indicating that all of SN's explanatory power was already measured by and shared with SGL.

Additional Validation Studies

As a means of out-of-sample validation, we sought to assess the extent to which predictive validity generalized to large samples of people who varied widely in their skills (e.g., physicians and patients) who also had diverse cultural backgrounds. Study 3 estimated the ability of the SGL scale to predict preferences for receiving and communicating health risks using graphs and reported clinical behavior. Study 3 also investigated emotional reactions toward the SGL scale. Study 4 examined predictive accuracy for estimations of risk reduction and psychological mechanisms involved in SGL.

STUDY 3

Participants

Participants were 175 active surgeons from 48 countries and 175 patients from 2 hospitals in the cities of Jaen and Granada (Spain). Seventeen surgeons and 24 patients did not answer all the questions or did not complete the study in a timely manner. These participants were excluded from analyses (cf. study 1). The final sample included 158 surgeons and 151 patients.

Surgeons were recruited by announcing the study on the webpage of the AO Foundation (see aofoundation.org). To participate, surgeons had to be fluent in English as all materials were presented

Table 3 Basic Attributes and Psychometric Properties of the Scales in Study 2

		OGL Scale	SGL Scale
Basic attributes	Mean	10.19	19.09
	Standard deviation	1.50	4.89
	Skewness	−.77	.19
Reliability	Cronbach's α	.61	.87
	Test-retest	.68	.86
Convergent validity	Correlation with scores in OGL		.58*
Discriminant validity	Correlation with scores in		
	ON	.47*	.31*
	SN	.28*	.49*
	Self-efficacy	.16*	.24*
	Need of cognition	.12	.05
	All personality traits	.02	−.07
	Stability	−.08	−.08
	Conscientiousness	.07	.02
	Agreeableness	−.03	−.09
	Extraversion	−.01	−.07
	Openness to experience	.09	.05

Note: OGL, objective graph literacy; ON, objective numeracy; SGL, subjective graph literacy; SN, subjective numeracy.
$^{*}P < 0.05$.

in this language. On average, surgeons were 42 years of age (range, 29–69 years), with 16 years of experience (range, 2–30 years); 90% were male; 34% were European, 32% were American, 25% were Asian, and the rest (9%) were African or Oceanian. Surgeons did not receive any compensation for participation.

Patients were recruited by the first author before treatment consultation. To be eligible for recruitment, patients had to have no previous formal medical training and be fluent in Spanish as all materials were presented in this language. Patients were not included if they had significant sensory impairment. On average, patients were 60 years of age (range, 30–80 years); 23% were male, 93% were Spaniards, and 49% had a chronic condition. Patients were educationally diverse, with 51% and 35% having no more than a high school degree or less than a high school degree, respectively. Fourteen percent had some college education. Patients received 5€ for participating in the study.

Materials and Procedure

Participants completed a 3-part computer-based questionnaire. In the first part, participants provided demographic information and completed the 5-item version of the SGL scale, the OGL scale,[10] and the Berlin Numeracy Test.[11] In addition, half of the participants answered the following 3 questions about their emotional reaction to the SGL scale, using 7-point sales: 1) To what extent did you enjoy answering the questions included in the scale, 2) to what extent did the questions in the scale make you feel stressed, and 3) to what extent did the questions in the scale make you feel frustrated (see Fagerlin and others[33] for a similar procedure)? The other half of the participants answered the same questions about their emotional reactions to the OGL scale. Answers to question 1 were reversed and combined with questions 2 and 3 for a composite score (Cronbach's α = .81).

In the second part of the questionnaire, surgeons (patients) had to imagine that they were reading a paper in a medical journal (a brochure about health) that included a graph reporting results. On 6-point scales, they indicated 1) whether they would read the text or the graph first and 2) whether they preferred to use numbers or graphs when they communicated health-relevant risk information to their patients (other people). On 6-point scales, they also indicated 3) whether they typically used numbers or graphs when they communicated health-relevant risk information to their patients (other people). The third part of the questionnaire is described in Study 4.

The materials were developed in English and were translated into Spanish and back-translated by 2 skilled translators. The Ethics Committee of the

University of Granada approved the method, and all participants consented to participation. There were no time constraints.

Results and Discussion

Basic attributes and reliability

Compared to the OGL scale, both surgeons and patients reported less negative emotional reactions toward the SGL scale (t_{156} = 6.86, P = 0.001, and t_{149} = 6.64, P = 0.001) and took less time to complete it (t_{157} = −47.04, P = 0.001, and t_{150} = −84.08, P = 0.001; see Table 4). Scores on the SGL scale were normally distributed in patients and showed only trivial negative skew in surgeons. Scores in the OGL scale showed slight negative skew in patients and moderate negative skew in surgeons. Cronbach's alphas indicated satisfactory internal consistency.

