Developing a Better and More User-Friendly Numeracy Scale for Patients

W. Trey Hill, PhD; Gary L. Brase, PhD; and Kevin L. Kenney, MS

ABSTRACT

Background: A person's ability to work with and understand numerical information (i.e., numeracy) is increasingly important in everyday health and other decision-making contexts. Several survey measures of numeracy have been developed to address this trend, including the widely used General Numeracy Scale (GNS), which is thematically focused on health decision-making and is assumed to measure a unidimensional construct of numeracy. **Objective**: The present research was designed to evaluate this proposed unidimensional structure of general numeracy, for which prior data have given mixed empirical support. **Methods**: Three samples completed the GNS, in different forms, and responses were analyzed in terms of underlying factor structure. **Key Results**: We show that both one-factor and four-factor models of numeracy are plausible based on the GNS (Study 1), and then develop a multiple-choice version of the GNS (i.e., the MC-GNS) that demonstrates some increased clarity in factor structure due to the consistent response format (Study 2). A further study evaluated the convergent and discriminant validity of the MC-GNS (Study 3), finding it to be as good as or better than the prior scale. **Conclusions**: Additionally, the MC-GNS is easier for people to take, likely to be less stressful, and easier for practitioners to score. Collectively, this research identifies a problem with the GNS measure, develops improvements to help address this problem, and in the process creates a way to more easily measure numeracy in practical settings. **[*HLRP: Health Literacy Research and Practice*. 2019;3(3):e174-e180.]**

Plain Language Summary: Numeracy is important across health contexts. Prevalent numeracy scales assumedly measure a single construct but empirical support for this is lacking. We find both one- and four-factor models are consistent with one scale and develop a revision that clarifies this structure without sacrificing validity. This revised numeracy scale is easier to administer and score, and therefore preferable in practical settings.

Understanding and using numbers is increasingly important in daily life and, in particular, in patients' understanding of medical situations. For instance, understanding that a 1 in 10 (10%) chance is a greater risk than a 1 in 100 (1%) chance could have important implications for one's health and well-being. A person's numerical literacy (or numeracy) is the ability to both understand numbers and know how and when to use them in novel situations. Numeracy is positively correlated with both general mental ability and rationality, but these correlations are low enough (all $rs \leq .41$) that numeracy is considered to be a distinct construct (Brooks & Pui, 2010).

Numeracy intersects crucially with shared medical decision-making (Frosch & Kaplan, 1999; Moumjid, Gafni, Bremond, & Carrere, 2007) because numeracy predicts the degree of patient passivity in shared decision-making situations, with people of low numeracy desiring a less active role than people of high numeracy (Galesic & Garcia-Retamero, 2011) and demonstrating less sensitivity to the risks described during the informed consent process (Couper & Singer, 2009). Indeed, numeracy is predictive of better medical decision-making across the lifespan (Wood et al., 2011) and predicts correct medical decision-making even when controlling for education and for cognitive ability (Brown et al., 2011; Wood et al., 2011). In sum, numeracy shows clear promise, but to realize that promise our measures of numeracy must be accurate and must have a solid underlying construct structure.

Although of great practical and theoretical importance, numeracy remains difficult to measure in a consensually agreed upon way (Reyna, Nelson, Han, & Dieckmann, 2009).

Several questionnaires have been created to assess numeracy or numeracy-related constructs, often with high face validity but with much less attention paid to construct validity and even less attention still spent on their theoretical foundations. This is a serious situation because these numeracy scores are often used as an individual difference measure with the assumption that numeracy is a unidimensional, homogenous construct (e.g., Chapman & Liu, 2009; Galesic & Garcia-Retamero, 2011; Hanoch, Miron-Shatz, Cole, Himmelstein, & Federman, 2010; Hess, Visschers, & Siegrist, 2011; Hill & Brase, 2012; Peters, Hibbard, Slovic, & Dieckmann, 2007). A better understanding of numeracy as a construct is warranted, and measures of numeracy should be aligned with that construct. The purposes of the present research are to empirically evaluate the structure of one of the more popular measures, the General Numeracy Scale (GNS) (Lipkus, Samsa, & Rimer, 2001), to improve the correspondence between this numeracy measure and the conceptual construct of numerical literacy, and to establish the validity and utility of this new measure.

