Exploring Undergraduate Mathematics Students' Math Identities Using Sentiment Analysis: Affordances and Precautions

Victoria Delaney
San Diego State University

Megan Selbach-Allen
Harvard University

Students' identities, orientations, and experiences in mathematics courses shape their participation and longevity in the discipline. Yet, undergraduate mathematics instructors do not often integrate pedagogy or curricula that help them build knowledge of their students' math identities. We believe that in many math courses this is a problem of time and scale. We propose that language processing methods can help alleviate this tension by giving instructors a baseline understanding of their students' math identities. In this study, we applied sentiment analysis to 28 mathematics "origin stories" written by undergraduate pre-calculus students. We compared the predictions from four language models with qualitative codes applied by two mathematics education researchers. We conclude this paper by discussing the strengths and limitations of sentiment analysis in our data, and discuss implications for undergraduate mathematics researchers and practitioners who may wish to apply these methods to their work.

Keywords: math identity, sentiment analysis, mixed methods, math instruction

Research has demonstrated the importance of *math identity*, how students see themselves in relationship to mathematics, as doers of mathematics, and their ability to persist in mathematics courses (Boaler, 2002; Cribbs et al., 2016). When mathematics instructors attend to students' math identities with culturally sustaining pedagogy, scaffolding curricula, or creating supportive course infrastructure, they can help students persist by diminishing issues of power and authority historically present in mathematics. However, an important first step in teaching responsively to students' math identities is building knowledge of students. While small class sizes make this a tractable task in K-12, organizational arrangements in undergraduate education pose difficulties for many teachers who wish to know their students' math identities. Undergraduate mathematics courses often enroll hundreds of students, and many instructors serve in a lecturer role. Even by incorporating pedagogical best practices that scale well to lectures, building knowledge of students' mathematics identities remains an intractable challenge. Furthermore, from a research perspective, undergraduate students' math identities are difficult to study at scale with qualitative methods that require humans to apply and agree upon qualitative codes.

To address these challenges, we explore the potential of using *sentiment analysis*, a technique from natural language processing, using community college mathematics students' math origin stories. Though limited to affect, sentiment analysis can give teachers a first glimpse at their students' positive/negative valences with respect to mathematics. We then investigate the limitations of sentiment analysis by comparing predictions from four language models to qualitative affective codes applied by the researchers. Our goal is to explore the extent, if at all, sentiment analysis can aid mathematics instructors and researchers in building knowledge of students' math identities (particularly in large math courses). We ask:

1. How accurately can sentiment analysis predict students' affect in math origin stories?
2. What limits exist for studying students' math identities with sentiment analysis?

Theoretical Framework

Math identity refers to “the ways that students think about themselves in relation to mathematics” including “a person’s beliefs, attitudes, emotions, and dispositions about mathematics and [their] resulting motivation and approach to learning and using mathematics knowledge” (Froschl & Sprung, 2016, p. 1). Building and acting upon knowledge of undergraduate students’ math identities can impact students’ perceptions of their ability to succeed in mathematics. Incorporating math identity requires teachers to inquire about students’ math histories and interrogate students’ ongoing tensions in seeing themselves as agentic doers and persistors in mathematics. Research has presented several strategies that can aid teachers build knowledge of students’ math identities and histories. For instance, Nasir and colleagues advocate for making explicit the relationship between mathematics knowledge, students’ cultures, and identities (2008). Rubel and Chu (2012) found that utilizing students' math histories as a starting point for enacting instruction led to increased participation and improved problem-solving skills among underrepresented minority students. However, both of these studies were conducted in K-12. It is unclear how the strategies presented in these studies would be implemented in larger undergraduate mathematics courses.

Implementing pedagogy based on students' math identities is significantly more challenging in large undergraduate mathematics courses. Hornsby and Osman (2014) highlight that the sheer size of lecture halls and high student-to-teacher ratios often impede instructors’ abilities to know their students personally. This conclusion is echoed by Cooper and Robinson (2000), who note that large class sizes can hinder the development of a classroom community that is conducive for students to share their personal mathematical experiences. Furthermore, Jaworski et al. (2017) point out that the diverse backgrounds of students in large lectures make it challenging to address all individual math identities effectively. To mitigate these issues, some researchers have suggested using technology to collect data on students’ math histories (e.g., Cotner et al., 2013), or using active learning pedagogies during lecture, such as group problem-solving, that could give students opportunities to share aspects of their math identity with peers (Freeman et al., 2014). However, these strategies require time investment and pedagogical knowledge, which may not always be available in large, traditional collegiate mathematics departments.