Validity

Both the SGL and the OGL scales predicted preferences for receiving and communicating health-relevant risk information using graphs, revealing reliable predictive validity (Table 4). SGL was a stronger predictor of preferences for receiving information than OGL, accounting for 98% and 32% of the total variance explained by a full regression model, respectively (with SGL, OGL, and ON as competing simultaneous predictors). SGL was also a stronger predictor than OGL in preferences for communicating information, accounting for 89% v. 41% of the total variance explained by the full model, respectively. Moreover, SGL explained a substantial portion of additional variance in these preferences after controlling for OGL (see Table 4). In contrast, OGL did not explain any unique variance after controlling for SGL (incremental $F_{1,306}$ = .784, P = 0.377, and $F_{1,306}$ = 1873, P = 0.172 for preferences for receiving and communicating information, respectively, when OGL was added to the model). Both SGL and OGL also predicted which format surgeons and patients used when communicating health-relevant risk information, accounting for 63% and 78% of the total variance explained by the full model, respectively (incremental $F_{1,306}$ = 10.809, P = 0.001 when SGL was added to the model).

STUDY 4

This study investigated the degree to which the SGL scale predicts accuracy of estimations of risk reduction. Participants received information about the risk of suffering a side effect after 2 alternative medical interventions. The risk was reported in samples of patients of different sizes (e.g., 190 patients were allocated to one intervention and 230 patients were allocated to the other intervention). Computing or otherwise representing proportions of affected patients within each intervention group allows accurate risk reduction estimations.[47] In contrast, considering absolute numbers of patients who suffer the side effect (rather than proportions) within each intervention group would produce inaccurate estimations.[18,48]

Study 4 also investigated the underlying psychological mechanisms involved in SGL. Higher levels of general abilities are often associated with differences in *metacognition* (e.g., thinking about thinking, self-assessment of risk comprehension, overconfidence), which can gives rise to more skilled and informed decision making.[49] Cognitive abilities such as numeracy tend to predict overconfidence, independent of accuracy (i.e., more numerate individuals show more appropriate confidence levels even when their other judgments are not correct).[50] Appropriate confidence levels also predict the accuracy of risk understanding[51] and can mediate the relationship between numeracy and superior decision making.[28,50] Accordingly, in study 4, we investigated whether differences in SGL predict differences in appropriate confidence (i.e., confidence calibration). We also investigated whether confidence calibration explains, at least partially, the effect of SGL on accuracy of graphically presented risk information.

Participants

This study employed the same sample of participants used in study 3. Demographics and characteristics of the sample are reported above.

Materials and Procedure

Participants were presented with a medical scenario. The scenario described the results of a randomized controlled trial testing side effects of a new type of anesthesia in patients who underwent surgery. Participants were provided with information about the risk of suffering postoperative deep vein thrombosis in 2 randomly selected groups of patients who underwent total hip replacement. One group of 190 patients was allocated to general

Table 4 Basic Attributes and Psychometric Properties of the Scales in Studies 3 and 4

		OGL Scale		SGL Scale	
		Surgeons	Patients	Surgeons	Patients
Basic attributes	Length	9 min 30 s	9 min 50 s	48 s	59 s
	Mean	10.63	7.87	19.87	18.34
	Standard deviation	2.40	2.67	5.94	4.57
	Skewness	−.80	−.27	−.43	−.03
	Emotional reaction	3.51	3.70	2.25	2.43
Reliability	Cronbach's α	.76	.94	.70	.86
Convergent validity	Correlation with scores in OGL			.37*	.55*
Predictive validity	Preferences for receiving information				
	Mean estimation (SEM)				
	First quartile	3.28 (.30)	3.38 (.32)	3.15 (.24)	3.07 (.29)
	Second quartile	4.16 (.35)	3.32 (.34)	4.05 (.25)	3.68 (.29)
	Third quartile	4.51 (.24)	4.10 (.30)	3.83 (.35)	4.25 (.33)
	Fourth quartile	3.91 (.21)	4.44 (.24)	5.00 (.17)	4.74 (.24)
	Standardized β coefficients (SEβ)				
	As single predictor	.17* (.08)	.26* (.08)	.38* (.07)	.34* (.08)
	With SGL/OGL	.04 (.08)	.10 (.09)	.37* (.08)	.29* (.09)
	With ON	.15 (.08)	.25* (.08)	.37* (.08)	.34* (.08)
	With all factors	.06 (.08)	.10 (.09)	.33* (.08)	.28* (.09)
	Preferences for communicating information				
	Mean estimation (SEM)				
	First quartile	2.63 (.23)	2.31 (.28)	2.61 (.19)	2.30 (.20)
	Second quartile	2.66 (.29)	2.68 (.27)	2.69 (.22)	2.66 (.25)
	Third quartile	2.71 (.16)	3.07 (.26)	3.07 (.32)	3.41 (.29)
	Fourth quartile	3.58 (.22)	3.33 (.22)	3.83 (.21)	3.47 (.26)
	Standardized β coefficients (SEβ)				
	As single predictor	.20* (.08)	.24* (.08)	.31* (.08)	.30* (.08)
	With SGL/OGL	.09 (.08)	.11 (.09)	.27* (.08)	.24* (.09)
	With ON	.14 (.08)	.24* (.08)	.27* (.08)	.29* (.08)
	With all factors	.06 (.08)	.11 (.09)	.25* (.08)	.23* (.09)
	Reported behavior				
	Mean estimation (SEM)				
	First quartile	2.56 (.22)	1.83 (.16)	2.72 (.17)	1.93 (.12)
	Second quartile	2.72 (.25)	1.97 (.12)	2.69 (.19)	2.03 (.17)
	Third quartile	2.68 (.16)	2.65 (.24)	3.07 (.30)	3.13 (.25)
	Fourth quartile	3.60 (.19)	2.88 (.18)	3.50 (.22)	2.74 (.21)
	Standardized β coefficients (SEβ)				
	As single predictor	.25* (.08)	.33* (.08)	.27* (.08)	.32* (.08)
	With SGL/OGL	.17* (.08)	.22* (.09)	.21* (.08)	.20* (.09)
	With ON	.22* (.08)	.33* (.08)	.25* (.08)	.32* (.08)
	With all factors	.16 (.08)	.22* (.09)	.20* (.08)	.20* (.09)
	Accuracy of estimations (numerical information)				
	Mean absolute distance (SE) to correct value				
	First quartile	.15 (.01)	.22 (.04)	.13 (.03)	.21 (.02)
	Second quartile	.23 (.04)	.20 (.02)	.14 (.04)	.20 (.06)
	Third quartile	.15 (.04)	.17 (.04)	.11 (.03)	.23 (.05)
	Fourth quartile	.08 (.02)	.27 (.06)	.13 (.05)	.25 (.07)
	Standardized β coefficients (SEβ)				
	As single predictor	−.14 (.12)	.04 (.11)	.01 (.12)	.07 (.11)
	With SGL/OGL	−.16 (.12)	−.01 (.14)	.05 (.12)	.08 (.14)
	With ON	−.10 (.11)	.07 (.11)	.14 (.11)	.10 (.11)
	With all factors	−.14 (.11)	.01 (.14)	.17 (.12)	.10 (.14)