The GNS (Lipkus et al., 2001), is an 11-item measure of numeracy that has become a favorite among judgment and decision-making researchers, including much of the medical decision-making community. Schapira, Wallaker, and Sedivyvy (2009) evaluated several measures of numeracy and concluded that the GNS performed well at discriminating between people with high and low numeracy across a fairly large range of intellectual ability.

To evaluate the psychometric quality of the GNS, Lipkus et al. (2001) performed a factor analysis of the 11 questions. The first factor accounted for 49.4% of the variance. Second and third factors accounted for 11.9% and 8.1% of the variance, respectively. Lipkus et al. (2001) suggested that much of the variance explained in the second and third factors was due to similar structures of the relevant questions loading under those factors. Therefore, Lipkus et al. (2001) suggested that their test measures only one truly psychological construct—global numeracy—and that the three-factor model suggested by their results was most likely an artifact of the question formats rather than measuring three separate constructs. There does not appear to have been any follow-up work, however, to assess or address these suggested artifacts.

STUDY 1

Objectives

Study 1 used a large dataset to examine if the GNS measures a unidimensional underlying factor of numeracy or if a larger number of factors emerge.

Methods

Data were collected from 467 undergraduate students at a large Midwestern university as part of Institutional Review Board-approved mass testing sessions for a number of different researchers. A packet, including the GNS, was given to large classes of up to 200 introductory psychology students. Although exact demographic information is unavailable, this sample was approximately 60% women and predominately age 18 to 20 years. Data from six questionnaires were excluded because four or more responses were omitted, leaving 461 usable questionnaires.

Due to issues with performing factor analytic procedures on binary data (Woods, 2002), these data were analyzed using tetrachoric, rather than Pearson, correlation matrices. Tetrachoric correlations assume an underlying latent correlation based on continuous data. These matrices were smoothed using procedures created by Uebersax (2007). Tetrachoric correlation matrices can often be nonpositive definite, but "noise" within the matrix can be systematically removed by

W. Trey Hill, PhD, is an Associate Professor, Department of Psychology, Fort Hays State University. Gary L. Brase, PhD, is a Professor, Department of Psychological Sciences, Kansas State University. Kevin L. Kenney, MS, is a Graduate Student, Department of Psychological Sciences, Kansas State University.

Address correspondence to Gary L. Brase, PhD, Department of Psychological Sciences, Kansas State University, 492 Bluemont Hall, 1114 Mid-Campus Drive, Manhattan, KS 66506; email: gbrase@ksu.edu.

Acknowledgment: The authors thank Katrina Ellis, PhD (Florida Institute of Technology), Jisook Park, PhD (Fort Hays State University), James Daugherty, PhD (Center for Army Leadership), Lora Adair, PhD (Brunel University), and Haley Dillon, PhD (State University of New York, New Paltz) for help with data collection, planning, and feedback during all phases of the project.

Disclosure: The authors have no relevant financial relationships to disclose.

Received: December 31, 2017; Accepted: November 30, 2018

doi:10.3928/24748307-20190624-01

smoothing methods, rendering the tetrachoric correlation matrix more appropriate for factor analytic methods (Knol & Berger, 1991). The TetMat smoothing software detected one problematic eigenvalue for this tetrachoric matrix.

After smoothing, the matrix was analyzed using SPSS syntax. First, a principal components analysis was performed specifying only to yield factors that achieve eigenvalues greater than 1. Second, because Lipkus et al. (2001) suggested that their scale assesses a single broad construct of numeracy, another principal components analysis was performed that was specified to yield only a single factor. Due to the increased likelihood that many of the items in the GNS were highly correlated, a direct oblimin rotation was performed on the multifactor analysis (no rotation was necessary for the single factor specified analysis). As such, the pattern matrix rather than the loading matrix was interpreted for each item's unique contribution to a single factor (Tabachnick & Fidell, 2007). (Note: Previous studies have used both factor analyses and principle component analyses, but most theoretical discussions used "factors" regarding dimensions of numerical literacy. Although factors extracted from factor analysis and components from principal components analysis are theoretically and statistically different, for convenience and ease of reading, both are labeled as "factors" throughout this article.)