Sentiment Analysis in Mathematics Education

Although limited to affect, sentiment analysis could give mathematics instructors and researchers initial signals about students’ math identities. *Sentiment analysis* is a supervised machine learning technique that has been used in social science research to study emotions and affect in large text-based data sets (Thelwall & Buckley, 2013). Within sentiment analysis, language models make inferences on input text by weighting linguistic features that are likely associated with affect (e.g., “great,” “fail,” or “!”) towards a positive or negative prediction. An input text assigned a “positive” prediction, for example, represents the model’s inference that features extracted from the input text align most strongly with patterns associated with positive sentiment, based on the data that the language model was trained upon. In the case of mathematics identity, sentiment analysis can be applied to predict the positive/negative valence in students’ writing samples about math identity, which could give suggestions about how students perceive their past and present experiences in mathematics courses.

To date, sentiment analysis has been applied in mathematics education research to analyze pieces of student writing *after* mathematics courses have been completed. For instance, Crossley and colleagues (2018) used sentiment analysis on students’ emails to an emotional support

"Genie" (intelligent agent) to study correlations between students' math identities and K-6 mathematics assessment scores. Other studies in higher education have used sentiment analysis to study students' dispositions in end-of-course reviews (Bringula et al., 2022; Hew et al., 2020). However, as large language models are now available to consumers, we believe that these tools may afford mathematics instructors and researchers opportunities to study students' math identities vis-à-vis writing artifacts produced *before* and throughout mathematics courses. Our goal is to study the affordances and limitations of this method with texts about math identities.

Context and Methods

The participants in this study are part of a larger research project in a community college undergraduate mathematics course sequence. The course sequence contains one semester of pre-calculus and one semester of calculus, taught by the same instructor. Students are STEM majors in a cohort pursuing 4-year degrees. Many come from diverse linguistic, cultural, and educational backgrounds and have had mixed experiences in mathematics prior to the study.

The second author requested that the instructor give students a "math origin story" assignment at the beginning of the semester. Here, we consider "math origin story" to be roughly equivalent to other math identity activities (e.g., "mathography," "math autobiography," and "math history"). 28 students submitted math origin stories. The prompt for this assignment was:

> "…Write a short paper (approximately 300 - 500 words) OR make a media recording (audio or video, 3-5 minutes) and talk about your experiences of mathematics [before this class]. Describe the ways that any important experiences you had made you feel about math and how they shaped your learning of math. You can draw from formal or informal experiences, in the home, school, or any other setting. Maybe you had conversations with a family member that was important to you? If you were fortunate, you may have had an amazing teacher or perhaps you were unfortunate enough to have had a negative learning experience. Be descriptive, this is not an academic paper, it is a chance for you to share your math history and for me to learn more about you and your relationship with math."

Data Analysis

To determine the usefulness of sentiment analysis from the perspective of college mathematics researchers and educators, we analyzed the data using mixed methods. First, both researchers individually coded students' math origin stories for positive, negative, and neutral sentiment at the level of the entire math origin story. Next, we met collectively to agree on sentiment codes and resolve outstanding disagreements (Cornish et al., 2016). Drawing from research on math identity, we placed importance on students' statements that described their relationship to mathematics, such as "I am excited by the idea of doing mathematics," or "Using mathematics, I see myself as a future engineer." We placed less emphasis, although still took into account, sentiments about students' experiences with individual mathematics subjects or teachers unless the description indicated a transformation about their relationship with mathematics (e.g., "After taking Ms. G's class, I no longer see myself as a mathematician."). Table 1 shows criteria that we used to assign positive, negative, and neutral sentiments with examples from the data.

Table 1. Qualitative sentiment codes applied by the researchers.

