* Although numeracy was initially described as a unidimensional construct, more recent research suggests that it is more complex. For example, measures of subjective numeracy predict different behaviors than measures of objective numeracy, and research by Ellen Peters unpacks numeracy into numeric confidence and ability. Further, there are criticisms that measures of numeracy are conflated with literacy and logic given that numeracy is often assessed via word problems. Review the various measures of numeracy and discuss their predictive validity, considering a larger umbrella of related constructs including health literacy (which often involved comprehension of numerical information) and graph literacy. Which measure (or measures) would you recommend including in studies of medical decision making? (these recommendations may differ by context)

Comprehensive Exam Answers: Victoria Shaffer

What measurements of numeracy exist? (objective, subjective, numeric confidence, word problems, etc.)

Understanding of numerical information is vital in many fields, however, given that cross-communication in differing fields of research is not yet widespread and thorough, many different measures of numeracy therefore exist. It is important to note that for these various measures of what is considered under the umbrella macro-construct of ‘Numeracy’, that in many cases, they are measuring slightly different concepts. A brief review of some of the various measures of numeracy in contemporary literature follows.

One of the original measurements of numeracy in the field of psychometric measurement, Subjective numeracy is also seen as one of the easiest ones to measure (Thompson, Mielicki, Rivera, Fitzsimmons, Scheibe, Sidney, Taber, and Waters 2022). Subjective numeracy is defined by Dr. Angela Fagerlin as the “self-assessment of quantitative ability”. As subjective numeracy relies on a self-assessment and not on actual mathematical calculations, the measure can be administered more quickly, and is easy to adapt for digital use (Fagerlin, Zikmund-Fisher, Ubel, Jankovic, Derry and Smith 2007). It is noteworthy that the Subjective numeracy, as measured using Fagerlin’s Subjective Numeracy Scale (SNS), is significantly correlated with the Lipkus objective numeracy scale, yet regularly is seen by participants as both less stressful and less frustrating to complete. Regarding practical applications, high subjective numeracy scores don’t just measure self-assessment about the ability to work with fractions and percentages, but ‘numerical confidence’ as well (Peters, Tompkins, Knoll, Ardoin, Shoots-Reinhard, and Meara 2019). This numerical confidence even has an interaction effect with actual numerical ability, as Peters research on financial and medical outcomes indicates. Humans with high confidence and ability, have the best financial and medical outcomes, and those with high confidence and low ability, have the worst outcomes. Independently, low subjective numeracy also predicts less persistence with difficult or impossible mathematical problems.

Another of the early measures of numeracy in the field of psychology, Objective numeracy, defined as “The ability to understand and use probabilistic and mathematical concepts”, has been used by psychologists to develop greater understanding of risk communication, especially how it relates to healthcare (Tompkins 2015). Subjective numeracy has generally been measured through various calculation exercises that consist of word problems and/or interpretation of tables (e.g., “If the chance of getting a disease is 10% how many people would be expected to get the disease out of 100?”; Lipkus, Samsa, Rimer, 2001). Some examples include the Lipkus Objective Numeracy Scale, the Rasch Numeracy Scale, and the Berlin Numeracy Test. In contrast to ‘subjective’ numeracy, objective numeracy is seen as more difficult and time intensive to measure. Considering these issues, why would we care to measure subjective numeracy? According to Organization for Economic Co-operation and Development (OECD) data, approximately a third of adult Americans are unable to locate basic quantitative data and utilize it to solve problems as simple as adding two numbers together (OECD, 2013). While numeric confidence doesn’t necessarily have to be tied to ability, we still need data on actual ability, and objective measures of calculation should exist to capture that nuance.