(continued)

Table 4 (continued)

	OGL Scale		SGL Scale	
	Surgeons	Patients	Surgeons	Patients
Accuracy of estimations (visual information)				
Mean absolute distance (SE) to correct value				
First quartile	.16 (.03)	.30 (.07)	.16 (.03)	.23 (.05)
Second quartile	.09 (.03)	.14 (.03)	.11 (.03)	.17 (.03)
Third quartile	.03 (.02)	.09 (.02)	.04 (.02)	.03 (.01)
Fourth quartile	.01 (.01)	.03 (.02)	.01 (.01)	.02 (.01)
Standardized β coefficients (SEβ)				
As single predictor	−.54* (.09)	−.52* (.11)	−.50* (.10)	−.48* (.11)
With SGL/OGL	−.39* (.10)	−.38* (.11)	−.31* (.10)	−.31* (.11)
With ON	−.43* (.10)	−.52* (.10)	−.42* (.09)	−.45* (.11)
With all factors	−.29* (.10)	−.40* (.11)	−.29* (.10)	−.27* (.11)

Note: OGL, objective graph literacy; ON, objective numeracy; SGL, subjective graph literacy; SN, subjective numeracy.
*$P < 0.05$.

anesthesia, while another group of 230 patients received a new type of anesthesia. The new type of anesthesia had a relative risk reduction of 36%. The task involved realistic risk as information was taken from a published study.[52]

Participants were randomly assigned to 1 of 2 information format conditions. In the *numerical* condition, they were told, ''52 of 190 patients allocated to general anesthesia suffered postoperative deep vein thrombosis. Compared to the group allocated to general anesthesia, 40 of the 230 patients in the group allocated to the new type of anesthesia suffered postoperative deep vein thrombosis.'' In the *visual* condition, participants received the same information represented via an icon array (see Figure 1).

The dependent variables included estimations of risk reduction (both in percentages and frequencies) and confidence. First, participants were asked to infer by what percentage the risk of suffering deep vein thrombosis was reduced when patients received the new type of anesthesia (question 1 [Q1]). Second, participants reported how confident they were about their answers to Q1, on a 7-point scale (question 2 [Q2]). Finally, participants answered the following 2 questions: How many of 1000 patients might have suffered postoperative deep vein thrombosis if they were allocated to general anesthesia (question 3 [Q3]), and how many of 1000 patients might have suffered postoperative deep vein thrombosis if they were allocated to the new type of anesthesia (question 4 [Q4])?

