Beyond the Classroom: Exploring Mathematics Engagement in Online Communities with Natural Language Processing

by

Michael Ion

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Educational Studies)
in the University of Michigan
2024

Doctoral Committee:

Professor Deborah Ball, Chair
Professor David Jurgens
Professor Chris Quintana
Professor Ying Xu

Michael Ion
mikeion@umich.edu
ORCID iD: 0000-0001-5364-5556

© Michael Ion 2024

DEDICATION

To Saba.

ACKNOWLEDGEMENTS

To my advisor and chair, Deborah Loewenberg Ball: Working under your guidance has meant to me more than I can express. Your relentless support, advocacy, and wisdom, alongside the generous sharing of your time, have been fundamental to my completion of the program. I also extend my gratitude to my dissertation committee—David Jurgens, Chris Quintana, and Ying Xu—for their encouragement, generous support, and vital feedback. David, your course on natural language processing, alongside the insights from your research as an interdisciplinary researcher using linguistic analysis and network science together to understand behavior in its natural social context has motivated and transformed the work I do as a scholar. Chris, your continued support for my interdisciplinary work and guidance on where it can fit in various conferences and journals has been extremely helpful. Ying, your methodological expertise and advice on how to refine my larger set of research questions and findings into deliverables was immensely valuable and tools I hope to take away with me into my future work. To my committee as a whole, I really enjoyed working with all of you, and deeply appreciate the kindness, care, and detail to attention you all brought into our collaborative discussions.

I would not be where I am today as a scholar if it were not for the time I spent in the GRIP lab. To Pat Herbst, I appreciate the dedication and passion you have for the field of mathematics education and for the work you do to provide opportunities for researchers to apprentice into this work. To Amanda Brown, thank you for all the time you spent apprenticing me into the work of qualitative coding, collaborating on writing projects. To Inese Berzina-Pitcher, thank you so much for all of the support you provided to me and the rest of the researchers in the lab. To the rest of my graduate student and postdoctoral researcher colleagues—Nic Boileau, Inah Ko, Claudine Margolis, Emanuele Bardelli, Mollee Shultz, Soobin Jeon, Irma Stevens, Gil Schwartz, Carolyn Hetrick, Srikanth Lavu, Enes Akbuga—my time with all of you was so valuable, and I truly appreciate the time I got to spend with all of you.

Special acknowledgement goes to the members of the statistics and machine learning reading group—Emanuele Bardelli, Michael Frisby, Anna Paulson, Jason Godfrey, and Matt Truwit—for fostering a space to explore new quantitative methodological topics that we were unable to cover through our coursework. Emanuele, in particular, deserves an additional shout-out for being such an

amazing mentor to me throughout the years. Thank you for all of the time you provided me, as well as many of the other graduate students in our program as a mentor, through thinking through our dissertation projects and other methodological issues we were running up against. The University of Michigan community was incredibly lucky to have you around.

The same can be said for the many classmates, work colleagues, and acquaintances that I've connected with at Michigan who made my time in this program an experience that I will always cherish. Mauricio, Mar, Alex, Edie, Darrius, Bill, Natalie, Forrest, Nic and Liza—your friendships, guidance, and support meant and continue to mean the world to me. To the Michigan Melee and Tasukete communities, during the unprecedented times of COVID, these groups not only facilitated connection through our shared passion for this game we play, but also showcased the remarkable ability of online communities to connect people, foster growth, and provide support around a shared interest. Your inclusivity and warmth provided me with times of joy and laughter when I was unwinding from my academic work.

I would like to thank my family and friends back in California for your patience, love, and support as Saba and I finished this program. All of the care packages, the letters, the wedding invitations, and continuing to be included in your lives as we worked on these degrees from across the country meant so much. Lastly, to my partner Saba Gerami—your dedication, patience, love, kindness, and generosity as we battled through this program together was absolutely essential to my ability to finish this degree, and I am so proud and happy to see where we are today and where our path will take us next.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE

LIST OF TABLES

TABLE

LIST OF ACRONYMS

API Application Programming Interface

CoP Community of Practice

JSON Javascript Object Notation

LLM large language model

LMS Learning Management System

MDS Mathematics Discord Server

ML machine learning

MOOC Massive Open Online Course

NLP natural language processing

STT secondary-tertiary transition

ABSTRACT

In an era where digital platforms increasingly shape the educational experiences of learners, this dissertation examines activity in the MDS, an expansive online learning community used by hundreds of thousands of mathematics learners worldwide. Daily interactions, numbering in the tens of thousands, brought by students in need of advice, comprise a dynamic environment for peer mentorship. The study investigated the phenomenon of online mathematics learning taking place in chat-based platforms by creating and analyzing *MathConverse*, a novel dataset of 200,000 structured conversations from the help channels on the MDS. This dataset, transformed from raw messages into a comprehensive repository of conversations with rich metadata, makes possible ways of understanding the complexity of real-time problem solving and cooperative learning that takes place when students look for help from others online. Beginning with tackling the complexities of transforming chat-based exchanges into analyzable data, this dissertation navigates the challenges of conversation disentanglement and contributes to the methodological and theoretical advancement of educational research in online spaces.

Central to this investigation are two primary objectives: First, to demonstrate and refine the application of methods from machine learning and natural language processing to study *text as data* in educational research, addressing the methodological gap in analyzing voluminous, text-based datasets. Chapter 2 provides details of the work involved in transforming extensive conversational data into structured datasets for analysis. In Chapter 3 and Chapter 4, I provide case studies using *MathConverse* to illustrate how techniques from natural language processing (NLP) can be used to draw rich qualitative insights from the texts we as social science researchers are surrounded by in our research. For example, once I determined a large language model could reliably categorize questions into question types paragraph 3.5.2.1.2, I used the model to classify a larger set of questions ($n = 120,362$) by question type.

Second, the dissertation aims to provide an illustration of the dynamics of engagement and learning within online mathematics communities, particularly the MDS. The creation, analysis, and public distribution of the *MathConverse* dataset empowers researchers to explore learning phenomena often obscured from our view as researchers and educators in traditional academic settings. The analyses in the study not only probe the types of inquiries posed by learners and

the nature of their interactions but also provide an example of the various ways a mathematical concept can be instantiated in a conversation through my closer look at the diverse conceptions of the derivative that showed up across the sample of conversations.

CHAPTER 1

Introduction

1.1 Motivation

1.1.1 From traditional classrooms to digital learning: The evolution of mathematics learning spaces

In their transition from secondary to university mathematics, students find themselves at a critical juncture where the familiar structures of academic support recede, forcing many to navigate their mathematics coursework more independently (Vollstedt et al., 2014). The study of this shift, often referred to as the secondary-tertiary transition (STT), has been studied extensively by scholars in mathematics education. Gueudet's (2008) review of the scholarship in this area highlights three main transitions involving individual, social, and institutional factors: the transition in ways of thinking, the transition to proof and the technical language of mathematics, and the institutional transition related to changes in the didactical contract. Furthermore, Bettinger and Long (2009) emphasize that many students entering university are academically under-prepared for such rigor, necessitating additional remedial support. Although students still engage with traditional forms of academic support, such as office hours, one-on-one tutoring, and discussion sections, many encounter barriers such as inflexible scheduling, financial constraints, and a hesitation to interact directly with authority figures, which can limit their access to these resources (Pepin, 2014).

In recent years, the escalating role of technology in education has become increasingly evident, leading to a shift in teaching and learning methods (Escueta et al., 2017). As traditional classroom approaches encounter their limitations in addressing diverse learning styles and adapting to a rapidly changing world, the need for innovative, technology-driven learning models has become more pronounced (Hiebert and Grouws, 2007). This shift towards digital learning platforms not only offers new avenues for engagement and interaction but also aligns with the evolving educational demands of the 21st century. The advent of these platforms marks a significant evolution in the ways mathematics is being taught and learned, and as education researchers, we must continue to investigate the role these platforms are taking on outside our traditional classroom walls.

In the current era, our access to extensive and detailed educational text data is unprecedented, sourced from online learning platforms, as well as transcripts of in-person classroom interactions and student work. This rise in data availability is paralleled by technological progress and research breakthroughs in fields such as NLP, computational linguistics, and social science. These advancements have catalyzed the creation of sophisticated methods for the quantitative analysis of large-scale text corpora (Fesler et al., 2019; Grimmer et al., 2022). Collectively, the expansion of data accessibility and the refinement of analytical methodologies significantly enhance our ability to identify novel patterns and validate emerging theories within the realm of education. In the context of these developments and the transition towards digital learning environments, Chapter 2 offers a new perspective for understanding how students engage with their mathematics coursework outside the traditional classroom setting by illustrating the creation of *MathConverse*, a dataset curated from an online mathematics learning community. This work provides education researchers with a dataset of over 200,000 conversations between students and helpers and illustrates the some of the initial steps (and potential) of doing research on learning that takes place in online platforms.

1.1.2 Online mathematics networks: Their growth and role in education

Online platforms for asking mathematics questions, such as the MDS, have not only provided a space for academic support but have also become instrumental in developing resilience and adaptability among students (Borba et al., 2018). These spaces have been shown to offer diverse benefits, accommodating different learning styles and promoting inclusivity (Pratt and Back, 2009). The sense of belonging and collective knowledge that emerges from these interactions is a testament to the theories of connectivism, which suggest that knowledge is distributed across a network of connections (Goldie, 2016).

The growing interest in studying online learning spaces within education research is reflected in their increased use as a help-seeking resource for students. Notable investigations, such as the study by van de Sande (2008) explore the dynamics of participation, community building, and the development of mathematical understanding within these forums. For instance, Van de Sande's (2008) analysis of discourse within the calculus help forum FreeMathHelp.com demonstrates how students actively engage in authoring content, questioning and challenging mathematical propositions, and contributing to collective knowledge construction. With the rise of participation across various platforms like MathHelpForum.com, StackExchange.com, and MathOverflow, the significance of these platforms in immediately being able to provide students with help on their mathematics problems is undeniable. As educators, researchers, and designers of educational technology, we must understand the interactions in these online learning spaces. Such understanding can help tailor our pedagogical approaches and research endeavors more effectively to the evolving learning modalities of our students.

In contrast to traditional forum-based websites for learning mathematics, learners can now turn to chat-based platforms on platforms such as Discord, which can help catalyze real-time interactions within specialized communities. The MDS, central to this dissertation, represents a departure from asynchronous forums, offering immediate dialogue and feedback—a recognized catalyst for learning (Hattie and Jaeger, 1998). In the MDS, students can pose questions in a set of help channels where they are met with a flow of nearly instant feedback that can be invaluable for clarifying concepts and correcting misunderstandings in the moment. This immediacy not only fosters a sense of community but also significantly enhances the learning process by aligning with the educational principle that timely feedback is essential for effective learning (Roth et al., 2008). Moreover, the volume of data generated by chat-based interactions on platforms like Discord presents an unprecedented resource for researchers, providing a wealth of real-time, naturalistic data that captures the nuances of learning as it unfolds.