Key Results

When a minimum eigenvalue of 1 was specified, the results yielded four factors that cumulatively explained a total of 71.36% of the variance. Because this was an exploratory analysis, standard guidelines on the number of items required for composition of a single factor were relaxed. Not surprisingly, the explained variance dropped substantially when specifying a single factor structure (down to 34.95%); however, if one uses a liberal loading criterion of .320 (Tabachnick & Fidell, 2007), all of the GNS items in the single factor analysis made the cut (see **Table A** and **Table B** for factor analysis details.)

These results have some interesting implications for the theoretical structure of the GNS. Although the first analysis displayed four distinct factors, the second analysis, which specified only a single factor, showed that all of the GNS items could be loaded to some extent onto a single factor. The plausible four-factor structure remains troubling, though, especially as there are theoretical explanations supportive of a four-factor structure (e.g., Liberali, Reyna, Furlan, Stein, & Pardo, 2012). If the scale truly measures a unidimensional construct of numeracy, a large data set should produce clear evidence of that. Instead, we found equally plausible single-factor and four-factor structures in the GNS. Study 2 explores two possible explanations for why a multiple-factor structure for the GNS may exist.

STUDY 2

Objectives

Some particular characteristics of the GNS may inadvertently lead to a multifactor structure. The GNS is composed of 11 questions, but they are heterogeneous in type. Most questions are free response or fill-in-the-blank, but two questions are multiple choice. Additionally, two other questions are similar in content, even arguably redundant. Having multiple response types in a single scale can produce misleading factor analytic results simply due to problems loading together because of problem difficulty and format similarity rather than theoretical similarity (Tabachnick & Fidell, 2007). Specifically, some response types (e.g., multiple-choice) are easier than others (e.g., open response), potentially producing an "easy" factor and a "difficult" factor. Because the GNS has two response types and two redundant questions, the purpose of Study 2 was to develop a revised General Numeracy Scale (i.e., the multiple-choice GNS [MC-GNS]) with a uniform multiple-choice response format and with some items modified to increase validity.

Using data from Study 1, the most common incorrect responses were identified for all free responses, and these were used as distractor options for our MC-GNS. Additionally, some of the questions in the original GNS were slightly modified to increase their difficulty, both because an often-reported characteristic of the GNS is that the distribution of total scores is negatively skewed and because a multiple-choice format could increase scores due to lucky guessing. Questions 8A and 8B were combined into a single question, bringing the total number of questions to 10. One additional risk comprehension question was then added.

If the factor structure of the GNS (Lipkus et al., 2001) in Study 1 is merely the result of different response types, then converting the scale to a consistent multiple-choice format should result in fewer factors and perhaps a clearer picture of a unidimensional construct of numeracy. However, if we continue to get several reasonably loading factors based on a numeracy questionnaire with all multiple-choice responses, it would suggest that the GNS may not necessarily measure a single construct.

Methods

Data were collected from 748 undergraduate students enrolled in introductory psychology courses at a large Midwestern university, as part of the departmental prescreening process. The sample contained data from 494 female students and 253 male students (mean [*M*] age = 18.83 years, standard deviation [*SD*] for age = 1.49); one participant declined to provide his or her gender. Participants completed the MC-GNS and other prescreening items via an online survey platform. The same data preparation methods were used as described for Study 1. No problems were detected in the correlation matrix. Again, two principal components analyses were performed, with one specifying an eigenvalue of at least 1 and a second specifying a single factor.

Key Results

Cronbach's alpha for the MC-GNS was .69, slightly higher than the Cronbach's alpha of .62 for the GNS data collected in Study 1. This brings the reliability close to an acceptable range (Nunnally, 1978), at least in the context of assuming a unidimensional structure. The mean numeracy score was slightly higher for the MC-GNS ($M = 8.36$, $SD = 2.23$) than for the GNS ($M = 7.49$, $SD = 1.88$), and the MC-GNS had a negative skew similar to the GNS distribution in Study One (**Figure 1**).

