Predicting Numeracy Using Speech Recognition

Introduction

Cognitive Reflection Tasks as powerful predictors of *Cognitive Abilities*:

A great deal of research (Cokely et al. 2012; Reyna et al. 2009, Sakurai & Kosaka, 2021,Hunter, 1986, Oechssler et al. 2009) has been aimed at evaluating the skills involved in performing tasks associated with perception, learning, memory, reasoning (cognitive abilities; American Psychological Association, n.d) in order to deduce reliable prediction patterns. Performance during these cognitive tasks often allows for the identification of diverse characteristics in everyday life. People with higher cognitive ability differ from those with lower cognitive ability in a variety of important ways (Frederick, 2005). On average, they live longer, earn more, have larger working memories (i.e., the ability to store and process information in real time), have faster reaction times, yet at the same time they are *more* susceptible to visual illusions (Jensen, 1998). All are factors that are crucial to humankind's ability to succeed. Thus, their prediction through different methods is one of the more valuable areas of research from a practical standpoint. It is no wonder then why so many studies (Cokely et al. 2012; Reyna et al. 2009, Sakurai & Kosaka, 2021, Hunter, 1986, Oechssler et al. 2009) have pursued this area of research.

For example, after having implemented the Cognitive Reflection Test (CRT), Frederick (2005) reported a correlation of .44 between performance on the CRT and SAT total scores, as well as a .43 correlation with performance on the Wonderlic IQ test (Frederick, 2005). The Addenbrooke's cognitive examination III (Bruno & Schurmann, 2019), a screening test that is composed of tests of attention, orientation, memory, language, and visual perceptual, has proved to be useful in the detection of cognitive impairments. Most commonly, many colleges and universities in the United States require students to submit scores from the SAT test as part of the application process to show colleges students' knowledge in regard to reading, writing and math comprehension and how well they can apply that knowledge (US Department of Homeland Security, n.d.). Those scores are often necessary requirement to important scholarships and "prestigious schools". Thus, performances in cognitive tasks have historically been shown to be powerful predictors of certain intellectual abilities in many different areas.

The Berlin Numeracy Test (Cokely et al. 2012) - which is a fast test of statistical numeracy that leverages available computing technology and internet accessibility (e.g., online data collection and scoring; accessible via smart phones and other internet ready devices) - has proven particularly effective in assessing risk literacy and decision-making abilities. Students' critical thinking was found to predict their academic performance after controlling for the effects of English fluency and general intellectual ability (Lun et al. 2010). The results of Bruine de Bruin et al.'s research (2007) revealed strong relations between intelligence and general decision-making skill. Although the different tests are intended to measure conceptually distinguishable traits, there are many likely sources of shared variance. For example, although the CRT (Frederick, 2005) is intended to measure cognitive thinking, performance is surely aided by reading comprehension and mathematical skills which the ACT and SAT also measure but only as measures of academic "success" (Frederick, 2005).

Numeracy Test as the best predictors of *Cognitive Abilitie*s

Of all the cognitive tests, those that are related to mathematical skills stand out as being more accurate and efficient than the others in overall prediction of superior decision-making performance (Cokely, 2016). Mathematics skills are among the most influential educational factors contributing to economic prosperity in industrialized countries (Hunt & Wittmann, 2008). This influence of mathematical skills on key financial behaviors is supported by several studies (Lusardi & Mitchell, 2011; Alessie et al. 2011, Fornero & Monticone, 2011). For example, Lusardi and Mitchell (2011) suggest that those who are not numerate are less likely to accumulate wealth. These skills are referred to as "Numeracy".

Numeracy

Numeracy can be defined as a set of numerical skills emphasizing mental and written calculations (Brown et al. 1998) and the perceived ability to perform various mathematical tasks (4). This definition encloses the two popular kinds of numeracy. The first one emphasizes mental and written calculations(Brown et al. 1998) and is known as Objective Numeracy whereas the second one focuses on the self-perceived ability to perform various mathematical tasks and is known as Subjective Numeracy.