The expansive datasets available from platforms like Discord provide new opportunities for research on conceptions, as data from these platforms can be collected with unprecedented detail and scale. Given that students are gravitating towards these online networks to get help in their mathematics courses, understanding the nature of these interactions can be a useful area to build research studies. Knowledge from these studies can help us better understand what these resources are offering them that traditional venues of help (e.g., office hours, mathematics tutoring centers, supplemental workshops) might not be. The interactions present in the *MathConverse* dataset, central to the analysis in Chapter 3 provides a window into the ways in which users are interacting in the MDS. Unlike traditional, asynchronous forums, the MDS facilitates real-time dialogue and feedback, aligning with modern educational principles that prioritize immediate, contextual learning support. This immediacy not only fosters a dynamic learning community but also offers a unique perspective on student engagement and problem-solving strategies outside conventional classroom parameters. Through an extensive analysis of over 200,000 student-tutor conversations spanning millions of turns of talk, Chapter 3 looks at the varied ways students and peer tutors engage around mathematics content and within within the MDS platform, offering insights into the types of questions posed, the topics discussed, and the temporal dynamics of these interactions. By employing methodological techniques from machine learning and natural language processing (NLP), the case study presented in this chapter aims to provide an illustration of the dynamics that can exist in a *community of interest* (Henri and Pudelko, 2003) such as this one and what the implications might be for the teaching and learning for mathematics worldwide. The approach to use text-as-data methods on a large corpus on conversational texts not only provides a granular view of student and tutor behavior but also highlights the potential of digital platforms to supplement traditional educational models, offering educators and researchers a richer understanding of how students interact with mathematics in the digital age.

1.1.3 Analyzing conversations about calculus problems on digital platforms

The derivative, a foundational concept in calculus, stands as a symbol of the academic and social transitions that students undergo as they embark on higher education (Rasmussen et al., 2014; Strogatz, 2019). Online platforms like the MDS can help serve as an essential bridge during this transition, as students can get immediate access to help and feedback. The prevalence of calculus within these online communities offers a good portion of conversations for analysis, presenting an optimal case study for looking into the complex nature of mathematical conceptions as they unfold in real-time online interactions. While research in this area has traditionally explored the interplay between students' explicit knowledge of the derivative (as formalized by educational authorities like teachers and textbooks) and their implicit mental constructs (see Nurwahyu et al., 2020; Tall and Vinner, 1981), these investigations have often been limited to smaller, controlled environments. In contrast, this study extends the scope of inquiry to a much larger, more spontaneous context. By focusing on the derivative, the work in Chapter 4 not only probes a single mathematical concept but also sets a precedent for a broader examination of mathematical conceptions within online tutoring spaces. The goal of this chapter is not merely to dissect students' understanding of the derivative but also to offer a window into the varied and rich ways in which students conceptualize mathematics within a community of their peers.

Researchers in mathematics education have historically oscillated between exploring what students know (i.e., an epistemological perspective) and how they come to know it (i.e., a cognitive perspective) (Brousseau, 1997). *Conceptions*, or dynamic understandings of mathematical concepts, are shaped by interactions with learning environments (Balacheff, 2013). Methodologically, mathematics educators have used interviews, classroom observations, analysis of student work, and textbook evaluations, mathematics educators to explore student conceptions. Building upon Zandieh's (2000) comprehensive framework for the derivative, which categorizes the concept across different contexts and layers of mathematical reasoning, Chapter 4 builds on the the previous chapter by filtering out the conversations that pertain to the derivative to look at at the various conceptions of the derivative that emerge in the conversations, using Balacheff's (2009) definition. Studies focused on the varied ways students can think about the derivative have emphasized the necessity to examine the interplay between students' knowledge of the derivative stated in terms of definitions provided by their teachers and textbooks (concept definition; Tall and Vinner, 1981) and their mental constructs they have associated with it (concept image; Nurwahyu et al., 2020). In these studies, researchers attempt to approximate students' concept images by dissecting their explanations, actions, or sequences of the steps they take while in a problem-solving situation. Compared to the work done in this chapter, research examining students' understanding of the derivative through problem-solving has primarily been conducted in smaller-scale settings. My goal here is to try to see whether I can use text-as-data methods to see how conceptions of the derivative are emerging in this large online platform.

The derivative's prevalence in conversations on the MDS offers a useful case study to explore the multifaceted nature of mathematical understanding as it unfolds in real-time, peer-to-peer interactions. The goal here is to provide a guiding framework that can help provide scaffolding to other education researchers on how they might study student conceptions by working with concepts that are of interest to them, either through the use of *MathConverse* or with their own, 'unstructured' datasets. By analyzing real-time dialogue data from Discord, the goal of the analysis in Chapter 4 is to provide a look into how often the conversations in the MDS that bring up the concept of the derivative go beyond the notion of "applying rules to take a derivative". That is, to see whether and how many of these discussions have students thinking about the derivative in the multifaceted ways described in Zandieh's (2000) framework. This chapter uses recent advances in the training and implementation of large language models to showcase how to classify large amounts of conversations into discrete categories reliably. In the next section, I provide a brief primer on understanding the use of text-as-data for analysis, as the analysis of text data shows up throughout all three of the studies.

1.2 Text as Data

The participants in the MDS primarily communicate via messages that contain text and images. Text is often referred to as 'unstructured data', as it is a recording of a verbal activity meant to communicate something (Benoit, 2020). In order to be able to treat these messages as data that can be understood by a machine, I needed to extract the messages from the platform, establish a 'unit of measurement' for each analysis, and transform these texts into numerical representations. As shown in Figure 1.1, this process turns the 'unstructured', yet very meaningful, text exchanges that transpire in the MDS, and transform these texts into data in a way that is amenable for analysis at scale.

NLP is a subfield of machine learning and artificial intelligence encompassing a set of techniques that makes human text accessible to computers (Eisenstein, 2019; Hirschberg and Manning, 2015). Goals of the field of NLP include using computers to perform useful tasks such as sentiment analysis and question answering. We can see NLP in action through several commercial applications we use in our everyday lives: our emails providing suggestions for how to complete a sentence, or our

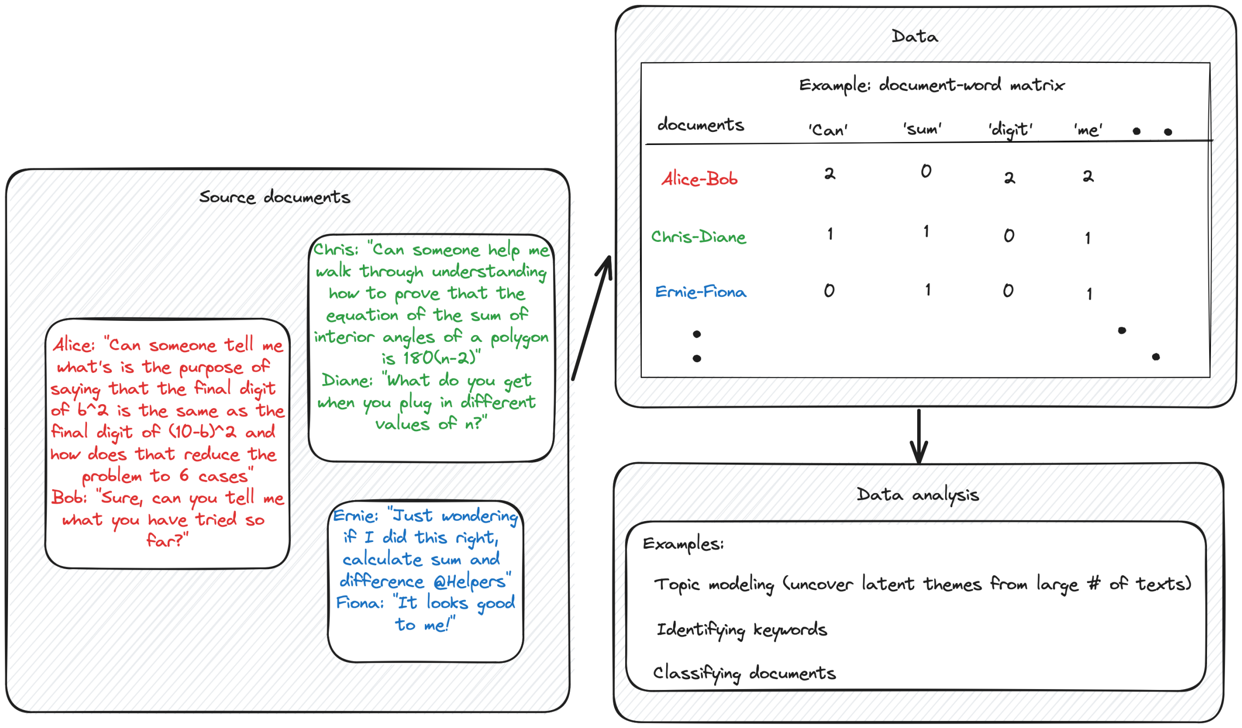


Figure 1.1: From text to data to data analysis; an adaptation of a figure from Benoit (2020)

.

ability to ask Google how to translate an English phrase to French. Additionally, applications of NLP extend to academic research settings. For example, political scientists have used NLP to infer political leanings of text based on their language choices (Yu et al., 2008), and social scientists used NLP to analyze police officer discourse to show racial disparities in officer respect (Voigt et al., 2017). In mathematics education, scholars have used topic modeling to analyze five decades of articles from two prominent education journals (*Journal for Research in Mathematics Education* and *Educational Studies in Mathematics*) to determine how the major areas of focus in mathematics education have shaped and shifted (Inglis and Foster, 2018). In this dissertation project, I leverage techniques from the field of NLP to construct representations of the data found in *MathConverse* in order to answer the research questions I pose in Chapter 3 and Chapter 4. The results from the findings of these chapters help demonstrate the benefits gained by augmenting our work with unsupervised and supervised machine learning methods. That is, combining qualitative coding with machine learning techniques can provide researchers with deep understanding of text-based interactions that are challenging to do without the use of computers.

1.2.1 Text classification and its relevance to this study

Among the diverse applications of NLP, text classification plays a useful role in the transformation of text into data for analysis. This common technique from machine learning automatically categorizes open-ended text into one (or many) predefined categories, such as in sentiment analysis to determine textual tone, language detection, and distinguishing between questions and statements. Text classification has been crucial in various sectors, notably in developing AI systems like chatbots, where accurately identifying user input as questions or statements is essential for effective response mechanisms. The investment in developing models for such tasks highlights the importance of text classification in both commercial and academic contexts. In the context of this dissertation, text classification serves as an essential analytic technique for addressing many of the research questions. For example, in Chapter 3, to explore what students and helpers discuss in the MDS, I use a pre-trained model to classify messages from *MathConverse* as either questions or statements, followed by further classification of questions by type using a large language model (LLM) provided through an Application Programming Interface (API) (RQ 2a of Chapter 3). Similarly, to identify different student conceptions of the derivative, a topic model is used to label conversations by topic, and then the conversations labeled as being about the derivative are then classified with another LLM to determine which (if any) conceptions of the derivative are present in the conversation (RQ 1 of Chapter 4). This dissertation study provides these two inquiries as case studies on how text classification with machine learning models can be used to enable the systematic analysis of educational interactions at scale, providing a contribution for understanding mathematical discourse within online platforms.

1.3 Research Objectives

One main goal of this dissertation project is to provide a window into a learning space that is used by hundreds of thousands of mathematics learners from around the world, one where there are tens of thousands of messages are exchanged everyday in its help channels where students can get guidance from peer mentors on their mathematics problems. My time as a doctoral student and researcher at the University of Michigan has provided me years of focused time to observe, learn from, as well as learn *how to learn from* learning environments like the MDS. In this study, I present a novel application of using text-as-data methods for analyzing large-scale educational interactions. Building upon the evolving landscape of students engaging with mathematics online and the significant potential of online learning communities to augment traditional educational supports, this dissertation aims to explore the nuanced ways in which students engage with mathematical concepts in these new environments. At the heart of this investigation are two primary objectives.