Data Analysis

By subtracting Q4 from Q3 and dividing it by Q3, we computed estimations of risk reduction in frequencies.[12] We then computed the absolute difference between these estimations and the correct value (i.e., accuracy of estimations of risk reduction in frequencies). Second, we computed the absolute difference between answers to Q1 and the correct value (i.e., accuracy of estimations of risk reduction in percentages). Finally, we computed judgment calibration bias as the difference between confidence (Q2) and accuracy of estimations of risk reduction in percentages (Q1).[28,50]

We analyzed the SGL and OGL scales with respect to their ability to identify participants who would benefit from graphical representations of health risks. We conducted multiple regression analyses with accuracy of estimations of risk reduction in frequencies (Q3 and Q4) as a dependent variable when the information was presented either numerically or visually. We investigated whether judgment calibration bias was related to individual differences in SGL and OGL. We also investigated whether judgment calibration bias explained the effect of SGL and OGL on accuracy of graphically presented risk information. We conducted mediational analyses with SGL and OGL as predictors, accuracy of estimations of risk reduction in frequencies (Q3 and Q4) as the criterion, and judgment calibration bias as a potential mediator (Q1 and Q2).

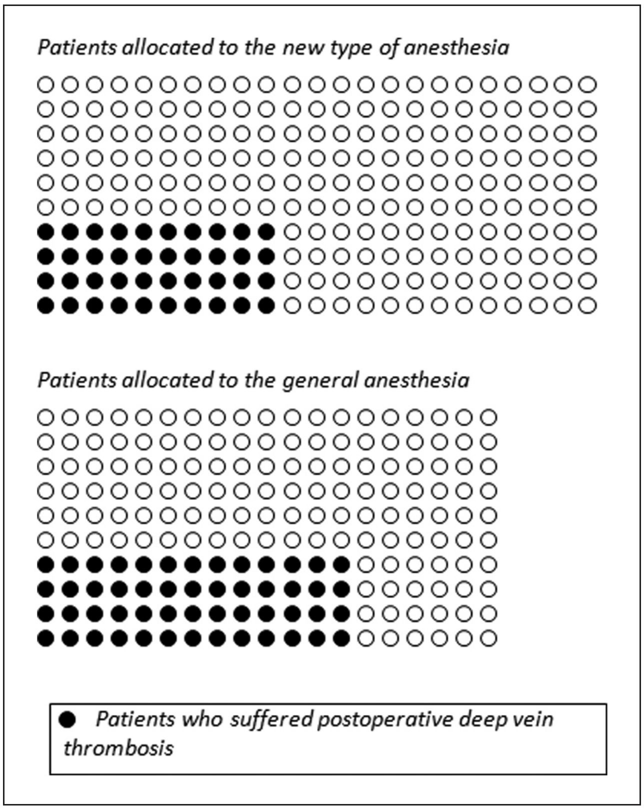

Figure 1 Icon array representing numerical information about risk reduction.

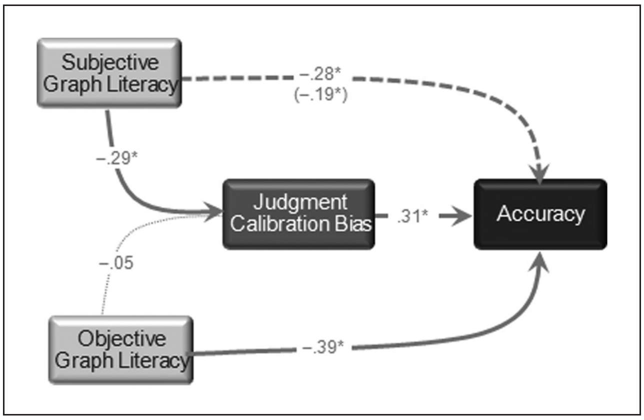

Figure 2 Path analyses of the effect of subjective graph literacy and objective graph literacy on accuracy of estimations of risk reduction (absolute distance to the correct value) and the mediational effect of judgment calibration bias when the risk information was provided visually. Standardized coefficients are shown. *$P < 0.05$.

Results and Discussion

Accuracy of estimations of risk reduction was similar in more and less graph-literate participants when the risk information was provided in numbers (see Table 4). In contrast, when the risk information was represented visually, participants with moderate or high SGL and participants with moderate or high OGL provided more accurate estimations than those with low SGL or OGL. We included SGL, OGL, and ON as competing predictors in a series of regression models. The unique predictive power of the SGL was found to be substantial, accounting for 63% of the total variance explained by a full regression model (incremental $F_{1,145} = 12.244$, $P = 0.001$ when SGL was added to the model; cf. 78% unique variance for the OGL). Mediational analyses were consistent with the hypothesis that SGL is partially independent yet also robustly related to OGL (see Figure 2), indicating that most people have some general understanding of how graph literate they are.

Both SGL and OGL were robustly related to judgment calibration bias when they were included as single predictors in regression models ($\beta = -.32$, $t_{146} = -4.08$, $P = 0.001$ and $\beta = -.20$, $t_{146} = -2.51$, $P = 0.013$, respectively), with more graph-literate surgeons and patients showing less bias. When SGL and OGL were included as competing predictors, the effect of OGL on judgment calibration bias became unreliable and nonsignificant ($\beta = -.05$, $t_{145} = -.52$, $P = 0.61$ and $\beta = -.29$, $t_{145} = -3.18$, $P = 0.002$ for OGL and SGL, respectively) indicating that all of OGL's explanatory power was already measured by and shared with SGL. When judgment calibration bias was included in the regression analysis, the effect of SGL on estimation accuracy was reduced ($\beta = -.19$, $t_{144} = -2.44$, $P = 0.02$). The result of the Sobel test[53] indicates that judgment calibration bias partially mediated the influence of SGL on estimation accuracy ($z = -2.61$, $P = 0.009$).