Since the 1990s, research on the role of mathematical skills in decision making has grown from practical efforts to improve risk communications, particularly in health and medicine (Cokely et al. 2012 ; ,Lau et al. 2022, Reyna, 2009). Numeracy research is generally concerned with effective everyday problem solving for activities like evaluating medical treatment options, political claims, and financial products (Cokely et al. 2012; Reyna, 2009; Cokely et al. 2012, Cokely et al. 2014; National Research Council, 1990). Numeracy has two components: Statistical Numeracy (calculations with percentages) and conventional numeracy (Geometry and Algebra). Numeracy skills have been found to be strong predictors of key cognitive skills such as decision-making skills or risk assessment (Cokely et al. 2012;), more so than other measures of cognitive ability (Cokely, 2016). Full scale numeracy (i.e., statistical numeracy + conventional numeracy) as measured with the Berlin Numeracy Components Test alone accounted for 34% of the total variance in overall decision-making skill (Figure 1). In comparison, the best combination of all other cognitive ability and numeracy instruments (e.g., fluid intelligence, cognitive reflection, health literacy) accounted for about 30% of overall decision-making skill (Figure 1). Thus, despite taking more than ten times longer to complete, all other cognitive ability tests combined offered significantly less predictive power compared to the single brief Berlin Numeracy Components test. Analyses further indicated that the statistical numeracy (i.e., probability + operations) sub-test alone explained 33% (Figure 1) of the total decision-making skill variance (as a single predictor), such that 97% of the predictive power of full-scale numeracy was shared with the statistical numeracy subscale (Cokely, 2016). Thus, when the different types of numeracy are combined (and all other measures of cognitive ability excluded), the predictive power increases.

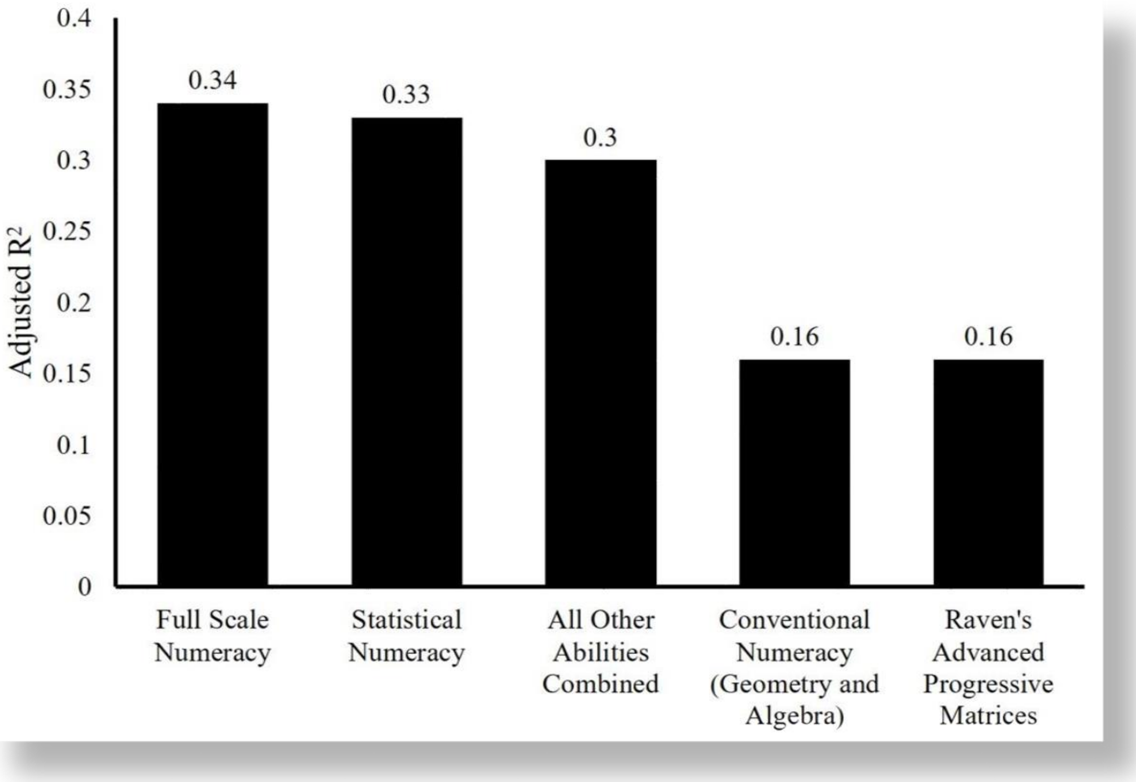

Figure *1*: The figure presents data from a sample of young adults displaying the total variance explained in overall superior making performance (i.e., general decision-making skill) by each of the single predictors alone (Cokely et al. 2016).