The first is demonstrating and refining the application of text-as-data methods and large language models (LLMs) in educational research, which I posit can help strengthen the methodological gap in analyzing large-scale, text-based datasets within the field of mathematics education. This includes detailing the process of transforming vast amounts of conversational data into structured datasets amenable to analysis and leveraging techniques from NLP to uncover patterns of engagement and conceptual understanding. The second objective is highlighting the type of engagement and learning that is happening in online mathematics communities, such as the Discord Server (MDS). My work in creating, analyzing, and making publicly available the *MathConverse* dataset can provide researchers a way in which to view this learning that is happening outside of the eye of many academics. This work involves a detailed analysis of the types of questions that are being raised in the community, the nature of interaction, and the topics of discussion, including a deeper dive into understanding the conceptions of the derivative that are emerging in the conversations.

1.3.1 Research questions

The objectives of the dissertation outlined above directly inform the following overarching research questions:

1. What processes are involved in converting large amounts of mathematical conversations into a structured dataset for analysis using text-as-data methods?
2. What are the characteristics of participant engagement and conversational dynamics within the MDS?
3. What conceptions of the derivative emerge in the data collected from the MDS?

These questions guide the work of each of the three chapters, respectively, and in Chapter 5, I check back in on the outcomes of these as they relate to the results of each study.

CHAPTER 2

MathConverse: The Design of a Large Scale, Multi-Subject Mathematical Dialogue Dataset

2.1 Introduction

This chapter introduces *MathConverse*, a dataset derived and constructed from a set of channels on an online mathematics platform dedicated to helping learners with mathematics problems. By capturing the conversations that take place in this online space into discrete, analyzable forms of data, it becomes easier to capture some of the intricate dynamics of student-tutor interactions in this unique educational setting that are often unseen by educators in K-12 and university classrooms. In the current digital era, the reliance on traditional classroom data, such as recordings and transcripts, has been pivotal for understanding and enhancing educational methods and student interactions (Major and Watson, 2018). However, as Suresh et al. (2022) point out, such data sources often face significant challenges related to scalability, privacy, and practicality. These factors severely limit their availability and the feasibility of sharing data, which can inhibit the replication and advancement of research in mathematics education. In stark contrast, *MathConverse* represents a novel approach to overcoming these obstacles. By leveraging data from a public online learning environment, *MathConverse* presents a scalable and diverse alternative that effectively navigates through many of the limitations inherent in traditional educational data sources. This significant advancement equips researchers with new avenues for exploring the ways students are engaging with the mathematics they are working on outside their formal classroom settings.

Online learning spaces have rapidly become integral to contemporary educational practices, offering an unprecedented avenue for students to engage, discuss, and explore mathematical concepts (Engelbrecht et al., 2020). These spaces, characterized by their accessibility and collaborative nature, represent a paradigm shift in how educational interactions occur outside the traditional classroom. The MDS, a focal point of this study, epitomizes this shift. It serves as an academic hub, facilitating rich dialogues and problem-solving exercises among diverse participants ranging from novices to experts.

MathConverse holds significance in understanding these novel educational interactions. By capturing conversations from the space in the MDS devoted to providing individual help to students on homework problems, this dataset provides a window into the real-time, dynamic process of learning and providing help with mathematical content. The work described in this chapter differs from conventional data collection methods, as it can help provide a large-scale, yet deep look into how students engage with mathematics problems outside of face-to-face learning environments. This dataset can not only help bridge the gap between theoretical understanding and practical application of what is going on when students leave our classrooms, but can also aid in helping identify patterns, challenges, and opportunities in online mathematics education.

The chapter begins by looking at some examples of online discussion groups that have evolved in digital communication platforms (e.g., FreeMathHelp, MathOverflow, Discord, Slack, WhatsApp) as well as how mathematics educators have studied mathematics teaching and learning in these settings. Next, I discuss prior work on constructing datasets in educational research, with a specific focus on datasets in mathematics education. Following this, I discuss the literature on *conversation disentanglement* highlighting the challenges and innovations involved in extracting coherent dialogues from synchronous chat environments. The chapter progresses to describe the composition of this study's dataset and its methodological underpinnings, followed by a discussion on the broader implications and potential applications of this dataset. I conclude by reflecting on the potential contribution of *MathConverse* to the ongoing discourse in online educational research, setting the stage for future explorations and implications.

2.2 Background

2.2.1 Learning in online communities

The digital era has significantly changed the way people talk with one another, with digital communication platforms such as Discord, Slack, and WhatsApp emerging as spaces for learning and interaction. These platforms offer unique opportunities for forming learning communities beyond traditional classroom spaces, catering to diverse educational needs and backgrounds. Discord's role extends beyond traditional social media, serving as a voice, video, and text chat app with server-based organization similar to group chats, complete with explicit rules and hierarchical structures (Roy, 2023). The structure of interactions on Discord resonates with Swales' (2016) concept of *discourse communities*, where the platform's server-based organization, explicit rules, and hierarchical structures not only foster goal-oriented communication but also support the development of shared practices and values among its users, embodying the evolving nature of digital discourse communities.

2.2.1.1 Authority and anonymity in online learning communities

Within these digital communities, authority is multifaceted. In his study of the University of Toronto Mississauga Math Server, Roy (2023) observed that authority was not only vested in administrative roles, such as Teaching Assistants and Professors, and the enforcement of server rules, but also emerged through participatory credibility. The latter is particularly critical as, as this shows that active participation builds trust and reinforces authority in these communities, diverging from a definition of authority as an intangible power granted through an institution Wardle (2004). Kim et al. (2023) observed similar dynamics in an introductory organic chemistry course, where Discord facilitated student-tutor interactions, leading to the formation of an active student learning community. This finding calls attention to the platform's capability to cultivate interactive educational environments that support student learning and engagement.

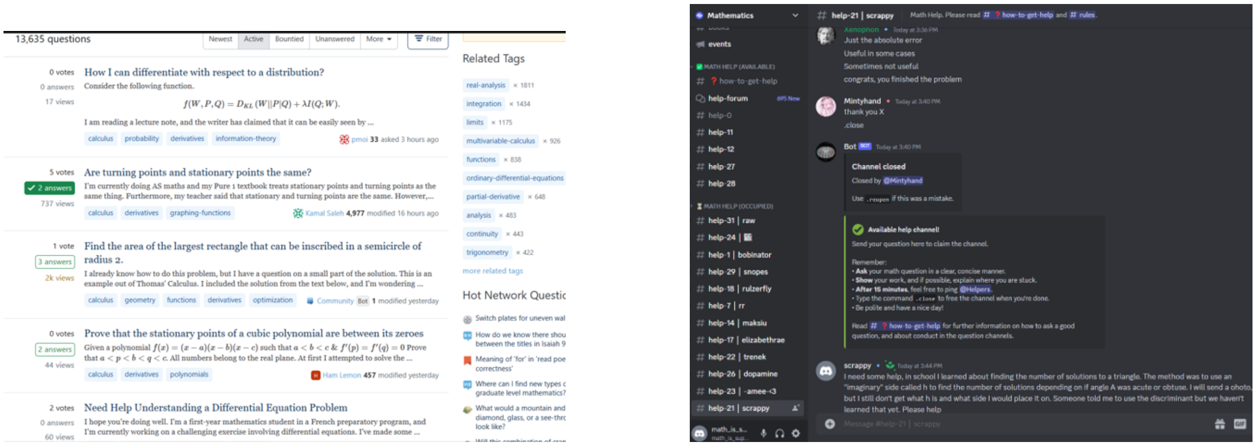
Effective interaction is key in these communities. In the UTM Math Server, the proper use of mathematical terminology and notation facilitates meaningful dialogue and integration into the community (Roy, 2023). This approach aligns with Duff's (2010) language socialization theory which emphasizes learning through interactions with more proficient members. Similarly, in the study by Kim et al. (2023), the importance of peer interaction in online environments is emphasized, showing how platforms like Discord can mitigate the challenges of traditional educational data sources by offering scalable, diverse alternatives. The structure of authority within Discord servers, such as those in the UTM Math Server, profoundly impacts the learning process. Roy (2023) identifies three types of authority: administrative roles like Teaching Assistants and Professors, the server administrator, and authority built through participatory methods. Each form of authority on the server is underpinned by online presence and trust, with the latter deriving entirely from participatory involvement. This emphasis on participatory authority illustrates the shift in traditional educational dynamics, where authority is not solely institutional but is also earned through active engagement and credibility within the community. Further extending the role of these platforms, Heinrich and Carvalho's (2022) research highlights how Discord supports professional identity formation. Unlike LinkedIn or Slack, Discord fosters informal communication, allowing students of different year levels, alumni, and staff to connect, engage in collaborative inquiry, and build confidence—a fundamental component of professional networking and identity formation.

The unique environment of online learning communities, as observed in platforms like Discord, foregrounds a significant shift in the traditional dynamics of authority and learning. In these digital spaces, students often experience a heightened sense of autonomy and comfort, which can encourage more open inquiry and participation. This autonomy is partly attributed to the anonymity that online platforms provide, alleviating the pressures and social anxieties often associated with face-to-face academic interactions (Jay et al., 2020). Such anonymity can be particularly empowering for students from traditionally underrepresented groups in STEM, who may face additional barriers in conventional classroom settings, like the fear of reinforcing stereotypes or threatening their self-image (Marchand and Skinner, 2007; Ryan et al., 2009).

Furthermore, the transformed perception of help-seeking in educational contexts, as highlighted by Nelson-Le Gall (1981) and others (e.g., Gonida et al., 2019; Schenke et al., 2015), aligns with this shift towards more autonomous and student-centered learning environments. Online platforms foster a normalization of seeking assistance, which is now recognized as a key factor associated with academic success. This is particularly evident in the context of mathematics education, where the pressures of grading and authority in face-to-face interactions with instructors may inhibit student engagement. The participatory nature of online communities therefore presents an empowering alternative, allowing students to engage more freely and confidently in their learning process. This transition to a more student-centered, interactive model of learning sets the stage for exploring the intricate dynamics of mathematical discourse and conceptualization within these digital communities, as detailed in the following sections of this chapter.

2.2.2 Forum-based vs. chat-based platforms

Online mathematics communities come in two forms: forum-based and chat-based. Forum-based platforms set the standard for structured knowledge exchange. Some examples of these forums, such as Mathematics StackExchange MathOverflow, and FreeMathHelp.com, offer users the ability to pose questions, provide answers, and engage in academic dialogue with a clear delineation of topics and responses. Their asynchronous nature affords users the time to craft thoughtful, in-depth contributions, thus fostering a more reflective and analytical exchange of ideas. However, this structure may also lead to slower response times, potentially impacting the immediacy of help and support that learners seek. Figure 2.1a illustrates the organized interface of Mathematics StackExchange, a prime example of such forum-based communities. Contrasting the forum-based model, chat-based communities like the MDS provide a fluid, conversational space for learners to interact. This synchronous mode of communication allows for a rapid exchange of ideas, resembling the immediacy of a live classroom discussion. Such platforms often attract students who prefer a more dynamic and instantaneous help system, which can enhance their understanding and retention of mathematical concepts. The less structured nature, as depicted in Figure 2.1b, encourages a more equitable participation across members, enabling a diverse range of voices and perspectives to contribute to the learning process.



