How we talk about math: Leveraging naturalistic datasets to define the discourse of math in contrast to other domains

Rachel Jansen
Department of Psychology
University of California, Berkeley
Berkeley, CA 94720 USA
racheljansen@berkeley.edu

Ruthe Foushee
Department of Psychology
University of California, Berkeley
Berkeley, CA 94720 USA
foushee@berkeley.edu

ABSTRACT

How do people talk about math? What point are we making when we contrast math with other topics? In studies of school performance, attitudes, and stereotypical beliefs, math is most frequently compared to language abilities and occasionally artistic qualities. Most studies about these topics administer assessments and closed-form surveys to make sense of how math ability or beliefs are different from similar constructs in other educational domains. In an analysis of Google search terms using Google Trends, "math" occurs in search queries far more frequently than "language" or "art" and —unlike searches about the other topics—the prevalence of "math"-related searches shifts in conjunction with the academic year. This project's goals are to (1) sample from diverse naturalistic text-based datasets to expose how math is referred to in non-experimental settings and (2) identify similarities and differences between math and the domains most frequently used as contrasts. We perform computational analyses on text derived from naturalistic sources written across a variety of different registers, from a journalistic source (NY Times) and a social media website (Twitter) to referential sources containing basic definitions (Merriam-Webster) and more informal descriptions (Urban Dictionary). We see that, across data sources, queries related to "math" refer more frequently to education-related themes and incorporate more disparaging terminology compared to content related to "language" or "art." This project is a first step in demonstrating that this methodology can aid in exploring more realistic discourse surrounding math and domains of comparison. This can inform and empower future researchers and practitioners interested in changing the discussion around math.

Keywords

math, language, art, naturalistic data, big data, NLP

Rachel Jansen and Ruthe Foushee "How we talk about math: Leveraging naturalistic datasets to define the discourse of math in contrast to other domains" In: *Proceedings of The 13th International Conference on Educational Data Mining (EDM 2020)*, Anna N. Rafferty, Jacob Whitehill, Violetta Cavalli-Sforza, and Cristobal Romero (eds.) 2020, pp. 624 - 628

Math is frequently discussed in a derogatory way: "Everybody hates math" is a major trope in popular media, with hundreds of instances across television sitcoms, comics, and movies.[1] In addition, when researchers investigate attitudes about math, they tend to focus their efforts on the study of math *anxiety* [2, 6], as opposed to positive feelings. These ways of portraying math, both in the media and in research and education appear unique to math. When researchers measure stereotypical beliefs [4] or attitudes about another subject [8], they are typically included as a contrast to math.

In public discourse, this differential treatment of math compared to other domains is perpetuated by *neuromyths*, or false ideas about the brain, such as the idea that "some of us are 'left-brained' and some are 'right-brained' and this helps explains differences in how we learn" [12]. Implying a natural contrast between math ability and art or language ability does a disservice to students, current and former: it encourages the belief that if we are "good" at one thing, we cannot be "good" at another. Beliefs about innate brilliance further amplify the folk distinction between math and art. Math is perceived as requiring the most brilliance out of any STEM discipline, and significantly more than all art and language-related fields, including English Literature, Art History, Linguistics, and Music Composition [10]. Such essentialist beliefs about domain-specific ability are bolstered by parallel sex disparities, with more "brilliant" fields like math including significantly fewer women.

In this project, we capitalized on naturally occurring data where people discuss math, and compare parallel discourse about other domains. We specifically analyze communications related to language and art, as these are frequently compared to math. We detail our rationale for each comparison domain in the next section.

0.1 Comparison domains

In research contexts, language very frequently serves as a comparison domain for math. Math is compared to language—and most often reading and writing skills—in research on ability [7], stereotyped beliefs [4], theories of intelligence [8]. [14] contrasts STEM performance (math and science) with non-STEM (language, humanities, and social science) and

[1] https://tvtropes.org/pmwiki/pmwiki.php/Quotes/EverybodyHatesMathematics

finds no actual performance difference nor evidence for gender differences in variability in academic grades.

On the other hand, art is much less studied in relation to math, in part because it is far more difficult to design art assessments than math or language assessments, and judging art ability is perceived as subjective. Some work has contrasted creative ability in math and art [9] and explored gender differences in stereotypical beliefs across the two areas [16]. More frequently, when conducting research on a task that involves some artistic expression in research about math ability, the idea of "art" goes unmentioned, as with drawing [5] or any study of spatial ability [1].