An Alternative to Traditional Numeracy: Subjective Numeracy

While the most common form of numeracy seems to be objective numeracy (based on mathematical performance only), the Subjective Numeracy Test which is a self-report measure of perceived ability to perform various mathematical tasks and preferences for the use of numerical versus prose information (CBSSM, 2020) was found to be highly correlated with objective numeracy with correlations between .63 and .68 (Fagerlin et al. 2007). To assess the results accurately and consistently, an eight-item measure, the Subjective Numeracy Scale (SNS), was developed through multiple rounds of testing (Fagerlin et al. 2007). The SNS consists of four questions asking respondents to assess their numerical ability in different contexts and four questions asking them to state their preferences for the presentation of numerical and probabilistic information (CBSSM, 2020). Even though there is very little evidence as to which provides better results between objective and subjective numeracy, the subjective numeracy scale represents a very good alternative to objective numeracy. The advantage of such a method compared to objective numeracy is that it can be administered in much less time, with fewer constraints, and was perceived as less stressful by respondents while offering fairly similar results. For example, 50% of participants who completed the SNS volunteered to participate in another study, whereas only 8% of those who completed the Lipkus and other objective Numeracy scales volunteered (Fagerlin et al. 2007).

In short, The SNS correlates well with mathematical test measures of objective numeracy but can be administered in less time and with less burden. In addition, it is much more likely to leave participants willing to participate in additional research and shows much lower rates of missing or incomplete data. (Fagerlin et al. 2007)

Speech Features to Predict Numeracy Skills

Studies (Darke, 1988; Humphreys & Revelle, 1984; Isen & Daubman, 1984; Isen et al. 1992) have shown that intense or even milder emotional states often have effects on social and economic interactions, decision making and rationality in general. Since improper self-control and excessive risk-taking (or avoidance) are often matters of material rationality (Pham, 2007), emotional states can help predict them. Most intense emotional states, except sadness, are accompanied by high levels of autonomic arousal, which is known to impair *working memory* capacity (Darke, 1988, Humphreys & Revelle, 1984). To make emotion recognition easier, Jin Q. et al (Jin et al. 2015) have generated feature representations from both acoustic and lexical levels. The experimental results show that late fusion of both acoustic and lexical speech features achieves four-class (angry, happy, sad, neutral) emotion recognition accuracy of 69.2% (Jin et al. 2015).

Similarly, with the proven effectiveness of numeracy tests to predict abilities such as decision-making skills, risk literacy or even financial behaviors, it would be interesting to consider ways to simplify and shorten the assessment of these skills as much as possible. The easiest and most suitable way for the most people could be the use of speech features. So far, although little to no research has focused on using speech features as means of predicting numeracy skills, there are certain strengths that make them very promising in this regard. The components of speech that are often used and that have offered acceptable levels of satisfaction are linguistic and acoustic features.

Speech Recognition Tools

Acoustic Features

Acoustic features have been the dominant features considered in speech emotion recognition literature (Jin et al. 2015). The acoustic features represent the pitch, unvoiced duration, shimmer, pause duration or speech rate (Mahajan & Baths, 2021) that someone can let appear during a speech. Acoustic features could be useful in the prediction of numeracy skills since they could allow one to evaluate a possible correlation between key factors such as confidence or spontaneity and the results of the numeracy tests collected from the participants especially in subjective numeracy tests. For example, a popular protocol for collecting speech samples is to ask volunteers to describe what they see in a picture. They are able to view the picture while they speak. The resulting acoustic speech files are then processed directly for acoustic features such as pauses and pitch contours (Sadeghian et al. 2021). To further optimize the results of these kinds of experiences, voice activity detectors are used. Voice activity detectors (VAD) are methods of detecting the presence of speech in an audio signal through algorithms. Voice activity detectors allow the evaluation and elimination of pauses or sequences of absence of speech that are not related to the ongoing study (Sadeghian et al. 2021). Several VAD algorithms are available (i.e., Likelihood Ratio Test-LRT-, Maximum Likelihood-ML-) (Jongseo et al. 1999). Acoustic features when used with another relevant feature are even more helpful and would allow a good prediction of numeracy performances and make it possible to simplify this process in an unprecedented way and would drastically reduce the time usually attributed to taking tests.