Results of the first principal components analysis, limiting factor extraction to factors with an eigenvalue of at least 1, showed three factors based on the correlation matrix. These factors explained a total of 62.92% of the variance (**Table C**). When the analysis was limited to a single factor the explained variance dropped to 40.55%. Examining the single-factor structure for loadings equal to or higher than .320 showed that all 11 of the MC-GNS items met that criterion (**Table D**). When only one factor was specified, the MC-GNS actually accounted for more variance than the original GNS. In the unconstrained analysis the GNS factors explained slightly more variance than the MC-GNS; however, the MC-GNS also yielded fewer underlying factors. Perhaps more importantly, despite the MC-GNS being composed entirely of multiple-choice questions and with arguably redundant items eliminated, a principal components analysis yielded three factors. The multifactor solution, it seems, cannot be attributed entirely to item format issues. Closer examination of the results shows that one of the factors contains only two items, which can be hazardous and misleading (Tabachnick & Fidell, 2007). Limiting the number of factors to two produces a more stable result with more items per factor. However, in many cases, items load heavily on both factors, lending support to the concept of unidimensional numeracy as measured by the MC-GNS. It is, unfortunately, not possible to further reduce the factor structure statistically via a second data reduction analysis because these data were analyzed using a matrix rather than raw data.

Aside from the theoretical issue of the factors possibly underlying numeracy, there is a practical sense in which the MC-GNS has improved utility. It is much easier to administer and score. Ease of use is important because many researchers have proposed links between numeracy and real-world decision-making situations, including the process of shared decision-making in health decisions. If medical doctors are to be expected to administer a numeracy measure prior to discussing test results with a patient, it should take the form of something quickly and easily scored. The MC-GNS offers that functionality.

STUDY 3

Objectives

Study 3 assessed the MC-GNS for convergent and divergent validity by comparing MC-GNS scores with participants' ACT (American College Testing) examination scores. If the MC-GNS measures numeracy, it should be positively correlated with more comprehensive measures of mathematical ability (e.g., ACT Math subscores) but not with other similar measures of subject achievement (e.g., ACT Reading and English subscores).

Methods

Data were collected from 141 undergraduate students enrolled in introductory psychology courses at a large Midwestern university, using an online survey platform. Participants included 67 female students and 74 male students (M age = 19.46 years, SD age = 1.86). All participants were given the MC-GNS and then also asked for permission to access their ACT scores through the university's record system. Participants were then thanked and shown debriefing information.

Key Results

Only 63 of the 141 participants provided permission to access their ACT scores, and subject scores were not always present in the student information system. Of the 63 participants used in the mediation analyses, one person did not have a Reading subject score and four did not have Math subject scores. Regression methods were used to fill in these gaps (missing subject scores were computed using a regression equation containing MC-GNS scores and ACT composite scores).

To test convergent validity of the MC-GNS, we examined the correlation between the MC-GNS and ACT Math score. Indeed, these two measures were positively correlated ($r[61] = .54$, $p < .001$). To test discriminant validity, participants' MC-GNS scores were compared to ACT English and ACT Reading scores, two theoretically unrelated constructs from mathematical ability. Interestingly, both ACT English ($r[61] = .43$, $p = .001$) and ACT Reading ($r[61] = .47$, $p < .001$) were also positively correlated with MC-GNS, posing a potential problem for the multiple-choice numeracy scale.

It is plausible, however, that comprehension of the English language (ACT English) and reading ability (ACT Reading)

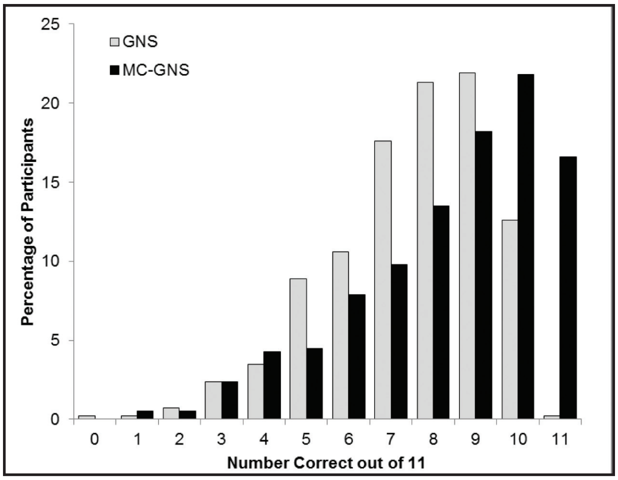

Figure 1. Distributions of scores for the GNS in Study 1 (Lipkus et al., 2001) and MC-GNS in Study 2. GNS = General Numeracy Scale, MC-GNS = Multiple-Choice General Numeracy Scale.