(a) Forum-based help platform (b) Chat-based help platform

Figure 2.1: Online mathematics platforms for helping learners with problems

2.2.3 Prior work building datasets in educational research

Contemporary research and practice on the teaching and learning of mathematics in classrooms are deeply informed by data derived from recordings of classroom activities, which encompass video, audio, and transcript formats. Such data are not only pivotal for understanding current educational interactions as highlighted by Major and Watson (2018) and Kim et al. (2023), but they also play a vital role in the development and training of educational technologies that are emerging. The construction and utilization of high-quality datasets in educational research have become increasingly important, especially in mathematics education. These datasets enable the analysis of student-teacher interactions, which are critical for developing teaching strategies and educational technologies. One such study, the "Million Tutor Moves Observatories Project" represents a significant advance in digital educational data science. The goal of this project is to crowd-source one million quality student-tutor interactions that are machine-readable, in hopes of training intelligent tutoring systems tailored for diverse learning needs (Reich et al., 2023). Another study by Demszky et al. (2021) focuses on a framework for computationally measuring teachers' conversational uptake, which is when a teacher builds on a student's contribution in dialogue. They released a dataset of 2,246 student-teacher exchanges from US math classroom transcripts, annotated for uptake by domain experts. The uptake is formalized using pointwise Jensen-Shannon Divergence (PJSD), compared to unsupervised measures, and correlated with educational outcomes. This dataset and methodology facilitate large-scale analysis of teacher-student interactions and aim to improve educational practices through better understanding of conversational dynamics. Suresh et al. (2022) work on the TalkMoves dataset is another pivotal contribution. This dataset comprises 567 annotated transcripts from K-12 mathematics lessons, derived from both in-person and online classes. It provides a rich source for analyzing teacher and student discourse, facilitating the development of educational technologies that can improve classroom interactions.

The work developing open-source datasets for by scholars like the ones I have cited above sets a new precedent in educational data science research. These datasets, with their extensive, detailed annotations of mathematical dialogues, are invaluable for building more reliable models that will be put into use one day in our mathematics classrooms. In the next sections, I explore the challenges and innovations in dataset construction and the role of AI and bots in data collection and processing, shedding light on the complexities and advancements in educational data science.

2.2.4 Conversation disentanglement

When using data from a synchronous chat platform (e.g., Slack, Google Hangout, Internet Relay Chat (IRC), or Microsoft Teams) for research, it should be noted that the messages are not necessarily organized into threads like they are in other asynchronous chat platforms (e.g., Stack Exchange, Quora, or Yahoo Answers). The conversations are entangled, meaning all messages appear in one space. When users enter these spaces, although search functionality is typically available, it is up to them to determine where conversations begin and end. Most chat disentanglement models have been based on chats extracted from IRC channels (Kummerfeld et al., 2019; Shen et al., 2006), but until 2019 it was done with small datasets (less than 2500 messages) or non-open source datasets. Kummerfeld et al. (2019) manually annotated a dataset of 77,563 messages for conversation disentanglement and provided their model, which has served as a training set for many studies working to develop their own disentanglement models (see Chatterjee et al., 2020; Liu and Cohen, 2021). To train the model, they engineered features at the message and pairwise message level. Some message level features included timestamps and answers to questions such as: how long ago this user last wrote, did the same user wrote right before or after, are they a bot, is this message targeted, and was the previous message targeted. Additionally, some pairwise message features included number of intervening messages or answers to questions like: is the user the same, did the user address another user, and do the two messages have the same target? Their results were state of the art at the time, and since then, researchers have been working to build and improve on this work given the new strengths of technology that have come out in the past two years. Figure 2.2 presents an example from the MDS, where three separate conversations are entangled in the same thread.

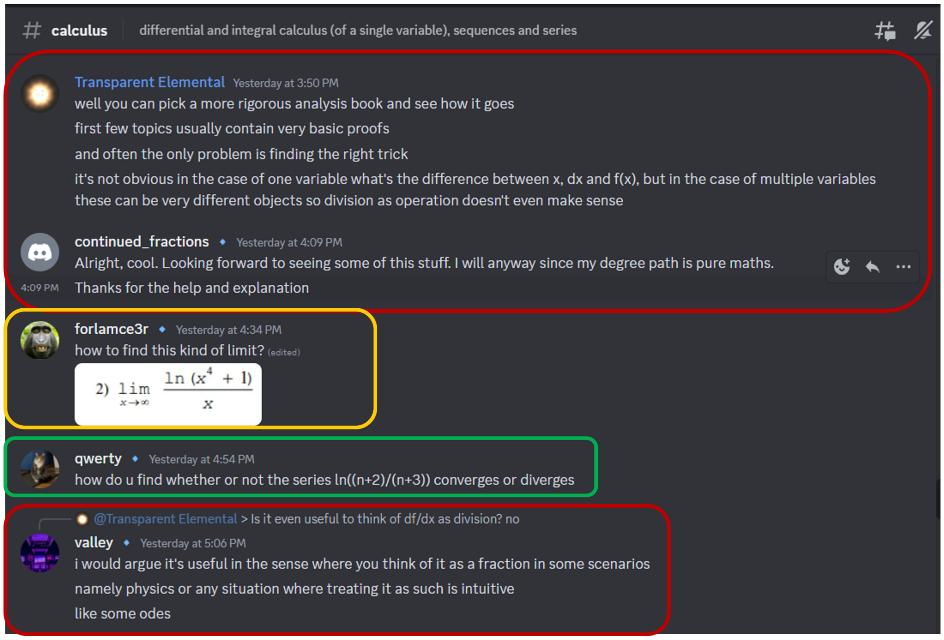


Figure 2.2: Entangled data with 3 conversations

2.3 Data

2.3.1 Site choice and description

The dataset for this dissertation was sourced from the Mathematics Discord Community, an active online platform facilitated by Discord. Discord's capabilities in supporting text, video, and voice communication have made it a popular choice for diverse groups, including those focused on education. The MDS, launched in January 2017, represents a unique convergence of individuals with a shared passion for mathematics (Figure 2.3). With over 38 million messages exchanged among more than 181,000 registered users (as of November 2023), the server provides a rich source of data for examining contemporary mathematical discourse and learning practices. This server was chosen due to its active engagement and substantial user base, which provide a diverse range of mathematical discussions and interactions. The broad yet focused nature of the community's rules has shaped the discourse, making it an ideal setting for this study. The University of Michigan Institutional Review Board (IRB) has granted an exemption from continuous review for this research, acknowledging its adherence to ethical research standards.

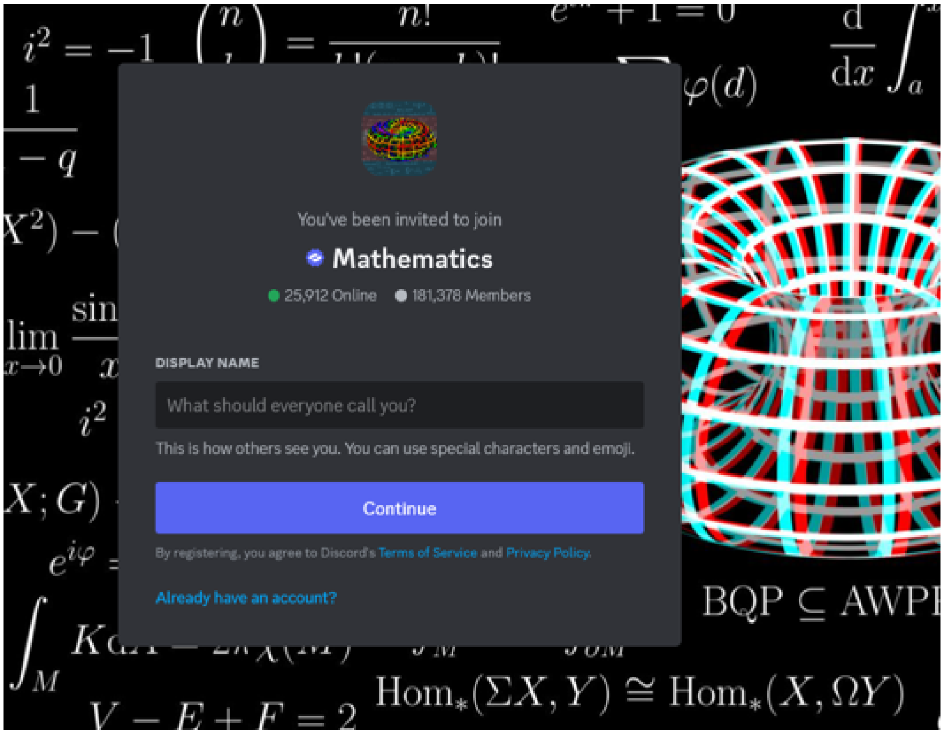


Figure 2.3: What users are shown upon joining the server, offering a glimpse into the community structure and entry process.

2.3.2 Community norms as an embodiment of an online community of interest

The MDS, with its structured yet dynamic environment, shares some elements of what Wenger-Trayner and Wenger-Trayner (2015) have succinctly encapsulated in their more recent writing of Communities of Practice (CoPs). They describe CoPs as groups bonded by a shared domain of interest, engaging in collective learning while learn how to do it better as they interact regularly. While the MDS provides a space for people who share an interest in mathematics, to engage in collaborative interaction with thousands of other like-minded people from all over the world, and share an abundance of resources for learning, I turn to the work of Henri and Pudelko (2003) in wanting to to note the MDS does not meet the requirements of being classified as a CoP. They identify three principal components of the social context of the activity of virtual communities: the emergence of intention (goal of the community); the methods of initial group creation and the temporal evolution of both the goals and the methods of group creation. I describe their theory in more detail in Chapter 3; however, for now, I say that a main component not met by the MDS is that a majority of their participants are going there for the goal of individual learning; that is, they go for help on their own mathematics problems. Using their framework, I can identify the MDS as something closer to a *community of interest*.

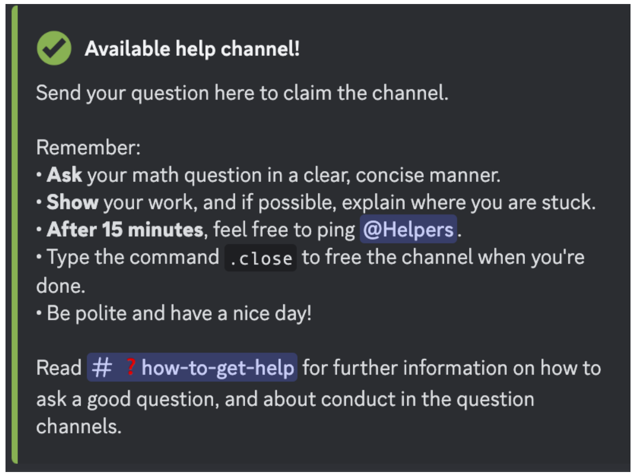


Figure 2.4: A screenshot showing the implementation of a bot that shows up in every encounter showing the norms of how students should ask for help.

All of this said, the server goes beyond being a mere repository for questions; the way that the server has been set up helps cultivate a dynamic environment conducive to collaborative problem-solving and the sharing of knowledge. This is particularly evident in the lively book recommendation channel, where members engage in diverse discussions that span topics from Riemann integration to machine learning, reflecting a community committed to mutual academic support and the collective creation of knowledge. Moderation plays a critical role in guiding these interactions. Moderators, wielding discretionary authority, are key to maintaining the server's scholarly tone and ensuring civility within discourse. Their influence is apparent in the quality of the interactions and, consequently, in the quality of data collected for this study. A notable moderation initiative was the establishment of 'help channels' in late 2019, which realigned the server's norms to direct question-asking into these channels, reserving topic-specific channels for broader discussions on subject matter. One of the rules established by the moderators that helpers are not supposed to directly give answers to those giving questions, and there is a bot that shows up in the available channels explicitly telling the 'question-askers' how to ask for help (Figure 2.4). Beyond this, moderators can recognize and assign 'roles' to users based on their contributions to the community, remove content deemed unsuitable, and enforce account suspensions to uphold the integrity of the platform and its users' experiences.