It is important to note that the field of psychometrics does not have a monopoly on the concept of measuring numerical ability; Math cognition researchers have been measuring what many psychologists would consider Numeracy under the broader umbrella of “Math Cognition Measures” for decades. These consists of several measures from the field of math cognition that directly measure various sub-elements of numerical cognition skill. General magnitude understanding, for example, can be operationalized as precision when estimating values on a number-line, with larger percentage of absolute error indicating worse numerical ability. Precise understanding of large magnitudes is tested by determining where to place a 1-million-unit marker on a line between zero and 1 billion (shockingly, as many as half of adults incorrectly believe one million sits at the midpoint between 0 and 1 billion; Landy, Silbert, and Goldin 2013). Ability to reason with fractions and ratios understanding more generally, can be tested by arranging a mix of various irregular fractions into correct order. Accuracy in comparing magnitudes can be tested by briefly looking at two fractions and being asked to select which one was the largest, greater correct proportions indicate better numerical skill at assessing relative magnitude. Arithmetic accuracy is generally measured with three sub-measures, fractional arithmetic, whole number arithmetic, and multi-step arithmetic. Fractional arithmetic is assessed with performance on basic addition, subtraction, multiplication, and division, involving one whole number and one fraction, in operation with each other. Greater accuracy indicates more fractional arithmetic ability. Whole number arithmetic is assessed with a ‘race’ to complete as many two-digit whole-number arithmetic problems as possible, split evenly between addition, subtraction, and multiplication. Greater volume of answers completed correctly indicates more whole number arithmetic fluency. Multi-step arithmetic is assessed with problems wherein common denominators for numbers need to be determined, and then these fractions need to be added or subtracted to each other. Again, greater accuracy indicates more multi-step arithmetic ability. There are many advantages of “Math Cognition Measures’, that is in contrast to measures of objective numeracy designed by psychometricians. One large advantage is that many of these math cognition measures can work perfectly fine with natural numbers (positive integers), whereas the fractions and percentages needed for many measures of objective numeracy can often be more confusing (Thompson 2022). Furthermore, many objective numeracy measures are presented using word problems (e.g. “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 1,000 people each by a single ticket to BIG BUCKS?”), which means that there are additional non-math barriers to solving the problem, making it difficult to disentangle the ‘pure’ effect of mathematical ability. Math cognition researchers furthermore believe that the psychometric objective and subjective measures of numeracy predict health-decision choices so well because they all relate to the fundamental mathematical ability of understanding various ratios. We can see that the existing objective measures directly assess the ability to calculate with ratios, and the subjective measures comprise entirely of self-assessments about one’s willingness to work and understand ratios. Medical decision making regularly involves rational numbers and requires judgement of risk, for oneself and others. Given these characteristics, direct “Math Cognition Measures” that address these can be seen as particularly valuable.

While measuring numeracy in a vacuum does have some appeal to researchers, numeracy is also looked at for it’s predictive validity in various other fields and measures, and value that can be obtained by having strong predictive validity. Some of these findings are relatively obvious, for example, weak numeracy skills earlier in life relate to poorer performance in math as a subject in later childhood (Siegler, Duncan, Davis-Kean, Duckworth, Claessens, Engel, Susperreguy, and Chen 2012). Likewise, weaker subjective numeracy skills predict less confidence in using math skills, less willingness to use them often, less persistence with numeric tasks and less understanding of numerical information in general (Peters 2019; Peters 2020). Indeed, it is vital to note that this weaker understanding of numerical information in general is particularly harmful in the medical context, as there is worse understanding of the magnitude and relative risk of side effects (Garcia-Retamero 2017). Weak subjective numeracy doesn’t just affect perception and willingness to use numbers, Peters finds that those with weak subjective numeracy also make normatively ‘worse’ choices. Likewise, greater numeracy is linked to better choices, however, this is not just due to some normative theory, but due to several steps that those with more numeracy methodically take. Those with greater numeracy make more explicit expected value calculations as well as verbalize those calculations, including transforming probabilities and translating percentages. There is also greater and more elaborative processing as a whole, with more time deliberating over choices, not just making calculations more quickly. Surprisingly, there is also more emotional reactions to probability, which can be seen as higher sensitivity to changes in probability.