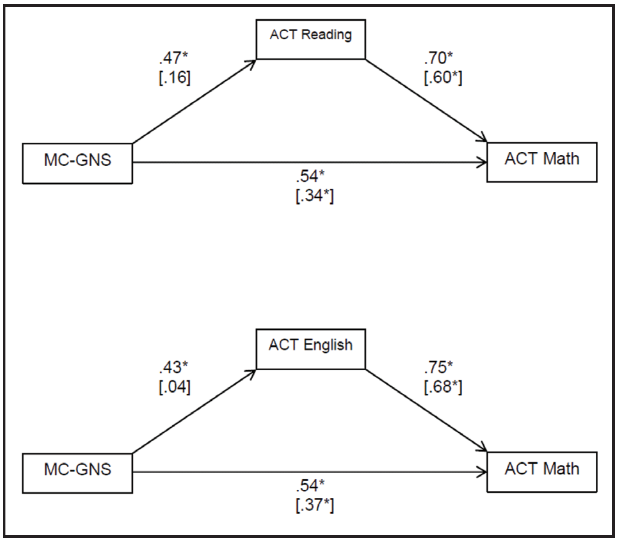

Figure 2. Zero-order and partial correlations (in brackets) between the revised MC-GNS and ACT subscores. ACT = American College Testing; MC-GNS = Multiple-choice General Numeracy Scale.
*$p < .05$.

are related to performance on the MC-GNS (as well as other numeracy scales) because these scales are composed of word problems. Careful reading and comprehension are at least a part of the response process. The question, then, is whether the MC-GNS (and other numeracy scales) are measuring more than just a person's ability to read and comprehend words. The most parsimonious explanation would be that the MC-GNS has a relationship with both ACT Reading and ACT English scores, but that those relationships are fully mediated by ACT Math.

Bivariate and partial correlations were calculated between the four variables. When controlling for ACT Math, ACT Reading and ACT English were no longer significantly correlated to MC-GNS. However, even when controlling for both ACT Reading and ACT English, the MC-GNS showed a significant positive relationship with ACT Math scores (**Figure 2**).

Thus, only the Math ACT scores were significantly correlated with the MC-GNS once the influence of other ACT scores was removed. This suggests that the MC-GNS in fact measures something above and beyond mere English competency and reading ability (although those are also relevant); specifically, the MC-GNS is measuring what it is intended to measure—numeracy.

CONCLUSIONS

The ubiquity and significance of numerical information in areas such as health risk and medical decision-making increasingly demands a functional level of numerical understanding. Numeracy varies considerably within the population, and this variability empirically accounts for differences in decision-making regarding health and finances (Finucane & Gullion, 2010; Peters, Hart, & Fraenkel, 2011). However, as pointed out previously by others, many numeracy scales have not been rigorously evaluated and existing evaluations have been heterogeneous in approaches (Brooks & Pui, 2010; Liberali et al., 2012; Schapira et al., 2009; Schapira et al., 2012). The current studies assessed the empirical and theoretical content of the GNS, developed improvements to this scale, and documented that these were useful advances.

Numeracy, as assessed with the GNS, can be characterized by both one- and four-factor models. Although the four-factor model seemed stable and patterns were relatively consistent with both the original study (Lipkus et al., 2001) and additional research (Liberali et al., 2012), the single-factor model also fit well and showed loadings for all 11 GNS items. The existence of two plausible factor structures may be problematic.

One possible explanation for a multifactor structure of the GNS was that it is composed of mixed question types, and this can lead to factors based on similar question format rather than question content (Tabachnick & Fidell, 2007). The factor structure of a multiple-choice version of the GNS (the MC-GNS), however, still indicated multiple factors (although up to two factors may appear to be eliminated in comparison to the GNS). Thus, the multifactor results for the GNS scale cannot be ruled out as solely the result of varied question format.

In addition to being easier to administer and score, the MC-GNS exhibits psychometric properties at least as good as, if not better than, those of the GNS. The MC-GNS has good convergent and discriminant validity (Study 3). The MC-GNS produced a slightly higher distribution of scores compared with the GNS, but the relatively lower difficulty of the MC-GNS may make it more usable for general population samples. Further research with broader samples than those used here (i.e., college students) can bolster these assessments. Collectively, the present studies indicate that the MC-GNS is better or equal in psychometric quality relative to the GNS, but with greater utility in terms of scoring ease in health and other applied settings. As a practical matter, applied uses of the MC-GNS may calculate a single composite score for the scale based on the evidence of a plausible single factor structure. The true structure of numeracy could be more complex, but at this point that structure is still uncertain.