In studies comparing math ability or perceptions to parallel constructs in another domain, no justification is given for the choice of alternative domain. It may seem obvious to us that reading ability is the most direct contrast to math ability, but this is precisely what we are interested in accessing in this project. The assumptions about concepts and distinctions among them present themselves in our communications. Therefore, investigating naturally occurring data may provide us with the justification we need for domain comparison choices across different contexts. We assert that though art is only sporadically studied in relation to math, there are many reasons why it may serve researchers as an appropriate foil. For example, math is a required course throughout schooling and necessary for attending college, while under budget shortages, art classes are the first to be cut from curricula. However, mathematicians regularly enjoy drawing comparisons between artistic and mathematical abilities [11, 3]. This perceived distinction in the relative utility of math and art, paired with experts' regular likening of the two suggests an interesting avenue of future work.

0.2 Measuring math talk

Human attitudes are typically explored via closed-form surveys. But sampling bias as well as the wording of the questions can impact responses. We propose using naturally occurring datasets to supplement existing research about math attitudes and as a guide for developing new theories and experimental paradigms [15].

The goals of this paper are 1) to source data from non-experimental contexts to examine naturalistic discourse surrounding math and its comparison domains and 2) to identify how math is discussed that may be distinct from related domains in similar contexts. We hope to make an empirical case for comparing math to specific domains: why and when do we measure math against language or art, and what might be the appropriate choice based on how people represent these domains? We locate several sources of communication spanning a range of genres (e.g., journalistic, social media, and references), and registers (from more formal to informal writing styles).

1. METHODS

We identified a variety of online sources with freely accessible APIs (Application Programming Interfaces). We first used the Google Trends and English Lexicon Project as measures of frequency of term usage. Next we collected a selection of articles from the New York Times, tweets from Twitter, and definitions from the Merriam-Webster dictionary and Urban Dictionary that related to the search terms "math," "language," and "art."

1.1 Data sources

Though Google does not provide access to search history data, the company built an online interface, Google Trends,[2] for observing both fluctuations of searches for specific keywords or topics over time and across locations [17]. We focused our observations exclusively on the US. As another overview of frequency of specific terms, we used the English Lexicon Project (ELP).[3] These sources provide a very general sense of how these topics are thought about differently. We next explore actual word usage in multiple other sources, namely the New York Times, Twitter, and two different online references (Merriam-Webster and Urban Dictionary). Two of these may be seen as relatively objective (Times and Merriam-Webster), though a computational analysis of word usage will show whether this is truly the case.

We used the "Article Search API" from the New York Times (NYT)[4] to collect all hits that include the terms "math," "language," and "art." The NYT API provides the headline, keywords, date, word count, and lead paragraph for all articles that come up for a specific search term. The NYT Article Search API yielded 441,773 searches for "math", 367,707 for "language" and 1,276,036 for "art." We sample approximately 2,000 results for each search term.

Twitter is the only social media company that offers easy access to their data, in part because posts are all expected to be public anyway. [5] In order to align results with the data we obtained from the NY Times, we used the twitter package for Python[6] to load 2,000 tweets per search term.

Merriam-Webster additionally offers easy access to their definitions. [7] Preliminary data mining returned just the definitions for each term of interest, but it is meaningful that "math" has three definitions, while "language" and "art" each have ten. From Urban Dictionary, we downloaded all existing results for each term, which was 856 for math, 262 for language, and 876 for art (see Table 1 for total documents used in analyses for each term and each data source).

1.2 Text Analyses

With each data source, we created Naïve Bayes classifiers to contrast word usage for documents about math, language, and art.[8] Prior to text analyses, we ran a series of standard text pre-processing techniques: a) removing stopwords b) removing punctuation and c) reducing words to their roots (stemming and lemmatizing). To test the accuracy of each classifier, we shuffle the data and separate it into a training set consisting of 80% of the data and a test set comprising the remaining 20%. We train the classifier on the training

[2]https://trends.google.com/trends/?geo=US
[3]https://elexicon.wustl.edu/
[4]https://developer.nytimes.com/
[5]https://developer.twitter.com/
[6]https://pypi.org/project/twitter/
[7]https://dictionaryapi.com/
[8]We employ the NaiveBayesClassifier function from Python's Natural Language Toolkit (`nltk` version 3.2.2) package https://www.nltk.org/

set, then report the classifier's accuracy predicting responses on the test set, alongside a subset of informative features (words that are more common for one specific subgroup).