Some of the predictive validity of low subjective numeracy is not nearly as obvious as making ‘worse’ choices or being less able to use math. Ciampa, Osborn, Peterson, and Rothman (2010) find, for example, that medical patients with low subjective numeracy perceive that they receive low quality communication from healthcare providers. Additionally, weaker numeracy skills are linked to more interest in a ‘paternalistic’ model of medical decision making, with less interest in traditional ‘shared decision making’ and a desire for more ‘passivity’ than those with higher numeracy (Garcia-Retamero 2017; Galesic and Garcia-Retamero 2011). The combination of deference to physician recommendation, and the simultaneous perception that these physician recommendations are poor quality is very alarming. When we additionally consider the fact that those with low numeracy are more easily biased by the way health-related numerical information is framed, physicians must significantly consider how they communicate with patients (Garcia-Retamero 2009). Fortunately, physician numeracy can directly address some of the issues with lower patient numeracy (Petrova, Kostopoulou, Delaney, Cokely, and Garcia-Retamero 2018). Physicians with high numeracy skills have 7 times the odds as compared to low numeracy physicians of presenting the risks of screening in a complete and balanced way. This was due to a combination of being less likely to mistakenly believe that increased detection or survival rates from screening show that screening saves lives and being more likely to communicate mortality rates and risks from overtreatment. In contrast, physicians with weak numeracy provide worse Medicare recommendations, inaccurate inferences, and inability/reluctance to communicate numerically with patients.

Another major field in which Numeracy skill is very predictively valid is that of risk perception in general. Keller and Siegrist (2009) find that, when looking at various risk communication formats (paling perspective scale, pictogram, ratio chart, table, etc.), those with lower numeracy did not distinguish between high and low risk for either down syndrome or colon cancer, when give nay of the formats. Those with high numeracy were able to distinguish between high and low risk when provided information in a ratio format, but not the others. Similarly, Hess, Visschers, Siegrist, and Keller (2010) find that there is strong correlation between subjective numeracy and visual perception of risk communication. Higher subjective numeracy is linked to more time spent looking at risk visualizations, and a higher total number of ‘gaze events’ where attentiveness is clearly focused on a single element. Those with greater subjective numeracy generally can look at and absorb more risk information faster, in a shorter period of time (more efficiently). Another common risk circumstance is being presented with medical test results. Zikmund-Fisher, Exe, and Witteman (2014) find that in these circumstances, those with weaker numeracy have trouble recognizing that a jump in A1C from slightly to moderately out of range was worth immediate response. In practice, there is an important distinction drawn here between patients “knowing” their test numbers, versus grasping the personal meeting of that data. In general, *a-priori,* patients cannot get a feeling of ‘goodness’ and ‘badness’ from merely seeing their data, especially those with weak numerical skills.

“Math Cognition Measures”, apart from the objective and subjective measures of numeracy that originate from psychometrics, also have strong predictive validity for many vital concepts. For example, performance on the ‘number line test’, considered a ‘hard objective numeracy measure’, correlated strongly with future math achievement. Furthermore, it predicts health decision-making performance, above and beyond objective and subjective measures commonly used in psychology (Thompson 2022). Furthermore, in the specific context of health decision making, greater performance on ‘magnitude estimation’ and ‘multi-step arithmetic’ tasks predicted greater health-decision making accuracy, whereas traditional measures of objective and subjective numeracy did not add any additional predictive validity.

Considering how many aspects of life can be affected by numeracy, it is no surprise that there is a great deal of related constructs that conversely have an impact back on numeracy. Understanding how these various constructs interact with numeracy in all its forms is necessary to build a holistic understanding of the meaning of numeracy. A brief review of several constructs follows.

Math anxiety can be defined as “A feeling of tension, apprehension, or fear that interferes with math performance” (Ashcraft 2002). Math anxiety is a worthwhile related construct to examine, as while it is associated with performance on objective and subjective measures of numeracy, it provides additional predictive validity above and beyond just that (Thompson 2022). It is interesting to note, that High math anxiety is associated with poor risk comprehension, however, when controlling for objective numeracy, this effect disappears. Likewise, math anxiety is negatively related to subjective numeracy, reducing risk-comprehension ability (Rolison, Morsanyi, and Peters 2020).