The structure of numeracy, and the MC-GNS in particular, is currently more of a theoretical issue. The MC-GNS (and GNS) could be a single construct or it could have a multifactor structure of some sort. Liberali et al. (2012) suggested a mapping of the factor structures from multiple numeracy-type scales onto a dual-process model, but the present results do not fit well onto this previous mapping. More recently developed numeracy scales continue this bifurcated view of underlying structure, with some scales moving toward a multifacted approach (e.g., differentiating domains of number sense, tables and graphs, probability, and statistics) (Schapira, et al., 2012) and other scales moving toward a unidimensional approach (Cokely, Galesic, Schulz, Ghazal, & Garcia-Retamero, 2012). For the single factor approach to succeed there must be an account of why multiple factors consistently show up. For the multifactor approach to succeed, it must have a compelling explanation of what these factors are and why they exist. If multiple dimensions of numeracy do exist, an immediate applied question becomes which dimensions are most central for assessing health literacy.

In the meantime, the MC-GNS provides a more structured measure of numerical literacy with many potential benefits. The MC-GNS is easier to administer, with no loss of psychometric quality. It also provides clearer and easier coding of both correct and incorrect responses, the latter of which may be useful for evaluating not only numerical literacy but diagnosing aspects of illiteracy.

REFERENCES

Brown, S. M., Culver, J. O., Osann, K. E., MacDonald, D. J., Sand, S., Thornton, A. A., . . . Weitzel, J. N. (2011). Health literacy, numeracy, and interpretation of graphical breast cancer risk estimates. *Patient Education and Counseling*, *83*, 92-98. doi:10.1016/j.pec.20010.04.027

Brooks, M. E., & Pui, S. Y. (2010). Are individual differences in numeracy unique from general mental ability? A closer look at a common measure of numeracy. *Individual Differences Research*, *8*(4), 257-265.

Chapman, G. B., & Liu, J. (2009). Numeracy, frequency, and Bayesian reasoning. *Judgment and Decision Making*, *4*(1), 34-40.

Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The Berlin numeracy test. *Judgment and Decision Making*, *7*, 25-47. doi:10.1037/t45862-000

Couper, M. P., & Singer, E. (2009). The role of numeracy in informed consent for surveys. *Journal of Experimental Research on Human Research Ethics*, *4*, 17-26. doi:10.1525/jer.2009.4.4.17

Finucane, M. L., & Gullion, C. M. (2010). Developing a tool for measuring the decision-making competence of older adults. *Psychology and Aging*, *25*, 271-288. doi:10.1037/a0019106

Frosch, D. L., & Kaplan, R. M. (1999). Shared decision making in clinical medicine: Past research and future directions. *American Journal of Preventive Medicine*, *17*, 285-294. doi:10.1016/S0749-3797(99)00097-5

Galesic, M., & Garcia-Retamero, R. (2011). Do low-numeracy people avoid shared decision making? *Health Psychology*, *30*, 336-341. doi:10.1037/a0022723

Hanoch, Y., Miron-Shatz, T., Cole, H., Himmelstein, M., & Federman, A. D. (2010). Choice, numeracy, and physicians-in-training performance: The case of Medicare Part D. *Health Psychology, 29*(4), 454-459. doi:10.1037/a0019881

Hess, R., Visschers, V. H. M., & Siegrist, M. (2011). Risk communication with pictographs: The role of numeracy and graph processing. *Judgment and Decision Making*, *6*(3), 263-274.

Hill, W. T., & Brase, G. L. (2012). When and for whom do frequencies facilitate performance? On the role of numerical literacy. *The Quarterly Journal of Experimental Psychology*, *65*(12), 2343-2368. doi:10.1080/17470218.2012.687004

Knol, D. L., & Berger, M. P. (1991). Empirical comparison between factor analysis and multidimensional item response models. *Multivariate Behavioral Research*, *26*, 457-477. doi:10.1207/s15327906mbr2603_5

Liberali, J. M., Reyna, V. F., Furlan, S., Stein, L. M., & Pardo, S. T. (2012). Individual differences in numeracy and cognitive reflection, with implications for biases and fallacies in probability judgment. *Journal of Behavioral Decision Making*, *25*(4), 361-381. doi:10.1002/bdm.752

Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making*, *21*, 37-44. doi:10.1177/0272989X0102100105

Moumjid, N., Gafni, A., Bremond, A., & Carrere, M. O. (2007). Shared decision making in the medical encounter: Are we all talking about the same thing? *Medical Decision Making*, *27*, 539-546. doi:10.1177/0272989X07306779

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.