Sentiment	Description	Example from Data

Positive	The text indicates affirmative behavior and/or thoughts in relationship to math identity, such as self-affirmation, seeing oneself as a doer of math or persisting in math. When positive and negative sentiments exist, positive sentiment is stronger.	"...I did notice at an early age that math was something that came natural to me even though I didn't apply myself as much." (Student 6)
Neutral	The text does not indicate sentiment direction, or, is not about the subject's math identity.	"The next math class I took was algebra 2." (Student 3)
Negative	The text indicates disaffirming behavior and/or thoughts in relationship to math identity, such as poor experiences, not seeing oneself as a doer of math or not persisting in the discipline. In the case of positive and negative sentiment, the negative sentiment is stronger.	"There were very few instances in math where I felt that what I was currently learning was genuinely …something that would … help me." (Student 17)

Both researchers held at least six years of teaching and research experience in mathematics education that informed how we conceived of math identity in undergraduate mathematics. The first author taught statistics and calculus for seven years in high school, and the second author taught calculus for nine years in higher education. Both authors taught mathematics to populations of students who have indicated complex relationships with "doing math."

After we applied sentiment codes, the first author pre-processed the text in students' math origin stories to ensure students' stories were compatible with the language models (Haddi et al., 2013). All program-specific acronyms were converted to their full name to reduce the likelihood that the language model would incorrectly infer alternative linguistic meanings. Then the corpus of words was spell-checked, and some words were edited for clarity (e.g., "highschool" was converted to "high school"). Any non-math symbols were removed from the data.

The first author then compiled a corpus of data containing students' origin stories to be analyzed at the level of each story. Sentiment analysis was performed with four language models: TextBlob, a python language processing package chosen for its accessibility; BERT, selected as a former state-of-the-art model for language inference tasks; GPT-4, chosen because it is a current state-of-the-art large language model accessible to consumers for free, and GPT-4o (ver. Sept. '24), selected because it is a current state-of-the-art large language model available to consumers (via OpenAI's API) for a monthly subscription fee. We then analyzed the nature of the errors each model made with confusion matrices and examined patterns of misclassified sentiments. The first author then created a new data corpus with all math identities misclassified by the highest performing model, at the level of individual sentences. We then performed sentiment analysis again with the highest performing language model.

Results

We present the findings from our study by research question. First, we give aggregate performance by each language model; then we analyze the sentiment of misclassified examples to further investigate differences between machine and human-applied sentiments.

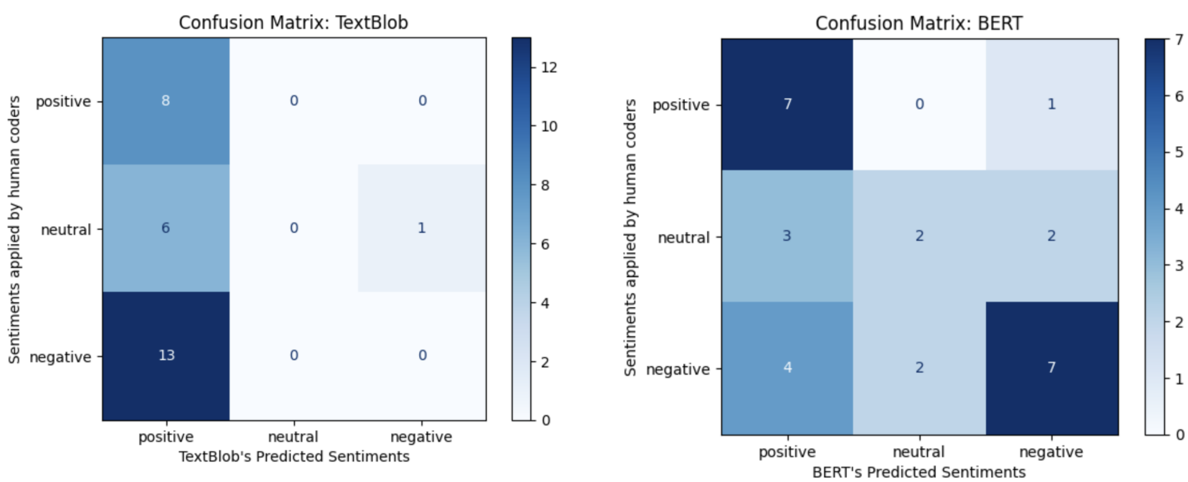
RQ1: How accurately can sentiment analysis predict students' affect in math origin stories?