Linguistic Features

The other important speech features are the linguistic ones. The linguistic features refer to the structure of the language rather than its form (i.e., Nouns, adjectives, adverbs, etc.) (Sakurai & Kosaka, 2021). To use linguistic information of speech, it is necessary to convert speech into text using an automatic speech recognition (ASR) system. The text can also be manually transcribed for slightly better performance relative to an ASR, but when acoustic and linguistic features were used together this error was reduced (Sakurai & Kosaka, 2021).

The acoustic and linguistic features have proven to be very effective when used together rather than when used separately (Sakurai & Kosaka, 2021). In a study on the recognition of emotions through speech of 400 utterances (10 sentences × 4 emotions × 10 speakers), when only linguistic features were used, the recognition rate was 45.0% for the transcribed text and 39.75% for the ASR results; when both acoustic and linguistic features were used, the performance was 79.0% for the transcribed text and 77.25% for the ASR results (Sakurai & Kosaka, 2021).

The use acoustic and linguistic features in numeracy prediction could allow for shorter test times and eliminate any potential participant bias (trying to sound smarter, fearing being judged…) in their answers. Importantly, it would also provide a stress-free alternative to traditional numeracy tests to obtain answers that reflect the full potential of the participants. Acoustics and linguistics would make numeracy testing easier than ever by processing only speech snippets and delivering the corresponding results to testers or assessors instantly.

Method

Participants and procedure

A total of 400 participants, age 18 or older, are expected to take part in the study. No distinction of sex/gender will be made, and no monetary compensation will be offered. Participants will be recruited from across the HU campus through the university's internal email list, flyers, posters, etc.… Recruited participants will take part in three tasks: A picture description test, a text reading, and a Berlin numeracy test. Participants will complete the following tasks: Speech Recognition Tasks:

Picture Description:

For this task, we use an image (upklyak, n.d) retrieved from Freepik.com (and retouched using magiceraser.io & Canva.com) that each participant will be asked to describe in their own words. This will allow us to examine verbal fluency and take advantage of any unvoiced duration, shimmer, pause duration or speech rate (Mahajan & Baths, 2021) that appears during a participant's speech. We will clean the resulting text files by removing all marks of repetitions, hesitations, incomplete words, and pauses; consider 2-

consecutive nouns as a single noun and discard all part-of-speech tags that are not in the linguistic patterns (Hernández-Domínguez et al. 2018). The picture to be described is shown in Figure 2.