In summary, this server represents more than a digital space for mathematical discussion; it is a growing, evolving space that has weekly talks, community members that participate daily helping learners who ask for help, and a set of norms that provide guidance on how to ask and provide help to learners. Research on this space can help provide a window into the digital evolution of how students are looking for help when they leave our classroom halls, which sets of questions they find most difficult, amongst a large set of questions we might find interesting to explore. In order to do this work, it must first be made into a dataset that can be analyzed at scale with appropriate metadata.

2.3.3 Characteristics about the data

Figure 2.5 illustrates a conversation among four community members and a bot (an 'automated' user) about a problem related to the derivative. This figure highlights six key features (numbered in the list below) that provide structure to the cleaning, representation, and modeling of the conceptions.

1. Multi-modality: The conversation initiates with a student providing a problem through a screenshot and adding clarification by typing in some clarification, demonstrating the need to use multiple modes of communication—symbolic and graphical—to convey their question effectively in the absence of face-to-face interaction with the person helping them.

2. Timestamps: Each message in the dataset includes a date and time record that adheres to the ISO 8601 international standard, marking the exact moment it was posted. This precise time-stamping is essential for maintaining uniformity in the temporal data, which is fundamental for analyzing the sequence and timing of interactions. It provides a framework for examining the rhythm and pace of the community's collaborative problem-solving efforts, revealing patterns in engagement and responsiveness over time.

3. Linked replies: Here we notice that users can reply directly to a message rather than just continuing to reply in the channel. This is relevant metadata when there are more than two participants present if someone is directing a message specifically towards someone. If a person says, "do you understand this", and it is a linked reply, the metadata can tell us who they are talking to.

4. Multiple Participants: This particular conversation involves several peers contributing to the problem-solving process, with one particularly active member providing most of the guidance. The data call attention to the collective aim and implementation to learning in this community, with multiple individuals partaking in a single thread of discussion.

5. Conversation Closure: The use of a ".close" command illustrates a systematic way in which conversations are concluded, with a bot formalizing the end of an interaction. This feature, significant to the structure of the server, offers a clear endpoint to conversations, was one of the key factor in the ability to disentangle the conversations.

6. Automated Bots: The presence of bots highlights the role of automation in managing discussions and providing support, such as typesetting with LaTeX or moderating conversation flow. As noted above, the bots can help ensure that the conversations come to a close and

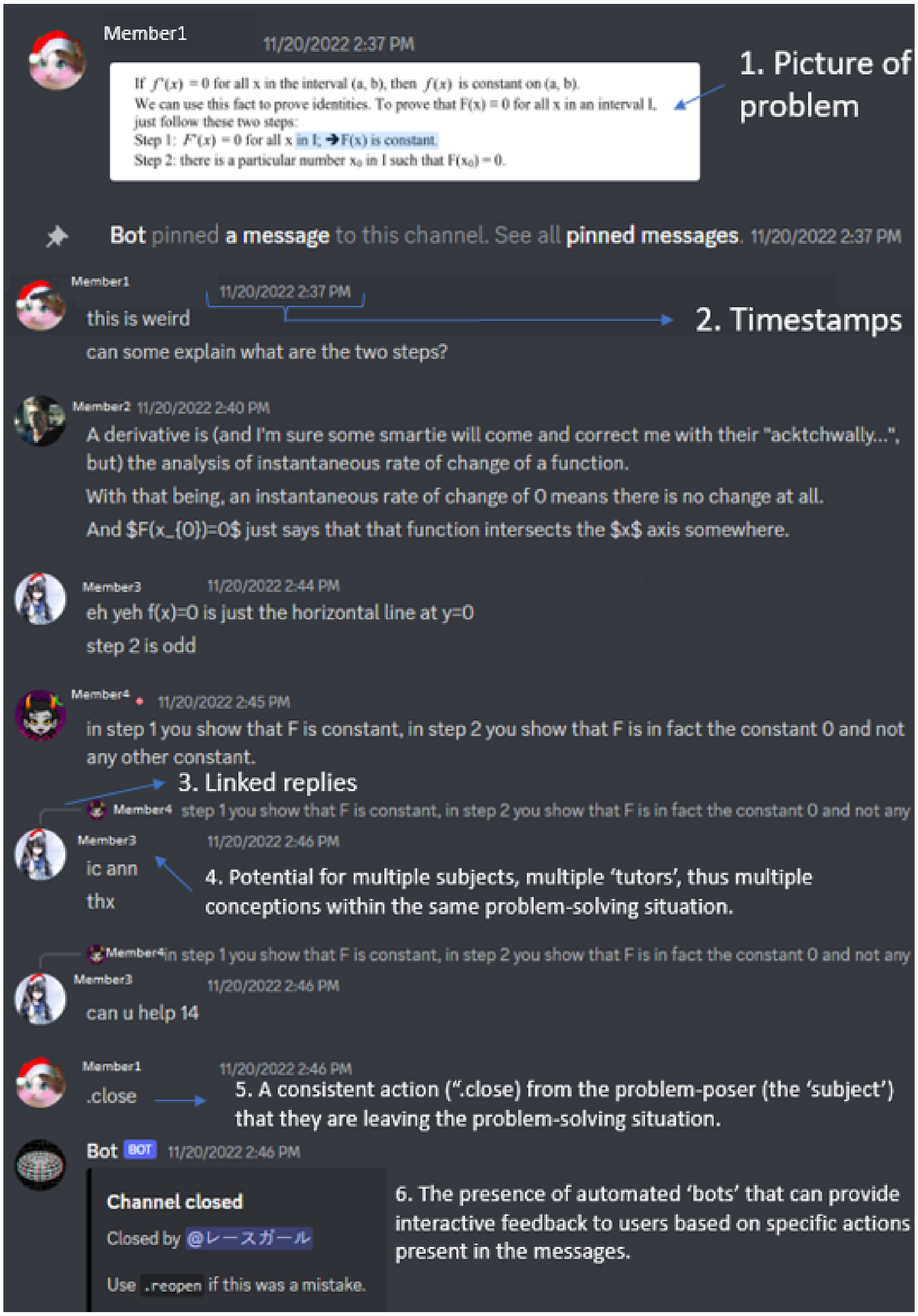


Figure 2.5: Example conversation from the dataset with some highlighted features. Usernames have been deidentified.

make the channels ‘available’ again for other learners.

These features collectively paint a picture of the important features present in the conversation and are reflective of some of the larger pedagogical and social dynamics present within the MDS. In the next section, I provide an overview of how I approach embracing some of these features into my methodological approach to the study.

2.4 Methodological Approach

The field of mathematics education is increasingly harnessing digital platforms to understand how students engage with and learn mathematical concepts. Platforms like Discord have emerged in order to support students looking for help in topics like mathematics, physics, and programming, and as a social science researcher interested in text-as-data methods, I see immense potential of being able to study student learning behaviors and interactions from these data. This section outlines the methodological approach undertaken of this study, which focuses on extracting and disentangling the data from the MDS into separate conversations for in-depth analysis.

The methodological approach employed in this study is significant for a number of reasons. It allows for exploration of authentic student interactions in a naturalistic setting, providing a window into how mathematical concepts are discussed, questioned, and talked about outside of the classroom. Additionally, the methodology is instrumental in managing the vast and complex nature of data intrinsic to online communication platforms. Online chat rooms, especially those used for the sake of learning, produce large volumes of inherently unstructured and varied data, including text, images, and interactive exchanges. Addressing this challenge requires a systematic approach to organize, categorize, and analyze the data in ways that render them meaningful for research. By transforming the raw, unstructured conversational threads into structured datasets, the approach I describe in this chapter is to describe a way to create a way to uncover patterns, themes, and insights that might otherwise remain hidden. This includes analyzing how mathematical concepts emerge while discussing homework problems, identifying the types of questions that foster deeper engagement, and observing patterns of peer-to-peer interaction that contribute to learning.

2.4.1 Tools and technologies employed

2.4.1.1 Data collection using Discord Chat Exporter

The initial phase of data collection was facilitated by the Discord Chat Exporter command line interface, an application that helps extract information from Discord channels. This application helped export the messages of each of the help channels into JSON format, encompassing around six

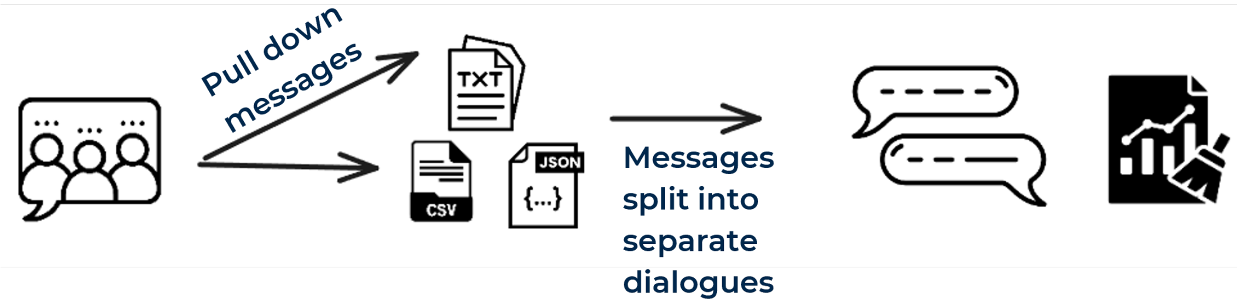


Figure 2.6: Overarching design of Chapter 2 study

million messages, which form the foundations of building the dataset for this study. Key components of the data included message IDs, timestamps, user information, and the content of the messages themselves.

Following data collection, the challenge of conversation disentanglement was addressed. This involved developing the algorithm which serves as the main finding of this chapter. This algorithm is encapsulated within a Python script, and serves to parse and organize the raw data into coherent, individual conversations.

2.4.2 Conversation disentanglement

As cited above in subsection 2.2.4, the development of an algorithm to disentangle the dataset was pursued to address the complex challenge of parsing and organizing a substantial corpus of educational data, specifically housed within a number of data formats that can handle the hierarchical, nested structure of the conversations. The primary objective was to structure the data in a way such that it could be clear to me and researchers in the future: (1) when the conversation started, (2) who started the conversation, (3) timestamps to identify when messages were sent and the duration of the message, alongside a number of other metadata that could be helpful for analyses by researchers in future projects. This process of figuring out the unit of analysis early on was important, as methodologically I want to make sure that the methods and models I am employing in the study can handle understanding the representation of the data I am creating. I aim for the study within this chapter to be both practical, in that the dataset can help facilitate mathematics education researchers to study how students engage with mathematics problems online in their own studies, but also in a theoretical or methodological sense, I aim to help my readers gain a sense of the challenges I have faced in constructing such a dataset, and what might be necessary in building a learning environment if one might wish to study the data at scale at future dates.