2. RESULTS

2.1 Google and ELP

From Fig 1, created from data generated in Google Trends to compare searches pertaining to the topics "math," "language," and "art,"[9] it is clear that, compared to the other topics, math is generally searched for at a higher rate, though takes steep dives in the summer months when school is no longer in session. This suggests that the term "math" is much more associated with education than other topics, and idea we explore in more detail in the other data sources. Contrary to Google search results, an analysis of term frequency in the English Lexicon Project reveals "math" to be significantly lower frequency (18,404) than terms relating to "language" (97,874) or "art" (62,513). According to this source, "math," "language" and "art" are estimated to be acquired at similar ages (5.56, 6.79, and 6.21 years, respectively), but "language" is rated as notably more abstract (that is, less concrete: language: 2.35, math: 3.15, art: 4.17).

	Math	Language	Art	Total
NY Times	1,541	1,946	1,835	5,322
Twitter	2,000	2,000	2,000	6,000
M. Webster	3	10	10	23
Urban Dict.	856	262	876	1,994

Table 1: Number of documents for each corpus.

2.2 Journalistic source

We excluded a set of "math" searches to ensure that the results would not be overly skewed. Specifically, 270 hits contained a daily math challenge and the lead paragraph began with "Test your math skills with today's question" and an additional 40 started with "Our weekly math problems are written by teachers at Math for America." For "art," we excluded 157 whose lead paragraphs began with "Our guide to new art shows and some that will be closing soon." There did not appear to be anything so consistent for searches relate to "language." Search results with blank lead paragraphs (159 math; 18 art; 64 language) were excluded from our training data. First, we analyzed the distribution of keywords. Of the 1541 math queries, 323 contained "school" (21%) compared to 81 of 1946 language queries (4%) and 14 of the 1835 art queries (0.8%). There was a similar pattern for the keyword "test," included in 149 math queries, compared to 6 and 1 for language and art, respectively.

We next looked at the text from each lead paragraph. We used all three sets of nonempty lead paragraphs for each topic to construct the classifier which included a total of 5,322 texts (1541 math; 1946 language; 1835 art). After removing all terms with roots "math," "language," or "art," the classifier achieved 70% accuracy on the test set. The most informative features for math included "test," "score," "grader," "improv," "educ," "competit," and "teacher" (e.g., "Growing up, I thought math class was something to be endured, not enjoyed. I disliked memorizing formulas and taking tests, all for the dull goal of getting a good grade"). The most predictive terms for an art-related hit contained "galleri," sculptur," "paint," and "noteworthi." For language-related queries, "speak," "translat," "dictionari," and "writer" were most informative. Though there are no apparent emotive terms, math arises much more frequently in documents related to school than does art or language in this context.

2.3 Social media

The average word count for tweets corresponding to each term was approximately equal (18 for math, 19 for language, and 18 for art), likely due to platform word count restrictions. In our classifier (accuracy: 62%), informative features for tweets about math included words very similar to those from the NYT, such as "test," "fail," "wrong," "class," and "science." For language, we saw "tiktok," "english," "speak," "video," and "utter." The set of most informative features for art contained "draw," "style," "anim," "design," and "cute." Here, we see many domain-specific similarities to the NYT data, but with the addition of terms conveying negative emotions related to math, such as "wrong," "fail," and "hard," which might speak to the greater subjectivity of the text source. The language- and art-related searches also appear to encompass more popular culture references. There was an interesting pattern of math being more ubiquitous in tweets relating to current events such as the election: 'berni,' 'vote,' and 'warren' (e.g., "Math says that Warren has a path") and the coronavirus outbreak: "million," (e.g., "It is simple math. The flu infects millions a year").

2.4 Reference materials

The Merriam-Webster dictionary produced few results, but the primary definitions themselves serve as a baseline set of relevant objective terms. Mathematics is defined as "the science of numbers and their operations," language as "the words, their pronunciation, and the methods of combining them used and understood by a community," and art as "skill acquired by experience, study, or observation." By their very definitions, math is a science (rather than an art) and art is not said to require innate ability.

The Urban Dictionary API yielded 856 definitions of "math," 262 of "language," and 876 of "art." The mean length of the math definitions was 36 words, 49 for language, and 58 for art (similar to the NYT results). Because this corpus was not evenly distributed across topics, we ran separate classifiers between each pair of topics, rather than over all 1,994 total definitions. For the math/art classifier ($n = 1732$), accuracy on the test set was 79% and the most informative words for math definitions were almost all negative: "abus," "number," "mental," "stress," "tortur," and "bore." For art on the other hand, informative features included "style," "emot," "draw," "amaz," "visual," and "color." The classifier comparing math to language yielded an accuracy of 85%[10] and primarily identified words that were informative of the language texts, as they represented a smaller proportion of our dataset. These included "speak," "talk," and "special," while terms more indicative of a math entry were "abuse," "mental," "human," and "bore."