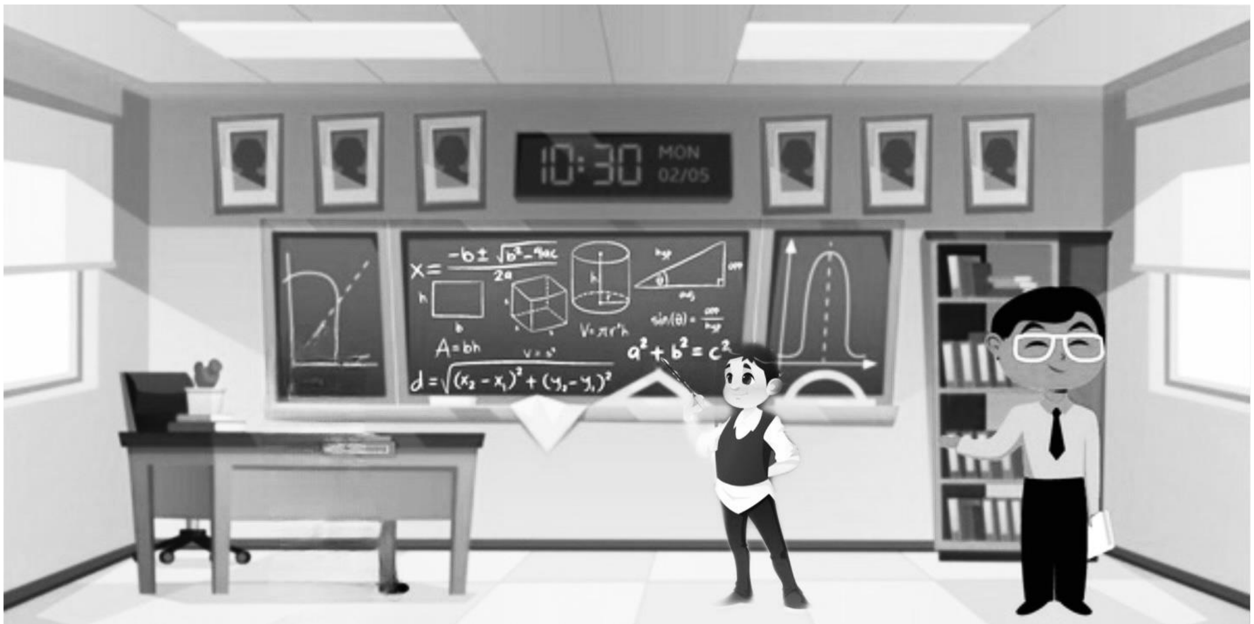

Figure 2: Retrieved from Freepik.com

<u>Text Reading Task:</u>

Similarly, each participant will have a text to read that would then be processed the same way as the picture description task. The text to be read is the following:

> *As of December 31, 2021, we had $49.7 billion of unsecured senior notes outstanding (the "Notes"). We issued $18.5 billion of Notes in May 2021, of which $1.0 billion was issued for green or social projects, such as projects related to clean transportation, renewable energy, sustainable buildings, affordable housing, or socioeconomic advancement and empowerment, and the remainder for general corporate purposes. We also had other long-term debt and borrowings under our credit facility of $924 million and $803 million as of December 31, 2020 and 2021. Our total long-term debt obligations are as follows (in millions):*

Amazon Annual Report (Amazon Inc., 2021)

Fever (temperature, ≥38°C) was reported after the second dose by 16% of younger vaccine recipients and by 11% of older recipients. Only 0.2% of vaccine recipients and 0.1% of placebo recipients reported fever (temperature, 38.9 to 40°C) after the first dose, as compared with 0.8% and 0.1%, respectively, after the second dose.

Systemic Reactogenicity (Polack et al. 2020)

Numeracy Test:

Based on the work of Cokely et al (2012), we first sought to find the best version of numeracy test that would fit the experiment. We used their Test Recommendation Tool and assigned 2-5 minutes as the time we can assign to each participant. The audience category chosen was "Highly educated sample" rather than "National or Less educated sample" given that the experience would be mainly pointed toward university student and faculty members cover the average person. Then, to make the test more accessible and the results easier to analyze, we chose the computer adaptive version of the Berlin Numeracy Test rather than the traditional version (on paper). The recommendation was the Adaptive Berlin Numeracy Test .

Instructions: Please answer the questions below. Do not use a calculator but feel free to use notes (i.e., scratch paper).

For Highly Educated Samples

Adaptive Berlin Numeracy Test Format

Instructions: Please answer the questions that follow. Do not use a calculator but feel free to use the scratch paper for notes.
[See Figure 1 for adaptive test structure.]

1. Out of 1,000 people in a small town 500 are members of a choir. Out of these 500 members in the choir 100 are men. Out of the 500 inhabitants that are not in the choir 300 are men. What is the probability that a randomly drawn man is a member of the choir? Please indicate the probability in percent.

______ %

2a. Imagine we are throwing a five-sided die 50 times. On average, out of these 50 throws how many times would this five-sided die show an odd number (1, 3 or 5)?
______ out of 50 throws.

2b. Imagine we are throwing a loaded die (6 sides). The probability that the die shows a 6 is twice as high as the probability of each of the other numbers. On average, out of these 70 throws how many times would the die show the number 6?
________out of 70 throws.

3. In a forest 20% of mushrooms are red, 50% brown and 30% white. A red mushroom is poisonous with a probability of 20%. A mushroom that is not red is poisonous with a probability of 5%. What is the probability that a poisonous mushroom in the forest is red?
____

Scoring = Based on answers to 2-3 questions following the adaptive structure. Correct answers are as follows: 1 = 25; 2a = 30; 2b = 20; 4 = 50.