GENERAL DISCUSSION

We developed and evaluated the SGL scale, a brief and reliable instrument that predicts graph comprehension by measuring people's judgment about their own ability to interpret graphs. In 4 studies using diverse samples from different cultures, we found robust reliability and validity that were generally comparable to that of the well-established OGL scale developed by Galesic and Garcia-Retamero.[10] Comparative analyses revealed that the

SGL scale tends to achieve a high level of psychometric performance while limiting test anxiety and improving assessment of risk communication preferences. The SGL scale also takes less than 10% of the time of the objective test (i.e., less than 1 minute on average).

Psychometric Properties of the Subjective Graph Literacy Scale

The SGL scale is well suited for estimating perceptions of graph comprehension and can be used to predict objective skills and decision performance (i.e., whether people are able to extract data and meaning from graphical representations). Convergent validity was established showing high correlations with the OGL scale.[10] Discriminant validity was established by showing moderate correlations with SN, ON, and self-efficacy,[11,33,44] suggesting that these constructs share some common ability elements. Discriminant validity analysis also showed that SGL was not systematically tied to unrelated subjective judgment of other psychological factors (i.e., need of cognition and personality).[45,46] Predictive validity was estimated by showing that the SGL scale offered both robust and unique predictive performance (i.e., independent of OGL), predicting preferences for receiving and communicating health-relevant risk information using graphs, and predicting self-reported behavior (i.e., information use). Predictive validity was also estimated by showing that the SGL scale predicted accuracy of performance in tasks involving interpretation of graphs in health risk communication—an essential aspect of a valid graph literacy assessment.

The evaluation and validity assessment of the SGL was based on a variety of studies conducted in diverse groups of people, including the general public, young adults, doctors, and patients from 30 to 80 years of age. Our analyses included samples of people from 48 different countries. These groups of people have different cultural backgrounds, education levels, languages, governments, and health systems. In addition, our studies examined risk communication in ecologically valid tasks that accurately reproduced the problems that people commonly encounter when they make health decisions. These ecological studies covered many topics, including estimations of health risk and risk reduction, confidence, emotional reactions, preferences for receiving and communicating health-relevant risk information, and self-reported health-relevant behavior.

Emerging Applications in Clinical Research and Practice

People from different cultures often have problems understanding health-relevant risks.[6,54] Previous research indicates that well-designed visual aids are often effective decision support tools when communicating these risks, so long as people have moderate levels of graph literacy.[9,10,16–19] The current research extends this literature in several ways. Results suggest that the SGL scale may be useful as a quick and simple measure of graph literacy in clinical research and practice, reducing respondents' burden while providing significant predictive validity. Theoretically, the OGL scale will likely offer slightly improved predictive performance in some contexts and may thus be better suited for certain applications. However, in clinically oriented applications, the extra time required and the increases in participants' anxiety caused by objective tests should make the subjective test a better choice (see Fagerlin and others[33] and Zikmund-Fisher and others[34] for a similar conclusion). In short, the SGL scale provides an effective balance between precision and usability that is well suited to the needs of clinical researchers. The SGL scale might also be useful for telephone and Internet surveys[38,55] and for designing interactive web pages and interventions with custom-tailored information that can be adaptively delivered according to participants' skills (e.g., "please answer one question so that we can present easier-to-understand information about your recent diagnosis").

The current research also converged with a growing body of work showing that visual aids tend to offer benefits that are robust to culturally variations.[5] This finding is particularly noteworthy given the increasingly globalized scope of science, including evidence-based medicine and associated risk communication practices. The robustness of transparent visual aids can be further improved with skills assessments and small-scale validation studies (e.g., using the SGL to assess graph reading level the same way numeracy tests are used to assess risk literacy or "risk reading levels" needed to interpret and understand common risk communications).

Cognitive Mechanisms of Subjective Graph Literacy

The theoretical construct of SGL is well characterized as primarily reflecting differences in 1) objective graph skills and 2) metacognition (e.g., overconfidence, effective deliberation, self-knowledge, self-reflection), whereas OGL appears to be more tightly linked to objective graph skills, as documented in study 4. In this study, more skilled individuals (with moderate/high OGL or SGL) showed less bias and more accurate estimations of their performance. However, only SGL was a reliable predictor of bias when both OGL and SGL were used as competing predictors, suggesting that SGL shares all the variance related to OGL skills as well as assessing unique variance associated with other metacognitive factors. Results also suggest that confidence calibration (i.e., reduced overconfidence/underconfidence) partially mediates the effect of SGL on accuracy of interpretations of graphically presented risk information. Confidence calibration tends to be related to differences in deliberation during decision making,[56] wherein more calibrated individuals often spend more time evaluating task-relevant information, thereby improving their memory for that information[57–59] and enabling more complex representation and reasoning (e.g., people who are overconfident often fail to check their assumptions or understanding).[11,28] Supplemental analyses of the results of study 4 further support these hypotheses, indicating that individual differences in SGL can be modeled as resulting from both skilled understanding of graphs (i.e., OGL; β = .42, $t_{305} = 7.78$, $P = 0.001$) and differences in metacognition (i.e., overconfidence bias; β = −.18, $t_{305} = -3.55$, $P = 0.001$). In contrast, SGL scores were largely independent of skilled understanding of numerical information (i.e., ON; β = .10, $t_{305} = 1.85$, $P = 0.07$). That is, there was no effect of ON above and beyond its influence on OGL.