[9]https://trends.google.com/trends/explore?date=all&geo=US &q=%2Fm%2F04rjg,%2Fm%2F04g7d,%2Fm%2F0jjw

[10]Chance would be 77% because 77% of definitions are math ones.

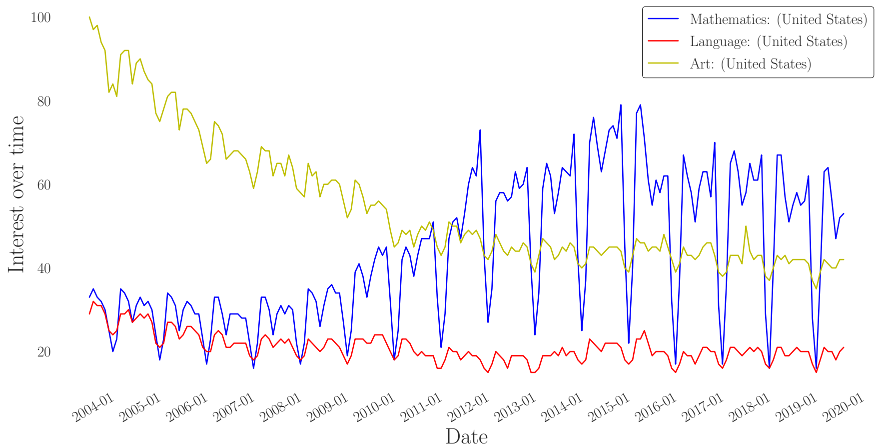

Figure 1: Google trends patterns of searching for the topics "mathematics," "language," and "art."

Finally, the classifier comparing art and language arrived at an accuracy of 85% and words from art-specific definitions included "music," "best," and "style," while definitions of language included terms "sign," "french," "wrong," and "number," which potentially likens math to language more than to art. Analyses of the example sentences included for each definition from Urban Dictionary produced comparable results, with notably more derogatory and profane terminology used to describe math than the comparison domains.

3. DISCUSSION

In a preliminary analysis of naturally occurring data sources, we have observed that math is more frequently written about and discussed in relation to education compared to language and art. In contexts where valenced language use is common (Twitter and Urban Dictionary), math is discussed using notably more unflattering terminology. Each data source we explored yielded different frequencies at which the three chosen topics were mentioned. Our Google Trends analysis revealed that math is searched for more frequently than language or art. However, though the NYT provided fewer "language" articles than "math" ones, there were more than double the number of hits related to "art" (owing to "arts and leisure" having its own section in the newspaper). In the references, "math" had fewer entries in Merriam-Webster compared to language and to art, but in Urban Dictionary, there were a comparable number of entries for "math" and "art," and this was more than triple the number of "language" entries.[11] Based solely on these simple search counts, we can identify important differences in how these topics are thought about: "math" appears to be defined more narrowly than the other domains (based on Merriam-Webster definition counts and shorter text lengths in the NYT and Urban Dictionary data) while emotions surrounding "math" and "art" are stronger than for "language" (based on the relative number of Urban Dictionary results). "Math" is also much more associated with education, a claim supported by the keywords from the NYT, our classifiers' informative features from the NYT, Twitter, and Urban Dictionary data, and from the cyclical nature of Google searches for "math."

[11] We were not able to acquire total hit numbers from Twitter.

4. FUTURE DIRECTIONS

This set of analyses only scratches the surface of what is possible with this methodology. We have many plans for further research, namely to conduct additional analyses on the data presented here, gather more data through clouds of novel search terms, and explore other naturally occurring data sources. First, to expand the findings from the data already gathered, we will conduct text-based sentiment analyses to search for systematic differences in overall valence associated with each term, and perform topic modeling over each set of documents. Second, "language" and "art" are only two of many possible domains to compare to math, alternatives to which we will pursue in future work. We aim to use the work done so far to refine our terms to determine what comparisons are more useful for different contexts. Finally, we intend to search deeper and with more specific intentions within the sources we have scrutinized thus far as well as among other potential sources of data.