REFERENCES

1. Alessie, R., van Rooij, M., & Lusardi, A. (2011). Financial Literacy, retirement preparation and pension expectations in the Netherlands. *Journal Of Pension Economics and Finance*, *10*(4), 525–545. https://doi.org/10.3386/w17109
2. Amazon Inc. (2021). *Annual reports, proxies and shareholder letters*. Amazon.com, Inc. - Annual reports, proxies and shareholder letters. Retrieved December 2022, from https://ir.aboutamazon.com/annual-reports-proxies-and-shareholder-letters/default.aspx
3. American Psychological Association. (n.d.). *Apa Dictionary of Psychology*. American Psychological Association. Retrieved October 17, 2022, from https://dictionary.apa.org/cognitive-ability
4. Brown, M., Askew, M., Baker, D., Denvir, H., & Millett, A. (1998). Is the National Numeracy Strategy Research-based? *British Journal of Educational Studies*, *46*(4), 362–385. https://doi.org/10.1111/1467-8527.00090
5. Bruine de Bruin, W., Parker, A. M., & Fischhoff, B. (2007). Individual differences in adult decision-making competence. *Journal of Personality and Social Psychology*, *92*(5), 938– 956.
6. Bruno D, Schurmann Vignaga S.(2019) Addenbrooke's cognitive examination III in the diagnosis of dementia: a critical review. Neuropsychiatr Dis Treat. 2019 Feb 15;15:441-447. doi: 10.2147/NDT.S151253. PMID: 30858702; PMCID: PMC6387595.
7. Cokely, E. T., Feltz, A., Ghazal, S., Allan, J. N., Petrova, D., & Garcia-Retamero, R. (2016). Skilled decision theory: From intelligence to numeracy and

expertise. *The Cambridge Handbook of Expertise and Expert Performance*, 476–505. https://doi.org/10.1017/9781316480748.026

8. Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Berlin numeracy test. *PsycTESTS Dataset*. https://doi.org/10.1037/t45862-000
9. Cokely, E. T., Ghazal, S., & Garcia-Retamero, R. (2014). Measuring numeracy. In B. L. Anderson & J. Schulkin (Eds.), *Numerical reasoning in judgments and decision making about health* (pp. 11-38). Cambridge University Press.
10. Cokely, E. T., Ghazal, S., Galesic, M., Garcia-Retamero, R., & Schulz, E. (2012). How to measure risk comprehension in educated samples. *In Transparent Communication of Health Risks* (pp. 29-52). Springer New York.
11. Fagerlin, A., Zikmund-Fisher, B. J., Ubel, P. A., Jankovic, A., Derry, H. A., & Smith, D. M. (2007). Measuring numeracy without a math test: Development of the subjective numeracy scale. *Medical Decision Making*, *27*(5), 672–680. https://doi.org/10.1177/0272989x07304449
12. FORNERO, E. L. S. A ., & MONTICONE, C. H. I. A. R. A. (2011). Financial Literacy and pension plan participation in Italy. *Journal of Pension Economics and Finance*, *10*(4), 547–564. https://doi.org/10.1017/s1474747211000473
13. Frederick, S. (2005). *Cognitive reflection and decision making*. Journal of Economic Perspectives. *19*, 25–42. https://pubs.aeaweb.org/doi/10.1257/089533005775196732
14. Hernández-Domínguez, L., Ratté, S., Sierra-Martínez, G., & Roche-Bergua, A. (2018). Computer-based evaluation of alzheimer's disease and mild cognitive impairment patients during a picture description task. *Alzheimer's & Dementia:*

Diagnosis, Assessment & Disease Monitoring, *10*(1), 260–268. https://doi.org/10.1016/j.dadm.2018.02.004

15. Humphreys, M. S., & Revelle, W. (1984). Personality, motivation, and performance—A theory of the relationship between individual-differences and information-processing. *Psychological Review, 91*, 153-184.
16. Hunt, E., & Wittmann, W. (2008). National intelligence and national prosperity. *Intelligence, 36,* 1–9.
17. Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior*, *29*(3), 340–362. https://doi.org/10.1016/0001-8791(86)90013-8
18. Isen, A. M., & Daubman, K. A. (1984). The influence of affect on categorization. *Journal of Personality and Social Psychology, 47,* 1206–1217.
19. Isen, A. M., Niedenthal, P. M., & Cantor, N. (1992). An influence of positive affect on social categori- zation. *Motivation and Emotion, 16,* 65–78.
20. Jensen, A. (1998). The G factor: The science of mental ability. *Choice Reviews Online*, *36*(04). https://doi.org/10.5860/choice.36-2443
21. Jin, Q., Li, C., Chen, S., & Wu, H. (2015). Speech emotion recognition with acoustic and lexical features. *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. https://doi.org/10.1109/icassp.2015.7178872
22. Jongseo Sohn, Nam Soo Kim, & Wonyong Sung. (1999). A statistical model-based voice activity detection. *IEEE Signal Processing Letters, 6*(1), 1–3. https://doi.org/10.1109/97.736233

23. Lau, N. T., Wilkey, E. D., Soltanlou, M., Lagacé Cusiac, R., Peters, L., Tremblay, P., ... & Ansari, D. (2022). Numeracy and COVID-19: examining interrelationships between numeracy, health numeracy and behaviour. *Royal Society open science*, *9*(3), 201303.