According to math cognition researchers, “Natural Number Bias” is thought to be the overall driver/mechanism behind what psychometricians would define as ‘poor numeracy’. Natural number bias can be defined as “The tendency to apply natural number properties in tasks with rational numbers, even when this is inappropriate” (Ni and Zhou 2005). Generally, this encompasses several sub-concepts of bias within itself, such as Ratio bias, denominator neglect, and the 1-in-X phenomena. For example, patient feedback on incorrect risk perception directly mention that they focused on numerators in isolation, and that covid lethality was undersold as it compared absolute numbers to flu deaths (Thompson 2022). In another example, the 1-in-X phenomena, people overestimate risks represented with one as the numerator (e.g., 1 in 4) relative to equivalent risks with a larger numerator (e.g., 30 in 70). Math cognition theory indicates that the underlying elements are due to two primary reasons. First, perceptual limits. A person can see 4 > 2 much faster than 104 > 102, even if objective difference is the same, the relative difference is much smaller. The second is that Natural numbers are very common! For example, ½ is seen much more often than 25/50 for example, as 1 and 2 are super common numbers. Even in adults, 100 pennies are seen as worth more than 1 dollar. Due to how common natural numbers are, people often incorrectly apply natural number properties incorrectly.

Gist understanding, primarily popularized by Valarie Reyna’s Fuzzy Trace Theory is seen as a parallel structure to some Math Cognition Researcher’s views on Numeracy (Reyna and Brainerd 1995). The core concept is that the general math cognition concept of ‘magnitude’ is seen as the building block of math, or to phrase differently, relative understanding of ‘magnitude’ is considered ‘gist understanding of math’ holistically (Thompson 2022). Generally, understanding the gist of magnitude allows for people to estimate what is needed for good decision making, and gist usage increases with age (Van Hoof, Verschaffel, De Neys, and Dooren 2020). Math cognition researchers describe the process of learning math as going from the gist understanding (rough magnitudes) to exact information, to using gist reasoning by considering approximate magnitudes. Indeed, dual process theory in general clearly explains how people can have all the required knowledge and skill to solve a task, but because they are affected by erroneous intuitive reasoning, make mistakes. In this case, people must inhibit their natural automatic bias towards using whole number responses to engage in the effortful process of leveraging rules to process the magnitude of the ratio. Even for those with high numeracy, when considering the effect of reaction time (system 1 vs system 2 thinking) people have greater accuracy on congruent (tasks where natural number knowledge will result in the correct answer, e.g. which fraction is bigger? 6/11 or 4/8) vs incongruent (tasks where natural number knowledge will result in the wrong answer, e.g. which fraction is the larger one? 6/11 or 3/4?). Likewise, even for participants with great accuracy, incongruent tasks still took longer than congruent ones. Notably, when time pressure was increased, there was a decrease in general accuracy overall, but a much greater decrease in accuracy for incongruent tasks vs congruent ones. We must acknowledge that there is a large difference between inaccuracy originating from lack of knowledge or misconceptions, versus mistakes made due to an intuitive response, as it is vital for determining the best pathway to address the problem moving forward.

On the final set of constructs related to numeracy, we look at the concepts of health and risk literacy. Health literacy is defined as “The degree to which individuals have the capacity to obtain, process and understand basic health information and services needed to make appropriate health decisions” (Nutbeam 2008). Lower health literacy is linked to less knowledge about medications taking, less ability to read and understand the labels of the medication itself, difficulties adhering to hospital instructions, and increased mortality. For older African-Americans with HIV specifically, weak health literacy independently is correlated with poorer treatment adherence and viral load suppression for (Gakumo, Raper, Cerice, Stand-Gravois, and Mugavero 2016). Likewise, research by Zikmund Fisher and colleagues (2014) shows that health literacy independently predicts difficulty with identifying out of range test results in health information. Health literacy is also strongly correlated with better accuracy when looking at urgency differences scores for potential medical issues, essentially being able to correctly identify that an out of range result is an actual problem (Zikmund Fisher, Scherer, Witteman, Solomon, Exe, Tarini, and Fagerlin 2017). Additionally, it is moderately correlated with numeracy. Note that this is not a unidirectional effect, 10% of subjects had lower numeracy and higher literacy, 6% had higher numeracy and lower literacy. Shockingly, the combined effect of lower literacy and lower numeracy more than halves probability of identifying out of range values (77% to 38% for diabetics and 65% to 30%). Lastly, the perceived usefulness of medical information was highly correlated with both health literacy and numeracy.