5. CONCLUSION

Using large-scale datasets of naturally occurring text, this work presents a preliminary exploration of how math is discussed, compared to its most frequent comparison domains. Our data confirm that math is generally spoken about in a manner that is both more limited (e.g., to educational contexts) and more negatively valenced. Previous work has shown that familiarity with an idea increases belief in that idea [18], which means that the restricted and unflattering ways in which we generally discuss math may progressively degrade public opinion about the topic. If—through the media and other sources—speakers continue to hear (or read) about math as a narrowly defined concept associated with negative emotions, this perception will continue to thrive, and be unwittingly transmitted to future generations [13]. Thus, this work also serves as a plea to limit unnecessary disparaging reference to math in mass communication.

6. REFERENCES

[1] D. Barner, G. Alvarez, J. Sullivan, N. Brooks, M. Srinivasan, and M. C. Frank. Learning mathematics in a visuospatial format: A randomized, controlled trial of mental abacus instruction. *Child Development*, 87(4):1146–1158, 2016.

[2] E. Carey, F. Hill, A. Devine, and D. Szücs. The chicken or the egg? The direction of the relationship between mathematics anxiety and mathematics performance. *Frontiers in Psychology*, 6:1987, 2016.

[3] G. J. Chaitin. *Conversations with a Mathematician: Math, Art, Science and the Limits of Reason*. Springer Science & Business Media, 2002.

[4] E. K. Chestnut and E. M. Markman. "Girls are as good as boys at math" implies that boys are probably better: A study of expressions of gender equality. *Cognitive science*, 42(7):2229–2249, 2018.

[5] J. E. Fan. Drawing to learn: How producing graphical representations enhances scientific thinking. *Translational Issues in Psychological Science*, 1(2):170, 2015.

[6] A. E. Foley, J. B. Herts, F. Borgonovi, S. Guerriero, S. C. Levine, and S. L. Beilock. The math anxiety-performance link: A global phenomenon. *Current Directions in Psychological Science*, 26(1):52–58, 2017.

[7] L. Guiso, F. Monte, P. Sapienza, and L. Zingales. Culture, gender, and math. *Science*, 320(5880):1164–1165, 2008.

[8] E. A. Gunderson, N. Hamdan, N. S. Sorhagen, and A. P. D'Esterre. Who needs innate ability to succeed in math and literacy? Academic-domain-specific theories of intelligence about peers versus adults. *Developmental psychology*, 53(6):1188, 2017.

[9] K.-N. Jeon, S. M. Moon, and B. French. Differential effects of divergent thinking, domain knowledge, and interest on creative performance in art and math. *Creativity Research Journal*, 23(1):60–71, 2011.

[10] S. J. Leslie, A. Cimpian, M. Meyer, and E. Freeland. Expectations of brilliance underlie gender distributions across academic disciplines. *Science*, 347(6219):262–265, 2015.

[11] P. Lockhart. *A mathematician's lament: How school cheats us out of our most fascinating and imaginative art form*. Bellevue literary press, 2009.

[12] K. Macdonald, L. Germine, A. Anderson, J. Christodoulou, and L. M. McGrath. Dispelling the myth: Training in education or neuroscience decreases but does not eliminate beliefs in neuromyths. *Frontiers in psychology*, 8:1314, 2017.

[13] E. A. Maloney, G. Ramirez, E. A. Gunderson, S. C. Levine, and S. L. Beilock. Intergenerational effects of parents' math anxiety on children's math achievement and anxiety. *Psychological Science*, 26(9):1480—-1488, 2015.

[14] R. E. O'Dea, M. Lagisz, M. D. Jennions, and S. Nakagawa. Gender differences in individual variation in academic grades fail to fit expected patterns for stem. *Nature communications*, 9(1):1–8, 2018.

[15] A. Paxton and T. L. Griffiths. Finding the traces of behavioral and cognitive processes in big data and naturally occurring datasets. *Behavior Research Methods*, 49(5):1630–1638, 2017.

[16] J. R. Steele and N. Ambady. "Math is hard!" The effect of gender priming on women's attitudes. *Journal of Experimental Social Psychology*, 42(4):428–436, 2006.

[17] S. Stephens-Davidowitz and H. Varian. A hands-on guide to google data. 2014.

[18] W.-C. Wang, N. M. Brashier, E. A. Wing, E. J. Marsh, and R. Cabeza. Knowledge supports memory retrieval through familiarity, not recollection. *Neuropsychologia*, 113:14–21, 2018.