Table 2 shows the performance accuracy of the four language models. While most models performed poorly (and Textblob performed worse than random guessing), GPT-4o's 82.19% accuracy is close to what is considered to be state-of-the-art performance (about 90%) among language interpretation tasks in English at the time of this writing.

Table 2. Sentiment analysis prediction accuracy.

Model	Python TextBlob	BERT	GPT-4	GPT-4o
% Accuracy	28.57%	57.14%	64.28%	82.14%

Figure 1 shows four confusion matrices that demonstrate how each model differed in sentiment predictions from the sentiment codes we applied as human researchers. Counts along the top-left, bottom-right diagonal indicate math stories where the model predicted the same sentiment as we did. All other areas of the confusion matrices indicate classification errors. For instance, TextBlob's low accuracy was almost entirely due to the fact that 27 out of 28 predictions were "positive," whereas we only labeled eight math origin stories as positive. On the other hand, GPT-4o's errors are due to predictions of neutral sentiment on positive math origin stories. We explore these misclassified results in detail in the second portion of the findings.



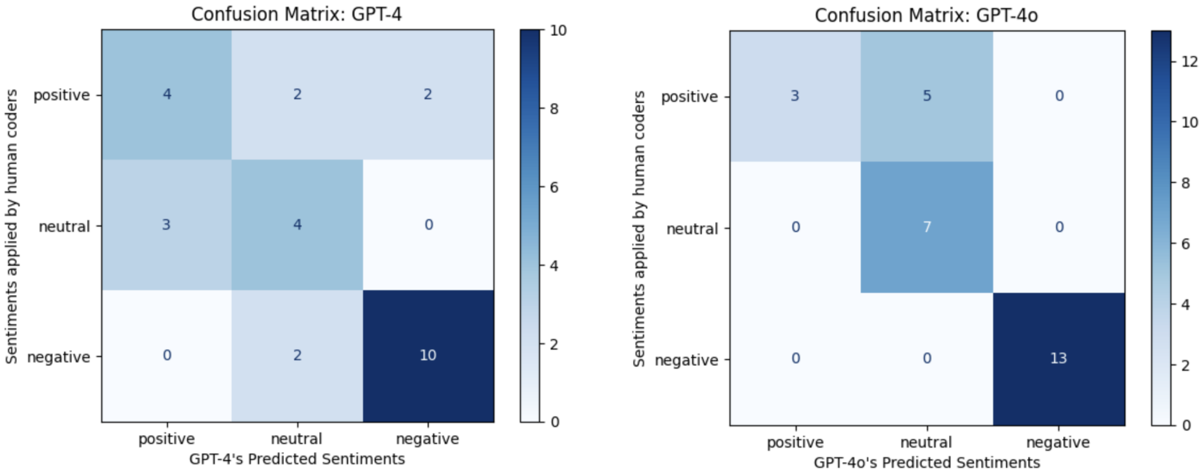


Figure 1. Confusion matrices for each language model.

RQ2: What limits exist for studying students' math identities with sentiment analysis?

We compared the sentiments of the misclassified math origin stories from GPT-4o by sentence to better understand why misclassifications occurred. Figure 2 shows the relative frequencies of positive, neutral, and negative sentiments among the 83 sentences in stories of five students whose positive sentiment was misclassified as "neutral." Results show that GPT-4o classified math origin stories as "neutral" if the ratio of positive to negative sentences was within ~20%. We examined misclassified data for common themes, and found three key distinctions.

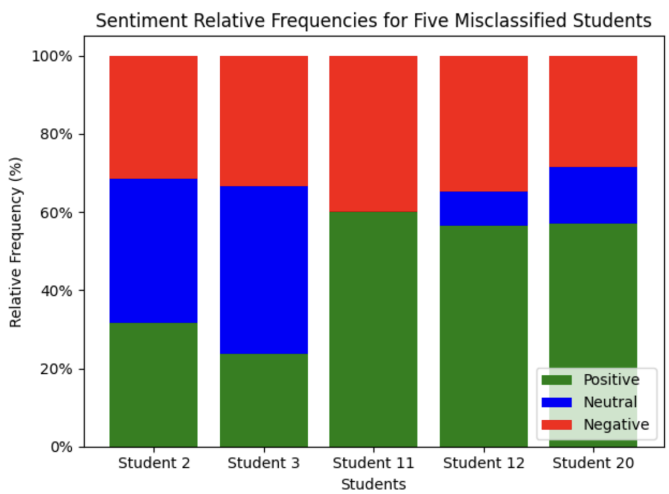


Figure 2. Relative frequencies of sentiments in five misclassified math origin stories.