Similarly to health literacy, graph literacy is defined as “The ability to understand graphically presented information”, and just like health literacy, graph literacy has significant implications with numeracy (Okan, Garcia-Retamero, Cokely, and Maldonado 2011). Generally, patients with high graph literacy can benefit more from information presented using icons and have greater confidence and accuracy in risk reduction estimates. For those with weak graph literacy, there was less improvement in accuracy and none on confidence when given icons. Thus, graph literacy can affect the extent to which people benefit from visual aids designed to overcome judgement biases due to weak numeracy! Zikmund-Fisher and colleagues (2014; 2017) also find that graph literacy is strongly linked to risk perception and recall for pictographs. Generally, those who were more numerate or graphically literate were able to have a stronger correlation between the actual risk of getting an illness and the risk perception they had of getting the illness (thus being more well calibrated). Graph literacy likewise has a strong relationship when predicting the ability for patients to distinguish between urgent and non-urgent anomalies in laboratory test results.

There are several measures that I would recommend for studies of medical decision making, with some differences in selection based on various contexts. First, we must note that the public has extremely poor baseline numeracy, by proportion for numeracy skills in the American population, 22% are below basic, 33% basic, 33% intermediate, and 13% are proficient (Health Literacy and Numeracy: Workshop Summary, 2014). Additionally, medical researchers believe that ‘proficient’ is seen as the level of skill needed to make good health-care choices in an open marketplace. Considering that only 13% of the American population has reached that point, this frames how we should measure numeracy for the average medical patient. It’s also important to note that patients in many cases are making choices under super high levels of affect, as relative differences in affective engagement can change the ability and resources available to engage in critical thinking. Additionally, health problems can both directly and indirectly impact the capacity for thinking in the dual process concept. Finally, when numeracy is measured under normal circumstances, if the patient finds themselves in abnormal circumstance (common when dealing with health problems), then there is a real risk of the patient’s ‘functional’ numeracy level being perhaps significantly lower than what was originally measured, resulting in perhaps different health choices or a different approach being better suited for that patient.

Generally, my belief is that a numeracy measure should be quick and easy to record, requires little additional resource or input from already overtaxed healthcare workers, and ideally can be provided/done digitally/virtually, as access to personal computing resources has obviously exploded in the past decade. Schapira, Fletcher, Ganschow, Jacobs, Walker, Smallwood, Gil, Faghri, Kong, Yen, McDunn, Marcus, and Neuner (2019) found that when using a computer test of health numeracy for breast cancer consultation, the tool was rated as very good or excellent by 65%, and fair by 33%. Additionally, as the tool was computer provided, it had an automatic report-generation feature that could output in both English and Spanish, helping to teach practitioners how to communicate to patients with higher or lower numeracy, as well as providing a physical visual aid. It is also noteworthy that physicians generally rated the tool as excellent or very good clarity, and relatively helpful. In fact, physicians ended up modifying their communication somewhat to a lot in approximately a third of their consultations! Given this, I believe that the ‘Number Line Test’ from the field of math cognition would work very well, and if needed or as an alternative, the 4-item Subjective Numeracy Scale would be also very appropriate. The number line would be a good measure, as number lines illustrate magnitude, which primes subjects to use the ‘correct’ skills, allows for comparison of magnitudes to each other very easily and directly measures understanding spatial-numeric relationships in an objective way, without requiring interpretation of word problems or generation of excess anxiety. The SNS scale would be a good measure as it captures the unique effects of numerical confidence, apart from actual numerical ability, in a short and easy to administer way. Additionally, both measures should be able to be done ahead of time using a simple online questionnaire or phone application, ideally.