In the data that has been extracted from the MDS, a *message* is defined as a discrete communication act—a student posing a question, a response from a peer, or a clarification from a mentor. These messages are the atomic units of communication, often rich with mathematical notation, diagrams, and discourse. Yet, the true essence of the educational dialogue is best captured when these messages are viewed in concert, forming what I refer to as conversations. A *conversation* in this context is an encapsulated exchange that begins from when a student introduces a problem and ends when the problem finds closure, whether through an answer, an explanation, or a conceptual understanding. Unlike messages, which are singular and often fragmented, conversations provide a comprehensive view, encompassing the ebb and flow of the interaction. The dataset reveals that these conversations are not confined to a singular domain of mathematics; rather, they traverse a spectrum ranging from high school subjects such as precalculus to graduate school subjects such as measure theory. This diversity reflects the server's role as a microcosm of the broader academic sphere, where students arrive with inquiries drawn from various mathematics courses. Each conversation, therefore, is a window into the unique challenges and learning trajectories within different mathematical domains.

Focusing on conversations as the primary unit of analysis unlocks the potential to observe and understand the dynamics of problem-solving and how students as they unfold in real time. This approach allows for an examination of the scaffolding that occurs as participants collaborate to navigate challenge mathematics problems. By dissecting these conversations, researchers can gain insights into not only the cognitive processes involved in mathematical problem-solving but also the social and collaborative dimensions of learning mathematics online. Therefore, the distinction between messages and conversations is an important one to make here—while messages are the building blocks, it is through the aggregation of these messages into conversations that researchers have a chance to capture what mathematics learners are grappling with when they pose questions online and engage with knowledgeable others, and what comes of these conversations.

Figure 2.7 provides a diagram of some of the over-arching pieces that go into disentangling the data. The algorithm starts by transforming the messages that happen in each channel into Javascript Object Notation (JSON) file. This step is critical for transforming the raw data into a structured format that can be effectively manipulated. Following this, the algorithm enters a loop, iterating through each message within the files. A key decision point in this loop is the identification of 'Channel closed' messages, which serve as markers to delineate the boundaries of individual conversations. When such a message is encountered, the algorithm increments a conversation identifier (conversation_id), which marks the start of a new conversation thread. This mechanism ensures that each conversation is captured as a discrete entity, reflecting the natural flow of dialogue as it occurred. To adhere to privacy concerns, the algorithm incorporates a function to replace the user names with pseudonyms. While the usernames already mostly protect the users

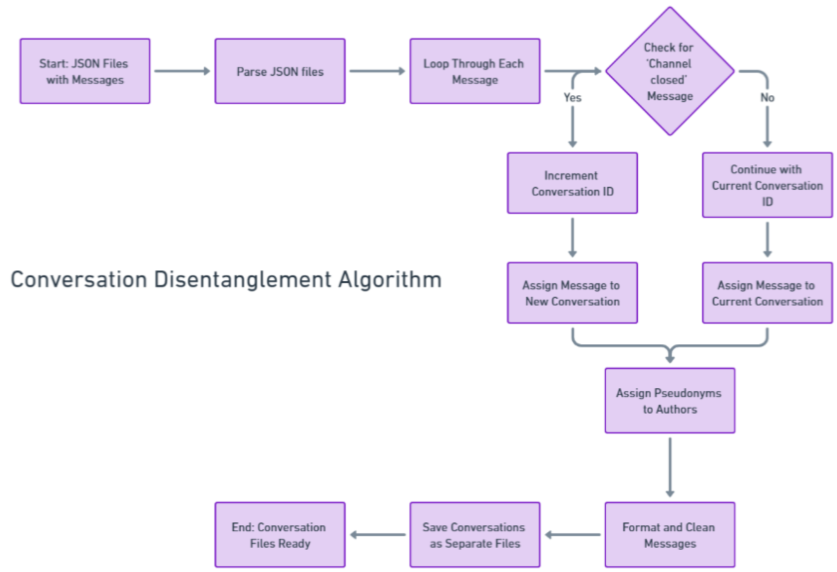


Figure 2.7: Flowchart diagram of the disentanglement algorithm

with an anonymous shield (they choose when they enter the platform a username to engage with the community), this step is essential for anonymizing the data, this step adds an extra step of making sure the users personal information isn't linked to their messages while still maintaining the integrity of the conversational context. The algorithm also performs a series of formatting operations on the messages, which include cleaning the text and standardizing the timestamp format. These operations are essential for ensuring data consistency and readability, thereby enhancing the quality of the subsequent analysis. To do this, I have made a spreadsheet of 500 common first and last names that have been extracted from a public list on the internet. This provides a total of $500 \cdot 500 = 250,000$ unique names to choose from, more than the number of unique author names in the MDS. The algorithm goes through the list of unique author names and assigns to each one of these unique first-last name pairs (e.g., Alice Jones). Upon completion of the processing loop, the algorithm finalizes each conversation by assigning it to a unique file. This segmentation into individual files is not only important for organization but also facilitates easier access and analysis of specific conversations. The output, therefore, consists of a series of files, each representing a distinct conversation, ready for further examination. This structured output is instrumental in enabling a comprehensive and detailed analysis of the conversational data, ultimately contributing to a richer understanding of student interactions and learning processes in mathematical education.

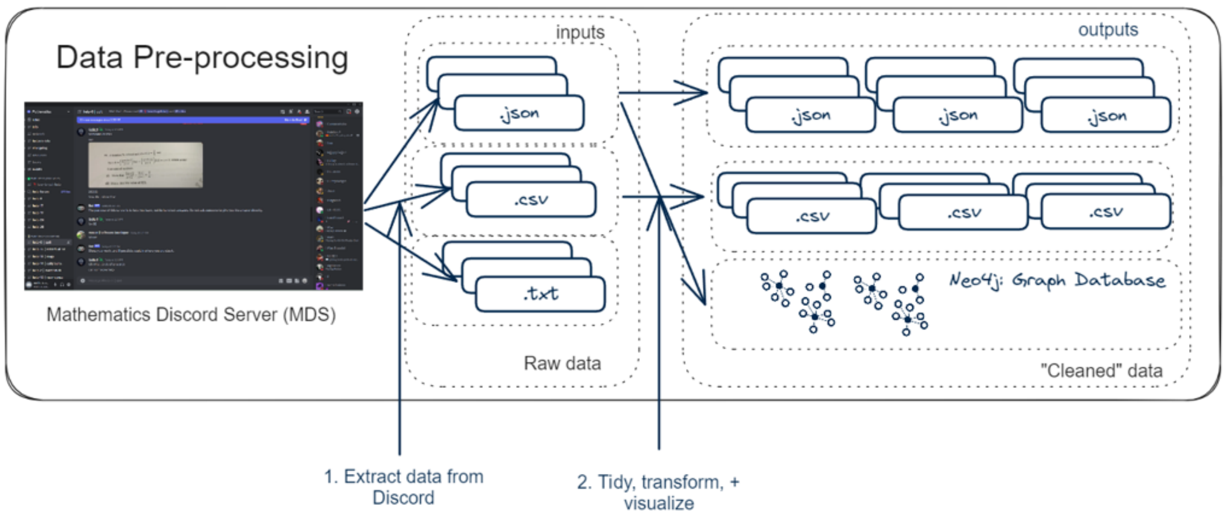


Figure 2.8: Schematic of the results of Chapter 2

2.5 Results

2.5.1 Overview of disentangled conversations

Following an exhaustive data disentanglement process detailed in the previous sections, the algorithm successfully parsed 6,384,642 messages from the MDS, structuring them into 205,885 distinct conversations. This transformation represents the initial objective of the study: to organize a vast corpus of mathematical exchanges into discrete segments, ready for analysis.

The dataset is organized into distinct conversation threads, each represented as a discrete file equipped with detailed metadata. This includes a 'conversation_id' that uniquely identifies each thread, along with the constituent messages. The metadata for each message provides information on the author ('author_id' and 'author_name'), the exact time of posting ('timestamp'), the message content, any associated attachments, and user mentions. This comprehensive metadata framework enables a nuanced analysis, capturing the multifaceted nature of interactions within the community. To ensure the integrity of the analysis, timestamps have been formatted to "MM/DD/YYYY HH:MM:SS" format in Eastern Standard Time.

2.5.2 Implications for research and educational practice

The disentangled dataset stands as a rich resource for educational researchers, offering an unprecedented opportunity to explore the mechanisms of online mathematical learning. Potential studies can leverage this data to investigate patterns of student engagement, peer-to-peer and mentor-student interactions, and the evolution of mathematical understanding within the community. Figure 2.9 provides a sketch of how the larger set of conversations can be queried to hone in on a subset of conversations of interest to a particular study. For example, in Chapter 4, I examine the various conceptions of the derivative that arise in the conversations, and one way to look at this diagram is to see that a script could be run to just pull out the conversations where students talk about derivative problems. If another researcher is interested in diagrams, they could write a script where only conversations where students talked about or included a diagram could be looked at. Furthermore, the dataset can serve as a benchmark for developing and testing machine learning models aimed at identifying educational outcomes and predictors of student success in online learning environments.

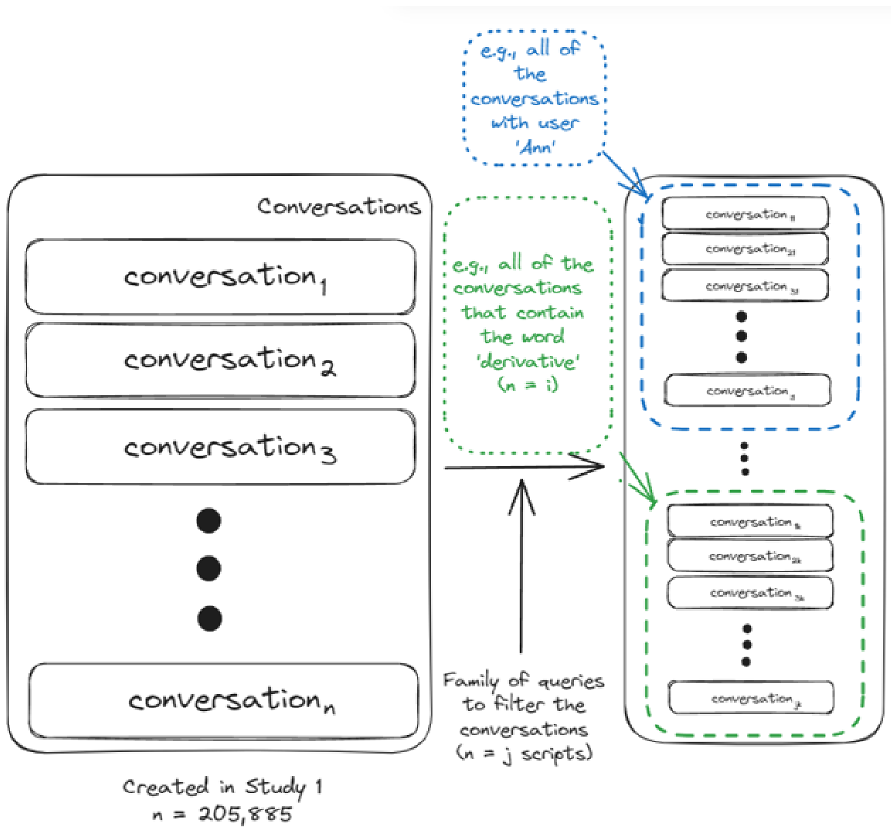


Figure 2.9: What can be done with *MathConverse*

Insights derived from this dataset have the potential to significantly influence online educational practices. By analyzing student behaviors and conversation patterns, educators can tailor their approaches to foster more effective online learning communities. The dataset may also provide valuable benchmarks for the development of automated tools to support teaching and learning in digital platforms. Future research could explore the application of natural language processing techniques to detect and support students' learning progress, enabling real-time interventions in similar educational communities. In summary, the methodological innovations and resulting dataset from this study contribute a foundational tool for advancing the field of mathematics education research, particularly within the context of studying how students are learning mathematics online. The data not only reflects the rich and complex nature of mathematical discourse but also provides a springboard for future investigations aimed at enhancing the educational experiences of learners who use online tools or platforms to learn and study mathematics.