Peters, E., Hart, P. S., & Fraenkel, L. (2011). Informing patients: The influence of numeracy, framing, and format of side effect information on risk perceptions. *Medical Decision Making*, *31*, 432-436. doi:10.1177/0272989X10391672

Peters, E., Hibbard, J., Slovic, P., & Dieckmann, N. (2007). Numeracy skill and the communication, comprehension, and use of risk-benefit information. *Health Affairs*, *26*, 741-748. doi:10.1377/hlthaff.26.3.741

Reyna, V. F., Nelson, W. L., Han, P. K., & Dieckmann, N. F. (2009). How numeracy influences risk comprehension and medical decision making. *Psychological Bulletin*, *135*, 943-973. doi:10.1037/a0017327

Schapira, M. M., Walker, C. M., Cappaert, K. J., Ganschow, P. S., Fletcher, K. E., McGinley, E. L., . . .Jacobs, E. A. (2012). The numeracy understanding in medicine instrument: A measure of health numeracy developed using item response theory. *Medical Decision Making*, *32*, 851-865. doi:10.1177/0272989X12447239

Schapira, M. M., Wallaker, C. M., & Sedivyvy, S. K. (2009). Evaluating existing measures of health numeracy using item response theory. *Patient Educational and Counseling*, *75*(3), 308-314. doi:10.1016/j.pec.2009.03.035

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*. Boston, MA: Pearson Education.

Uebersax, J. (2007). *TetMat user guide* (Version 2007.1.0.3).

Wood, S., Hanoch, Y., Barnes, A., Liu, P. J., Cummings, J., Bhattacharya, C., & Rice, T. (2011). Numeracy and Medicare Part D: The importance of choice and literacy for numbers in optimizing decision making for Medicare's prescription drug program. *Psychology and Aging*, *26*, 295-307. doi:10.1037/a0022028

Woods, C. M. (2002). Factor analysis of scales composed of binary items: Illustration with the Maudsley Obsessional Compulsive Inventory. *Journal of Psychopathology and Behavioral Assessment, 24*(4), 215-223. doi:10.1023/A:1020770831134

TABLE A

Pattern Matrix Loadings for GNS Data in Study 1 When Specifying a Minimum Eigenvalue of 1

GNS Item	Factor 1	Factor 2	Factor 3	Factor 4
1	–0.658	-	-	-
2	-	0.781	-	-
3	-	0.599	-	-
4	-	-	0.902	-
5	-	-	0.983	-
6	-	0.793	-	-
7	-	0.713	-	-
8A	-	-	-	–0.956
8B	-	-	-	–0.915
9	-	-	-	–0.739
10	-	-	-	-

Note. Data were rotated using a direct oblimin rotation (delta = 0). GNS = General Numeracy Scale.

TABLE B

Pattern Matrix Loadings for GNS Data in Study 1 When Specifying a Single Factor

GNS Item	Factor 1
1	.472
2	.558
3	.635
4	.539
5	.425
6	.506
7	.667
8A	.831
8B	.670
9	.648
10	.420

Note. GNS = General Numeracy Scale.

TABLE C

Pattern Matrix Loadings For MC-GNS Data in Study 2 When Specifying a Minimum Eigenvalue of 1

MC-GNS Item	Factor 1	Factor 2	Factor 3
1	-	-	0.626
2	-	-	0.568
3	-	-	0.701
4	0.829	-	-
5	0.764	-	-
6	-	0.885	-
7	-	0.893	-
8	-	-	0.494
9	-	-	0.500
10	-	-	0.634
11	0.912	-	-

Note. Data were rotated using a direct oblimin rotation (delta = 0). MC-GNS = Multiple-Choice General Numeracy Scale.

TABLE D

Pattern Matrix Loadings for MC-GNS Data in Study 2 When Specifying a Single Factor

MC-GNS Item	Factor 1
1	.661
2	.584
3	.457
4	.811
5	.676
6	.651
7	.599
8	.541
9	.665
10	.469
11	.789

Note. MC-GNS = Multiple-Choice General Numeracy Scale.