Theme 1: Math identity in mixed sentiments. Sentences with higher word count frequently contained both positive and negative sentiment. In some cases, GPT-4o classified these sentiments as "neutral." For example, student 20 wrote: "As soon as the semester was over, I immediately dropped [math] class and became a teacher's assistant for my cooking teacher." We inferred that GPT-4o predicted "neutral" because "Dropped the class" is a negative sentiment and "became a teacher's assistant" is positive. However, as mathematics educators, we believed that dropping a mathematics course was more consequential for students' persistence in math, and therefore was more closely related to the student's math identity than becoming a teaching assistant for a cooking class. We therefore rated this sentence, and others like it, as "negative."

Theme 2: Attribution of sentiments. Unlike human raters, GPT-4o did not seem to weight sentences that referred to individuals other than the student author with less importance. For instance, student 12 wrote, "The most important person to my journey understanding math is my mom. She was always very good at math and knew most things." GPT-4o rated both sentences with positive sentiment. We agreed; however, we did not believe these sentences were as closely related to the students' own math identity, as they described the mother's ability to know and do mathematics. We rated these statements as "neutral" because they described another person.

Theme 3: Prioritization of sentiments that directly described students' math identity. All sentences seemed to be weighted equally by GPT-4o toward the overall sentiment prediction of the math origin stories, regardless of the subject-matter within the sentence. It appeared that the model tried to give an "average" sentiment prediction in neutral cases. For instance, student 3's story contained five positively-classified sentiments, nine neutrally-classified sentiments, and seven negatively-classified sentiments, and was given a "neutral" label. We show a portion of student 3's story below, *italicizing* negative sentences and **boldfacing** positive sentiments:

> "…*My algebra 2 class was a joke. My teacher retired and we did not have a teacher for the rest of the year. We changed substitute teachers and I learned nothing.* The next math I took was pre-calculus. *I would always skip class so I really didn't learn anything during the last 2 years of math.* During all my high school math I have had nothing crazy or anything important enough to remember. *My math classes shaped me to know the basic understanding of math but did not prepare me to thrive.* **Even though I had poor math experience it does not scare me or deter me from becoming a computer science major and I am excited to complete calculus and move on to future maths."**

Since Student 3 described their overall relationship with math as positive in the final sentence, and indicated that they wished to study more math (indicating persistence), we labeled their math origin story as positive, despite the student's description of negative past experiences.

Discussion

Performing sentiment analysis on a class of community college mathematics undergraduates gave us insights about the affordances and limitations of the methodology for both researchers and practitioners. Although GPT-4o underestimated the amount of positive math origin stories, it accurately predicted negative stories. This implies that state-of-the-art language models could potentially assist researchers and teachers to determine how many students express negative affect toward math early in the term, affording opportunities to provide instructional interventions proactively. Of course, sentiment analysis is not instructors' only potential tool for assessing the affective dimensions (McLeod, 1992) of students' math identities. Kaspersen and colleagues (2017) constructed a validated survey to measure math identity in STEM students. We note that the underlying data sources that support surveys (student questionnaires) and sentiment analysis (math origin stories), as well as students' relationships to those data sources, are structurally different from each other. We are curious to explore the measurability affordances of surveys versus sentiment analysis in math identity research in future work.

Furthermore, sentiment analysis by sentence demonstrates that instructional interventions can go beyond mathematics performance and address common themes in students' math identities such as challenging teachers or struggles to understand math concepts. Doing so could signal to students that their teacher is receptive to negative sentiments about mathematics, and will help students persist in mathematics regardless (Cribbs et al., 2016).