24. Lun, V. M.-C., Fischer, R., & Ward, C. (2010). Exploring cultural differences in critical thinking: Is it about my thinking style or the language I speak? *Learning and Individual Differences*, *20*(6), 604–616. https://doi.org/10.1016/j.lindif.2010.07.001

25. Lusardi, A., and O. S. Mitchell. 2011. Financial literacy and planning: Implications for retirement wellbeing. In *Financial literacy: Implications for retirement security and the financial marketplace*, eds. Olivia S. Mitchell and Annamaria Lusardi. Oxford, UK: Oxford University Press, 17–39.

26. Mahajan, P., & Baths, V. (2021). Acoustic and language based deep learning approaches for alzheimer's dementia detection from spontaneous speech. *Frontiers in Aging Neuroscience*, *13*. https://doi.org/10.3389/fnagi.2021.623607

27. National Research Council. 1990. *On the Shoulders of Giants: New Approaches to Numeracy*. Washington, DC: The National Academies Press. https://doi.org/10.17226/1532

28. Oechssler, J., Roider, A., & Schmitz, P. W. (2009). Cognitive abilities and behavioral biases. *Journal of Economic Behavior & Organization*, *72*(1), 147–152. https://doi.org/10.1016/j.jebo.2009.04.018

29. Polack, F. P., Al., E., for the C4591001 Clinical Trial Group*, Author AffiliationsFromAffiliations From Fundacion INFANT (F.P.P.) and iTrials-Hospital Militar Central (G.P.M.), Longo, E. J. R. and D. L., Others, M.-F. Y. and, Others, T. L. C. and, Others, V. B. and, T. L. Cowger and Others, Others, S. C. and, & Y. Feng and F. Wang. (2020, December 31). *Safety and efficacy of the BNT162B2 mrna covid-19 vaccine: Nejm*. New England Journal of Medicine. Retrieved November 28, 2022, from https://www.nejm.org/doi/full/10.1056/nejmoa2034577#article_letters

30. 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*(6), 943– 973.

31. 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*(6), 943.

32. Sadeghian, R., Schaffer, J. D., & Zahorian, S. A. (2021). Towards an automatic speech-based diagnostic test for alzheimer's disease. *Frontiers in Computer Science*, *3*. https://doi.org/10.3389/fcomp.2021.624594

33. Sakurai, M., & Kosaka, T. (2021). Emotion recognition combining acoustic and linguistic features based on speech recognition results. *2021 IEEE 10th Global Conference on Consumer Electronics (GCCE)*. https://doi.org/10.1109/gcce53005.2021.9621810

34. Shane Darke (1988) Anxiety and working memory capacity, Cognition and Emotion, 2:2, 145-154, DOI: 10.1080/02699938808408071Darke, S. (1988a). Anxiety and working memory capacity. *Cognition and Emotion*, *2*, 145–154.

35. *Subjective numeracy scale*. CBSSM. (2020). https://cbssm.med.umich.edu/how-we-can-help/tools-and-resources/subjective-numeracy-scale

36. upklyak. (n.d.). *Classroom for physics learning with formula and graph on chalkboard*. Freepik. upklyak. Retrieved 2022, from https://www.freepik.com/free-vector/classroom-physics-learning-with-formula-graph-chalkboard_15241048.htm#query=cartoon%20math%20board&position=2&from_view=search&track=sph#position=2&query=cartoon%20math%20board.

37. Upklyak. *Free vector: Mathematics class in school with teacher and boy*. Freepik. Retrieved December 5, 2022, from https://www.freepik.com/free-vector/mathematics-class-school-with-teacher-boy_30895733.htm#page=2&query=math%20cartoon&position=45&from_view=search&track=sph

38. US Department of Homeland Security. (n.d.). *What is the Sat?* Homeland Security. https://studyinthestates.dhs.gov/2012/07/what-sat