Limitations, Open Questions, and Future Research

Reviewers have suggested that the main limitation of the SGL scale derives from the item selection process. That is, the scale is only based on the best 10 of the initial 45 candidate items generated by our expert panel. It is possible that other items could perform better in various ways. The selection of the final 5 questions was also based on item-total correlations, which led to an emphasis on subjective ability questions, which were highly correlated with each other, rather than assessment of subjective preference items. As is appropriate for all scales, future research will need to consider whether alternate sets of items measure various aspects of SGL more efficiently. Generally, results indicate that the current SGL scale provides a relatively robust, simple, and reliable tool for researchers and clinicians who want to quickly assess broad differences in graph literacy in culturally and demographically diverse samples.

Most standards for informed decision making hold that people should be able to deliberate about risks and consequences of their decisions in the light of their own values. That is, informed decision making generally requires representative risk comprehension. The current research showed for the first time that some aspects of the quality of one's deliberation in naturalistic risk evaluation can be robustly predicted by differences in SGL (i.e., metacognitive variables such as confidence calibration). Individual differences in SGL and OGL and their effect on risk comprehension may also shape the expression of decision-relevant factors such as anxiety, motivation, and general care or deliberation (e.g., reading and estimation latency). Future research can test these hypotheses using cognitive process tracing assessments (e.g., reaction times, functional magnetic resonance imaging, eye tracking, and think-aloud protocols).[58,60] These process-level assessments are essential components of test development and construct validation that help eliminate test bias and inform test and training development (i.e., mapping cognitive and affective processes that cause performance differences).[11,61,62] Although there is some research investigating individual differences in OGL,[22,24] beyond the current results, we know of no other cognitive process-oriented study on SGL or its relations to judgment and decision making.

Efficient design of decision support technologies requires considerable attention to the fit between people's skills, strategies, and environments (i.e., an interacting systems perspective).[57,63] The easy to use 1-minute subjective graph literacy assessment may be particularly useful when trying to quickly identify people who may benefit from various interventions with graphs (e.g., what types of decisions require more and less graph literacy). This kind of personalization is a cornerstone of efficient custom-tailored risk communication (for recent reviews, see Garcia-Retamero and Cokely[5,64] and Garcia-Retamero and Galesic[6]). Future research may also identify suitable strategies for communicating health risks to patients who are not graph literate. Ongoing work using analogies from people's everyday lives shows that these analogies are relatively undemanding in terms of graph literacy and numeracy, and they tend to be effective as well.[65]

Encouraging elaborative information processing by including reflective questions about graphs followed by accuracy feedback or using simple explanations to convey the meaning of important information in graphs has also been helpful for improving graph understanding in people with low graph literacy.[21]

Considering the role of graph literacy in other domains (e.g., education, economics, finance, climate science, and engineering),[10] we see many opportunities to leverage the SGL scale for research and educational benefit. Future research might investigate SGL in other health professionals (e.g., nurses), in business and legal professionals, and in larger samples of at-risk or underserved populations (e.g., minorities or rural community members). Based on the strength of the current findings, we predict that the SGL scale will prove to be an efficient assessment that often matches and sometimes outperforms longer, objective tests.

REFERENCES

1. Waters EA, Weinstein ND, Colditz GA, Emmons KM. Reducing aversion to side effects in preventive medical treatment decisions. J Exp Psychol Appl. 2007;13(1):11–21.

2. Zikmund-Fisher BJ, Fagerlin A, Ubel PA. Improving understanding of adjuvant therapy options by using simpler risk graphics. Cancer. 2008;113(12):3382–90.

3. Spiegelhalter D, Pearson M, Short I. Visualizing uncertainty about the future. Science. 2011;333(6048):1393–400.

4. Garcia-Retamero R, Cokely ET. Effective communication of risks to young adults: using message framing and visual aids to increase condom use and STD screening. J Exp Psychol Appl. 2011;17(3):270–87.

5. Garcia-Retamero R, Cokely ET. Communicating health risks with visual aids. Curr Dir Psychol Sci. 2013;22(5):392–9.

6. Garcia-Retamero R, Galesic M. Transparent Communication of Health Risks: Overcoming Cultural Differences. New York, NY: Springer; 2013.

7. Keller C, Siegrist M. Effect of risk communication formats on risk perception depending on numeracy. Med Decis Making. 2009;29(4):483–90.