We explored sentiment analysis on students' math identities as a proof-of-concept methodology with a small sample intentionally. Our goal was to unpack the nuances of misclassified sentiments and make conjectures about how human mathematics researchers weigh linguistic features and assign labels in comparison with machines. In future work, we aim to explore this methodology at scale. We also wish to investigate how, if at all, knowing students' math sentiments at the beginning of the course impacts instructors' pedagogical decisions.

In the meantime, we recommend that instructors and researchers wishing to experiment with sentiment analysis and math identity begin their inspections of the data with neurally-labeled sentiments. Throughout this analysis, we learned that "neutral" can mask "mixed" or "nuanced" statements about math identity, which was the case for Student 3. Yet, a balance of "good" and "bad" math experiences does not imply neutrality in one's math identity. Humans apply their own weights to past experiences to determine their present relationship and future aspirations when seeing themselves as doers of mathematics. Our students deserve culturally-responsive instructors and researchers that understand the mixed nature of students' sentiment toward math.

References

Boaler, J. (2002). The development of disciplinary relationships: Knowledge, practice and identity in mathematics classrooms. *For the learning of mathematics*, *22*(1), 42-47.

Bringula, R., Ulfa, S. A. I. D. A., Miranda, J. P. P., & Atienza, F. A. L. (2022). Text mining analysis on students' expectations and anxieties towards data analytics course. *Cogent Engineering*, *9*(1), 2127469.

Cornish, F., Gillespie, A., & Zittoun, T. (2014). Collaborative analysis of qualitative data. *The SAGE handbook of qualitative data analysis*, 79-93.

Cotner, S., Loper, J., Walker, J. D., & Brooks, D. C. (2013). " It's not you, it's the room"—are the high-tech, active learning classrooms worth it?. *Journal of College Science Teaching*, *42*(6), 82-88.

Cribbs, J. D., Cass, C., Hazari, Z., Sadler, P. M., & Sonnert, G. (2016). Mathematics identity and student persistence in engineering. *The International journal of engineering education*, *32*(1), 163-171.

Crossley, S., Ocumpaugh, J., Labrum, M., Bradfield, F., Dascalu, M., & Baker, R. S. (2018). Modeling Math Identity and Math Success through Sentiment Analysis and Linguistic Features. *International Educational Data Mining Society*.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the national academy of sciences*, *111*(23), 8410-8415.

Froschl, M., & Sprung, B. (2016). Organization spotlight: Furthering girls' math identity: The key to girls' math success. *Childhood Education*, *92*(4), 320-323.

Haddi, E., Liu, X., & Shi, Y. (2013). The role of text pre-processing in sentiment analysis. *Procedia computer science*, *17*, 26-32.

Hornsby, D. J., & Osman, R. (2014). Massification in higher education: Large classes and student learning. *Higher education*, *67*, 711-719.

Hew, K. F., Hu, X., Qiao, C., & Tang, Y. (2020). What predicts student satisfaction with MOOCs: A gradient boosting trees supervised machine learning and sentiment analysis approach. *Computers & Education*, *145*, 103724.
Jaworski, B., Mali, A., & Petropoulou, G. (2017). Critical theorising from studies of undergraduate mathematics teaching for students' meaning making in mathematics. *International Journal of Research in Undergraduate Mathematics Education*, *3*, 168-197.
Kaspersen, E., Pepin, B., & Sikko, S. A. (2017). Measuring STEM students' mathematical identities. *Educational Studies in Mathematics*, *95*, 163-179.
McLeod, D. B. (1992). Research on affect in mathematics education: A reconceptualization. *Handbook of research on mathematics teaching and learning*, *1*, 575-596.
Nasir, N. I. S., Hand, V., & Taylor, E. V. (2008). Culture and mathematics in school: Boundaries between "cultural" and "domain" knowledge in the mathematics classroom and beyond. *Review of research in education*, *32*(1), 187-240.
Rubel, L. H., & Chu, H. (2012). Reinscribing urban: Teaching high school mathematics in low income, urban communities of color. *Journal of Mathematics Teacher Education*, *15*, 39-52.
Thelwall, M., & Buckley, K. (2013). Topic-based sentiment analysis for the social web: The role of mood and issue-related words. *Journal of the American Society for Information Science and Technology*, *64*(8), 1608-1617.