CHAPTER 3

Characterizing Engagement on the MDS Platform: Analyzing Participants' Activity Patterns and Conversation Content

3.1 Introduction

The widespread use of smartphones and the internet amongst our learners and educators has transformed the teaching and learning of mathematics (Engelbrecht et al., 2020). The digital age has ushered in new modalities of learning where thousands of students can learn together in Massive Open Online Courses (MOOCs), interactive learning applications, and online learning forums, all of which have become fundamental in reshaping education at all levels (Haleem et al., 2022). This shift transcends traditional classrooms, with students worldwide turning to a diverse set of online resources for educational support. This trend is not confined to a single educational platform; it is a widespread phenomenon seen across countless online learning environments. Van de Sande's (2011) study of an open, online calculus help forum exemplifies a shift towards platforms that provide unrestricted access to students of diverse educational backgrounds and levels, facilitating questions and engaging in subject-specific discussions. Central to this study is the *MathConverse* dataset which offers a dynamic and collaborative space where students from across the world can engage together in mathematical discourse, encompassing everything from homework help to open-ended discussions on specific mathematics topics.

The history of distance learning has roots evolving from early forms in the 1700s to the multi-modal online platforms we see our learners gravitating to today. This history, which will be elaborated upon in the following section, sets the stage for understanding the current era of online learning, providing context for understanding the nature of the MDS that is studied in this dissertation. By situating the study within this historical continuum, I present how education has been reshaped by the internet, technology, and distance learning, leading to the ways in which we as education researchers are in an era of being able to collect and understand how students are engaging with their learning with these tools and platforms. Much of the data we collect and gather from these tools and platforms is 'unstructured' and text-based, and from this data there is enormous potential to see how students and educators are engaging with mathematics content.

Traditional tutoring research has predominantly focused on small-scale, in-person sessions and their effectiveness (see Roscoe and Chi, 2007). *MathConverse*, the dataset I described in Chapter 2 composed of 200,000 student-tutor interactions, provides an opportunity to study how students engage with mathematics problems at large-scale. Unlike data collected from conventional academic settings, this dataset's strength lies not just in its volume but also in its variety. Students and helpers are coming from all over the world, bringing in problems from a wide range of topic areas, and because they are looking for help that is not face-to-face, the ways in which they have to communicate about their work is different than it is when working in person. The data collected encompasses text, mathematical drawings, links to videos, screenshots, and more. The multi-modal conversations from *MathConverse* provide a comprehensive view of the tutoring process, capturing nuances of communication that extend far beyond traditional text-based analysis. While this study looks primarily at the text exchanges that take place between the students and the helpers, there is promise to integrate the other forms of multi-modal metadata that has been collected in *MathConverse*. Techniques from machine learning and natural language processing are used to navigate this complex, unstructured dataset, extracting meaningful patterns and insights that would be challenging to discern in smaller, more homogeneous datasets.

3.1.1 Research questions

Building upon the expansive scope of the *MathConverse* dataset and its potential for unprecedented insights into online tutoring, this study is structured around a series of research questions that aim to understand the dynamics of engagement within the Mathematics Discord server. The research questions of the study are as follows:

1. What does activity at the participant- and conversation-level look like in the Mathematics Discord Server (MDS)?
   (a) How often do users participate in the MDS across time, in terms of months, days, and times of day by asking questions or helping one another?
   (b) How do students and helpers converse on MDS (engage within conversations) in terms of time, that is, length of conversations and time between turns of talk?
2. What do students and helpers discuss in the MDS?

(a) What types of questions do the conversations address?

(b) What are the topics of the conversations?

The findings from these research questions will offer educators and researchers valuable insights into the dynamics of student engagement in external online mathematics communities in ways that could help align research and practice more closely with the observed needs and preferences in these independent online learning environments.

3.2 Background

3.2.1 Historical development of distance learning

The way students learn and teachers teach mathematics was transformed with the advent of distance learning, a phenomenon that spans centuries of innovation and adaptation. This section examines the historical progression from traditional forms of distance learning to the online learning platforms learners are using in the digital age. By tracing this evolution, I aim to illustrate the foundations upon which current online learning environments, including platforms like the MDS, are built. This exploration not only highlights key technological advancements but also highlights some of the shifts in the ways students and teachers facilitate learning that have led to the current state of learning from online resources.

3.2.1.1 From distance learning to digital education

The development of online learning platforms, which can be tied back to the notions of distance learning in the 1700s, has a history that predates the advent of computers or the internet. Early initiatives, such as Caleb Phillips' shorthand lessons via postal mail in 1728 and Anna Eliot Ticknor's correspondence school in the 1800s, were key in laying the groundwork for modern educational technologies (Harting and Erthal, 2005). These initial forms of remote learning, emphasizing accessibility and self-paced learning, have continually shaped educational methodologies and technologies, evolving into the diverse spectrum of online learning platforms we see today. This transition, marked by technological advancements throughout the 20th and 21st centuries, represents a continuum of innovation deeply rooted in distance education's core objectives. As Kwon et al. (2021) notes, this evolution led to the development of structured online courses and programs, most notably Learning Management Systems (LMS) like Blackboard and Massive Open Online Courses (MOOCs) like MIT OpenCourseWare which have become integral in managing course content and facilitating interactions in an organized, structured manner. These platforms, while technologically advanced, continue to echo the core objectives of early distance education, including overcoming geographical barriers, enabling flexible learning, and reaching diverse learner populations.

The historical evolution of online learning platforms, tracing back to the early initiatives of distance learning, has played a key role in shaping the current forms of online education. This development from traditional correspondence courses to contemporary, technologically advanced online learning methods represents a continuum of innovation deeply intertwined with the foundational principles of distance education. As we explore this progression, it becomes evident that the technological advancements of the 20th and 21st centuries have not only broadened the scope of distance education but also facilitated the emergence of dynamic, community-driven online learning spaces. Platforms like the MDS are prime examples of this evolution, marking a significant departure from structured learning environments to more fluid, collaborative spaces where knowledge is co-constructed by participants.

3.2.1.2 Technological advances shaping online learning

The transition from structured educational systems like Learning Management Systems (LMSs) to dynamic, interactive online communities signifies a pivotal shift in how students are learning online. While traditional LMSs, like Blackboard or Canvas, are designed for structured content delivery and management, focusing mainly on instructor-led activities, platforms like Discord offer a more informal, adaptable environment that can be more conducive for student-led interactions. Discord's design, facilitating real-time, peer-to-peer communication and collaboration in both academic and social contexts, aligns with the principles discussed by Kurianski et al. (2022) for humanizing mathematics education. It enables the creation of diverse channels for mathematical discussions, homework help, and social interactions, fostering a well-rounded community. This contrasts with the more rigid, content-focused nature of traditional LMSs. Such an environment, as highlighted by Kurianski et al. (2022), is essential for building community and encouraging communication, key aspects of humanizing online learning experiences. Discord's flexibility not only supports academic engagement but also promotes a sense of belonging and community among students, echoing the study's emphasis on intentional community building in virtual educational settings. Such an environment aligns with the Fehrman et al.'s (2021) emphasis on the importance of meaningful interactions in asynchronous online discussions (AODs), further supporting the shift towards more engaging and inclusive online educational paradigms. In examining the dynamics of online education, the study by Lin and Overbaugh (2007) provides valuable insights into the impact of allowing students to choose their mode of online communication. Their findings suggest that such flexibility can enhance student satisfaction and engagement, a concept that resonates with informal, community-centric learning on platforms such as the MDS. This aligns with my study's objective of exploring engagement patterns and the effectiveness of tutoring in environments where learners have autonomy in their interactions. Specifically, it touches on the core aspects of my research questions which aim to understand user activity, the nature of discussions, tutor responses, and the overall quality of messages within the server's community.

The MDS illustrates an example of a platform that exists in modern educational times where technology can facilitate spontaneous, peer-driven knowledge exchange, transcending conventional classroom dynamics. This study centers on the unstructured, community-driven platform of the MDS, which contrasts with more structured environments like how students learn with tutors in person or engage with question-answering in more structured online forums. By focusing exclusively on the *MathConverse* dataset derived from the server, my goal is to provide a window into the ways in which learners are engaging with online learning spaces, in this case one that focuses on mathematics. This approach allows for a concentrated analysis of what activity looks like in terms of participation and what is being talked about in a less formalized, yet very active online community. The insights gained will contribute to a deeper understanding of how such unstructured digital platforms can influence how students are learning mathematics, highlighting both the potential and challenges of online learning in contemporary educational settings.

This shift has significant implications for the nature of tutoring and collaborative learning in online environments. The flexibility and accessibility of platforms like the MDS enable ways of learning that can be more natural and spontaneous for learners, where traditional educational roles and norms that tend to be hierarchical and restrictive can be redefined. Here, the role of tutors evolves from being 'more knowledgeable others who transmit knowledge' to 'facilitators of knowledge', collaborating and adapting to the unique needs and identities of learners within these digital communities. I argue that by exploring the evolution of online learning platforms as well as taking note and doing more research on what students want and need out of their learning experiences, the field can gain a better understanding of how these environments are shaping tutoring dynamics and collaborative learning. This way, our research can inform the learning conditions for students in our classroom spaces. This research has been useful in contextualizing my study of the MDS within the broader narrative of online coursework, online learning, and what we need to do as educators in order to make sure our students can stay engaged if we transition more of our high school and university teaching to being through online platforms.

3.2.2 Dynamics of online learning: Opportunities and challenges

3.2.2.1 Enhancing student engagement and learning through online platforms

Platforms like the MDS highlight the evolving dynamics of education through online learning environments. These platforms demonstrate both the strengths and weaknesses of providing assistance remotely compared to in-person. While online communities offer unparalleled resource access, promote asynchronous learning, and can connect learners from across the world, they lack the immediacy of physical classrooms, relying instead on asynchronous, text-based interactions that can limit real-time guidance, and there is less infrastructure for understanding the quality of instruction our learners are receiving. Despite this, their multi-modal assistance, like sharing videos or diagrams, showcases their adaptability, balancing the pros and cons of online versus traditional educational settings. Interaction patterns within these online communities vary greatly, which can depend on the type of platform, as well as the engagement of the participants in the platform (Tareen and Tahir Haand, 2020). This variability can also span from members being completely engaged contributing to discussions to passive participants absorbing information, mirroring in-person engagement but with unique moderation and facilitation challenges (Borup et al., 2020). Understanding the diversity in these interaction patterns, shaped by the evolving norms and etiquette of the online community, is imperative for investigating user activity patterns and participation nature in online learning environments, aligning with research question 1 of this study.

3.2.2.2 Navigating challenges in online learning environments

In online learning environments, traditional educational hierarchies are undergoing a transformative shift. These platforms foster a collaborative and peer-to-peer learning ethos, democratizing the exchange of ideas beyond the conventional teacher-student model. This shift is evident in the diverse nature of discussions, ranging from specific problem-solving queries to broader exploratory dialogues and debates (Engelbrecht et al., 2020). Borba et al. (2016) highlight this transformation in their paper, noting how mobile technologies and social media have reconfigured the flow of knowledge, enabling a dynamic, two-way exchange between teachers and students. This evolution represents a significant departure from the standard didactical contract, as described by Herbst and Chazan (2012). In their conceptualization, the didactical contract traditionally delineates a more unidirectional flow of knowledge from teacher to student. The current shift, however, re-positions learners as active participants in their learning, thus challenging and reshaping the traditional boundaries of this contract.