8. Zikmund-Fisher BJ, Witteman HO, Dickson M, et al. Blocks, ovals, or people? Icon type affects risk perceptions and recall of pictographs. Med Decis Making. 2014;34(4):443–53.

9. Garcia-Retamero R, Galesic M. Who profits from visual aids: overcoming challenges in people's understanding of risks. Soc Sci Med. 2010;70(7):1019–25.

10. Galesic M, Garcia-Retamero R. Graph literacy: a cross-cultural comparison. Med Decis Making. 2011;31(3):444–57.

11. Cokely ET, Galesic M, Schulz E, Ghazal S, Garcia-Retamero R. Measuring risk literacy: the Berlin Numeracy Test. Judgm Decis Mak. 2012;7(1):25–47.

12. 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.

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

14. Anderson BL, Schulkin J. Numerical Reasoning in Judgments and Decision Making about Health. Cambridge, UK: Cambridge University Press; 2014.

15. Garcia-Retamero R, Andrade A, Sharit J, Ruiz JG. Is patient's numeracy related to physical and mental health? Med Decis Making. 2015;35(4):501–11.

16. Ruiz JG, Andrade AD, Garcia-Retamero R, Anam R, Rodriguez R, Sharit J. Communicating global cardiovascular risk: are icon arrays better than numerical estimates in improving understanding, recall and perception of risk? Patient Educ Couns. 2013;93(3):394–402.

17. Gaissmaier W, Wegwarth O, Skopec D, Müller A, Broschinski S, Politi MC. Numbers can be worth a thousand pictures: individual differences in understanding graphical and numerical representations of health-related information. Health Psychol. 2012;31(3):286–96.

18. Garcia-Retamero R, Cokely ET. The influence of skills, message frame, and visual aids on prevention of sexually transmitted diseases. J Behav Decis Making. 2014;27(2):179–89.

19. Okan Y, Garcia-Retamero R, Cokely ET, Maldonado A. Individual differences in graph literacy: overcoming denominator neglect in risk comprehension. J Behav Decis Making. 2012;25(4): 390–401.

20. Garcia-Retamero R, Muñoz R. How to improve comprehension of medical risks in older adults. Rev Latinoam Psicol. 2013; 45(2):253–64.

21. Okan Y, Garcia-Retamero R, Cokely ET, Maldonado A. Improving risk understanding across ability levels: encouraging active processing with dynamic icon arrays. J Exp Psychol Appl. 2015;21(2):178–94.

22. Okan Y, Galesic M, Garcia-Retamero R. How people with low and high graph literacy process health graphs: evidence from eye-tracking. J Behav Decis Making. 2015;29(2–3):271–94.

23. Okan Y, Garcia-Retamero R, Galesic M, Cokely ET. When higher bars are not larger quantities: on individual differences in the use of spatial information in graph comprehension. Spatial Cogn Comput. 2012;12(2–3):195–218.

24. Woller-Carter MM, Okan Y, Cokely ET, Garcia-Retamero R. Communicating and distorting risks with graphs: an eye-tracking study. Proc Hum Fact Ergon Soc Annu Meet. 2012;56(1):1723–7.

25. Rodríguez V, Andrade AD, García-Retamero R, et al. Health literacy, numeracy, and graphical literacy among veterans in primary care and their effect on shared decision making and trust in physicians. J Health Commun. 2013;18(Suppl 1):273–89.

26. Sharit J, Lisigurski M, Andrade AD, et al. The roles of health literacy, numeracy, and graph literacy on the usability of the VA's personal health record by veterans. J Usability Stud. 2014;9(4): 173–93.

27. Okan Y, Garcia-Retamero R, Cokely ET, Maldonado A. Biasing and debiasing health decisions with bar graphs: costs and benefits of graph literacy. Under review.

28. Ghazal S, Cokely ET, Garcia-Retamero R. Predicting biases in very highly educated samples: numeracy and metacognition. Judgm Decis Mak. 2014;9(1):15–34.

29. Reyna VF, Nelson WL, Han PK, Dieckmann NF. How numeracy influences risk comprehension and medical decision making. Psychol Bull. 2009;135(6):943–73.

30. Peters E. Beyond comprehension the role of numeracy in judgments and decisions. Curr Dir Psychol Sci. 2012;21(1):31–35.

31. Peters E, Bjalkebring P. Multiple numeric competencies: when a number is not just a number. J Pers Soc Psychol. 2015; 108(5):802–22.

32. Dolan JG, Cherkasky OA, Li Q, Chin N, Veazie PJ. Should health numeracy be assessed objectively or subjectively? Med Decis Making. 2015 May 6 [Epub ahead of print].

33. 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 Making. 2007;27(5):672–80.

34. 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 Making. 2007;27(5):663–71.

35. Galesic M, Garcia-Retamero R. Statistical numeracy for health: a cross-cultural comparison with probabilistic national samples. Arch Intern Med. 2010;170(5):462–8.

36. Hess R, Visschers VH, Siegrist M, Keller C. How do people perceive graphical risk communication? The role of subjective numeracy. J Risk Res. 2011;14(1):47–61.

37. Zikmund-Fisher BJ, Witteman HO, Fuhrel-Forbis A, Exe NL, Kahn VC, Dickson M. Animated graphics for comparing two risks: a cautionary tale. J Med Internet Res. 2012;14(4):e106.

38. Paolacci G, Chandler J, Ipeirotis PG. Running experiments on Amazon Mechanical Turk. Judg Decis Making. 2010;5(5):411–9.

39. Garcia-Retamero R, Cokely ET. Promoting shared medical decision making in doctors and their patients. Manuscript in preparation. 2016.

40. Schneider RH, Staggers F, Alxander CN, et al. A randomised controlled trial of stress reduction for hypertension in older African Americans. Hypertension. 1995;26(5):820–7.

41. Chan JM, Rimm EB, Colditz GA, Stampfer MJ, Willett WC. Obesity, fat distribution, and weight gain as risk factors for clinical diabetes in men. Diabetes Care. 1994;17(9):961–9.

42. Colditz GA, Willett WC, Rotnitzky A, Manson JE. Weight gain as a risk factor for clinical diabetes mellitus in women. Ann Intern Med. 1995;122(7):481–6.

43. Knowler WC, Barrett-Connor E, Fowler SE, et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. N Engl J Med. 2002;346(6):393–403.

44. Suárez PS, García AMP, Moreno JB. Escala de autoeficacia general: Datos psicométricos de la adaptación para población española. Psicothema. 2000;12:509–13.

45. Falces C, Briñol P, Sierra B, Becerra A, Alier E. Validación de la escala de necesidad de cognición y su aplicación al estudio del cambio de actitudes. Psicothema. 2001;13(4):622–8.

46. Gosling SD, Rentfrow PJ, Swann WB. A very brief measure of the Big-Five personality domains. J Res Pers. 2003;37(6):504–28.

47. Garcia-Retamero R, Galesic M. Communicating treatment risk reduction to people with low numeracy skills: a cross-cultural comparison. Am J Public Health. 2009;99(12):2196–202.

48. Garcia-Retamero R, Galesic M, Gigerenzer G. Do icon arrays help reduce denominator neglect? Med Decis Making. 2010;30(6):672–84.

49. Baron J. Rationality and Intelligence. Cambridge, UK: Cambridge University Press; 2005.

50. Garcia-Retamero R, Cokely ET, Hoffrage U. Visual aids improve diagnostic inferences and metacognitive judgment calibration. Front Psychol. 2015;6:932.

51. Bruine de Bruin W, Parker AM, Fischhoff B. Individual differences in adult decision-making competence. J Pers Soc Psychol. 2007;92(5):938–56.

52. Davis FM, Laurenson VG, Gillespie WJ, Wells JE, Foate J, Newman E. Deep vein thrombosis after total hip replacement: a comparison between spinal and general anaesthesia. J Bone Joint Surg Br. 1989;71(2):181–5.

53. Sobel ME. Asymptotic confidence intervals for indirect effects in structural equation models. Sociol Method. 1982; 13(1982):290–312.

54. Lipkus IM. Numeric, verbal, and visual formats of conveying health risks: suggested best practices and future recommendations. Med Decis Making. 2007;27(5):696–713.

55. Buhrmester M, Kwang T, Gosling SD. Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data? Pers Psychol Sci. 2011;6(1):3–5.

56. Jaeggi SM, Buschkuehl M, Jonides J, Shah P. Short- and long-term benefits of cognitive training. Proc Natl Acad Sci USA. 2011;108(25):10081–6.

57. Cokely ET, Kelley CM. Cognitive abilities and superior decision making under risk: a protocol analysis and process model evaluation. Judg Decis Making. 2009;4(1):20–33.

58. Cokely ET, Kelley CM, Gilchrist AL. Sources of individual differences in working memory: contributions of strategy to capacity. Psychon Bull Rev. 2006;13(6):991–7.

59. Garcia-Retamero R, Cokely ET, Wicki B, Joeris A. Improving surgeons' risk literacy with visual aids. Patient Educ Couns. In press.

60. Schulte-Mecklenbeck M, Kühberger A, Ranyard R. A Handbook of Process Tracing Methods for Decision Research: A Critical Review and User's Guide. New York, NY: Psychology Press; 2010.

61. Cronbach LJ, Meehl PE. Construct validity in psychological tests. Psychol Bull. 1955;52(4):281–302.

62. Messick S. Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. Am Psychol. 1995;50(9):741–9.

63. Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwartz LM, Woloshin S. Helping doctors and patients make sense of health statistics. Psychol Sci Public Interest. 2007;8(2):53–96.

64. Garcia-Retamero R, Cokely ET. Using visual aids to help people with low numeracy make better decisions. In: Anderson BL, Schulkin J, eds. Numerical Reasoning in Judgments and Decision Making about Health. Cambridge, UK: Cambridge University Press; 2014. p 153–74.

65. Galesic M, Garcia-Retamero R. Using analogies to communicate information about health risks. Appl Cognitive Psych. 2013; 27(1):33–42.