In online learning environments, technological features like discussion thread layouts, messaging capabilities, and resource-sharing tools play a decisive role in shaping user interaction and communication. Threaded forums, while popular, often encounter issues such as a lack of focus and minimal in-depth interaction, leading to predominantly surface-level discourse (Gao et al., 2013; Suthers et al., 2008; Thomas, 2002). Suthers et al. (2008) highlight the inherent challenges in collaborative knowledge construction that rely on threaded reply forums. In a study of community question-answering sites, Roy et al. (2022) observed that these forums, while popular, often face issues related to coherence and convergence, leading to predominantly surface-level discourse. This can result in incoherent or misaligned collaborations between learners and more knowledgeable others. In contrast, when learning communities are built on platforms like Discord and norms are constructed around how 'synchronous' the feedback is, it can lead to more focused and meaningful exchanges, aligning with the principles of the Productive Online Discussion Model proposed by Gao et al. (2013): (1) discuss to comprehend, (2) discuss to critique, (3) discuss to construct knowledge and (4) discuss to share. The roles of anonymity and identity in these platforms present a delicate balance between ensuring user security and maintaining accountability in discourse, further complicated by the global and culturally diverse nature of participation, introducing challenges such as language barriers and varying educational objectives.

In these online-mediated environments, tutors evolve from content deliverers to facilitators, leveraging diverse communication tools for asynchronous student engagement. This shift necessitates adaptive communication strategies and effective management of varied student inquiries. While the asynchronous nature of traditional forums, as discussed in Gao et al. (2013), allows for reflective and thoughtful participation, it can also lead to delayed responses and challenges in maintaining student engagement. In contrast, the Mathematics Discord Server's design promotes a more immediate, collaborative learning experience, encouraging dynamic interaction and participation. This approach is particularly relevant for tutors in collaborative and peer-to-peer models, who transition from knowledge providers to discussion facilitators. Such a transition necessitates strategies to foster inclusive and interactive atmospheres, aligning with RQ 1a and RQ 1b that examine how and how often users participate in the MDS in terms of time. The findings from these analyses will offer a distinct contrast to the typical dynamics of threaded forums as outlined in Chapter 2 (see subsection 2.2.2).

3.2.3 Social dynamics and diversity in online learning

3.2.3.1 Help-seeking behaviors in online learning spaces

The perception of help-seeking within education has undergone significant transformation over time. Historically, educators and students tended to view help-seeking negatively, often associating it with dependency or an avoidance of personal academic effort (Gonida et al., 2019). This viewpoint, however, began to shift, particularly following Nelson-Le Gall's (1981) seminal work on help-seeking. Nelson-Le Gall differentiated instrumental help-seeking—a proactive approach aimed at acquiring understanding and problem-solving skills—from executive help-seeking, which is characterized by the pursuit of direct answers over comprehension. Instrumental help-seeking has since been recognized as a key factor associated with academic success (Schenke et al., 2015), marking a significant change in the educational paradigm. As new digital and online technologies have emerged both in and outside the classroom, traditional methods of seeking help have evolved to meet the diverse and dynamic needs of today's students. Online platforms have emerged as pivotal tools in this transformation, offering alternatives that extend beyond the limitations of face-to-face educational interactions. These technological advancements have not only reshaped the ways in which students seek help but have also broadened the accessibility and inclusivity of educational resources.

3.2.3.2 Overview of tutoring interventions

Research on tutoring consistently demonstrates its positive impact on student learning. For instance, Elbaum et al. (2000) conducted a meta-analysis revealing substantial benefits of one-to-one reading interventions, especially for students at risk for reading failure. While many tutoring studies focus on reading and young children, the principles of effective tutoring, such as personalized instruction, are equally applicable to the subject of mathematics. In the context of mathematics tutoring, Cunningham et al. (2011) provide compelling evidence for the effectiveness of online homework-completion tutoring in remedial mathematics courses. Their study revealed that integrating online homework tutoring, even when introduced later in the semester, significantly improved student performance and attendance in comparison to traditional pencil-and-paper methods. Specifically, the analysis showed that students who engaged in online homework tutoring had notably higher COMPASS test pass rates and Math Lab attendance, with these results being statistically significant. This indicates that interactive online homework, which prompts active problem-solving rather than passive note-taking, can substantially enhance the learning experience and outcomes for students in remedial mathematics courses. Importantly, the study suggests that the earlier students begin receiving homework tutoring, the more pronounced the benefits, highlighting the value of timely and interactive tutoring interventions in mathematics education.

Despite the value of tutoring, there remains a notable gap in data regarding the moment-to-moment work of tutors, especially in online environments. As large-scale, detailed analyses of the improvisational strategies used by tutors are scarce, this gap highlights the need for more in-depth research into the dynamics of tutoring interactions. This dissertation focuses on an online tutoring environment to explore how students engage with mathematical concepts. Previous research like Bloom (1984) and Fuchs et al. (1997) mainly centered on program design and the efficacy of learning interventions. In contrast, this study investigates what goes on behind the scenes in these out-of-school tutoring interventions by understanding the conceptions and ideas emerging during tutoring conversations. This approach seeks to fill a gap in the literature identified by Roscoe and Chi (2007) by examining the content of conversations in online tutoring contexts, with a particular focus on student conceptions about the derivative. By investigating the nuances of peer-tutoring conversations in an online platform, this study contributes to a broader understanding of effective tutoring strategies and the role of conversation content in facilitating learning. It offers valuable insights into how students conceptualize and engage with mathematical problems, enhancing our understanding of the patterns and dynamics of mathematics discussions in peer-tutoring settings.

3.2.3.3 Role of online platforms in facilitating help-seeking

As the role and norms around help-seeking in education have evolved, a significant development has been the rise of online platforms, which have come to play a central role in facilitating modern help-seeking behaviors. These online environments have not only increased the accessibility of educational resources but have also redefined the nature of student interactions and support mechanisms. Online platforms, encompassing numerous digital resources such as interactive lessons, educational videos, and forums, have become integral in providing flexible and accessible learning opportunities. Their significance is particularly pronounced for traditionally underrepresented groups in STEM fields, including women, first-generation college students, and racial minorities (Jay et al., 2020). These groups often encounter additional challenges in conventional educational settings, such as the fear of reinforcing stereotypes or threatening their self-image, which can impede their willingness to seek help. The anonymity and normalized methods for seeking assistance offered by online platforms can potentially alleviate these barriers, contributing to a more equitable and inclusive educational environment. Additionally, unlike more cost-prohibitive spaces like university course forums, online chat spaces can provide universally accessible spaces where students can anonymously post specific assignment-related questions. This transparency not only fosters a community spirit but also expands the scope of learning beyond individual efforts. Covering a wide array of subjects, these platforms cater to learners at different academic stages, from primary education to postgraduate studies (Van De Sande, 2013), thereby enhancing the overall educational experience and transcending traditional academic boundaries. Lastly, focusing specifically on the teaching and learning of mathematics, online platforms can help offer learners solutions to specific problems but also foster a deeper understanding of mathematical concepts. The MDS, exemplifies one of these platforms as a space where students go for in help. It represents a dynamic and collaborative space where students engage in mathematical discourse, from seeking homework help to participating in open-ended discussions. When mathematics instructors give homework, there are a number of more cost-prohibitive resources for students to get help, including private tutoring, paying for online services for access to solutions, amongst other things. As educators, we must examine that students will go to the internet or their peers for help with their assignments, with or without explicit policies against such actions, and if we wish for them to get help from more official resources, we must continue to ask whether such resources are accessible for all.

3.2.3.4 Diversity and inclusion in online learning

The transition to learning in online learning environments is not without its challenges. Miller (2021) highlights that even experienced educators, who value student relationships, find maintaining these relationships increasingly challenging as curricular demands intensify, particularly in remote settings. Her paper highlights that during the shift to remote learning during COVID-19, educators worked to *(re)build relationships*, which involved setting clear expectations, responding to non-academic needs, and promoting peer-to-peer interactions. Efforts such as these are vital in keeping online learning platforms active, healthy, and successful, where establishing a positive tone and fostering a sense of community can significantly impact student engagement and success. However, Chapman et al. (2010) emphasize the persisting *digital divide*, noting the disparities in technology access and technical skills among teachers, particularly in high-needs schools, which can hinder effective implementation of online learning and professional development. This ends up having an effect on students, where those from socioeconomically disadvantaged backgrounds may lack the necessary resources for remote learning, which can affect their sense of belonging and academic success (Cleary et al., 2006; Miller, 2021).

In exploring the challenges faced in online learning, a critical issue that emerges is the persistent gender gap, despite the significant increase in women's participation in distance education (DE). Gnanadass and Sanders (2018) highlight that while women have increasingly enrolled in DE programs, surpassing men in both undergraduate and graduate levels, the culture of online learning is not devoid of gender biases. The fundamental challenge lies in the perception and design of technology and online platforms, which often fail to be gender-neutral (Patterson, 2009). This oversight leads to a replication of traditional gender norms and inequities in online learning spaces, contradicting the emancipatory goals of DE (Weatherly, 2011). Similarly, the evolving demographic of millennials and gen-Z in online education introduces unique dynamics and challenges in the learning process. This new age of learners, having grown up immersed in digital technology, exhibit distinct expectations from online learning environments. In her doctoral dissertation, Yonekura (2006) emphasizes that millenial students tend to favor learning environments that are interactive, flexible, and well-structured. Their study reveals that millennials, irrespective of gender, value aspects such as convenience, time management, and the ability to learn at their own pace. However, these students also express concerns over the lack of interaction, the role of the instructor, and the design of online courses. Female millennials, in particular, showed higher levels of satisfaction with online learning experiences compared to their male counterparts. These findings suggest that for millennials, the quality of online education hinges significantly on the effectiveness of course design, the instructor's engagement, and the opportunities for meaningful interaction. Consequently, educational platforms like the Mathematics Discord server must consider these preferences and challenges to cater effectively to the millennial demographic, ensuring that the server's structure and instructional approaches align with their distinct learning styles and needs.

In summary, online learning comes with both beneficial dynamics and distinct challenges. As we transition towards increasingly digital educational environments, the traditional paradigms of help-seeking, tutoring, and educational interactions are being reshaped. Online mathematics learning communities embody this transformation, offering new opportunities for engagement and learning while also highlighting the importance of addressing the digital divide, fostering inclusive communities, and catering to the unique needs of diverse student populations, including millennials and Gen-Z learners. As we navigate these complex dynamics, it becomes imperative to adopt a theoretical framework that comprehensively addresses these multifaceted aspects of online learning. In the subsequent section, I introduce the conceptual framework, focusing on aspects of connectivism and communities of practice (CoPs) as key frameworks for understanding the dynamics of online learning. Through the lens of connectivism, I explore how the digital learning network within the MDS facilitates the creation and navigation of diverse information sources, in line with Siemens's (2005) concept of learning as a process of network building and connection creation. Additionally, the exploration of some elements of CoPs, as conceptualized by Lave and Wenger (1991) and Wenger-Trayner and Wenger-Trayner (2015), will shed light on the social aspects of learning within the MDS. I examine how this online platform functions as a *community of interest*, fostering shared learning experiences, collaborative problem-solving, and a sense of community among its members.

3.3 Conceptual Framework: Some Takeaways from Connectivism and Communities of Practice

This study provides a unique opportunity to observe a learning space that is different from a classroom or office hours. In this space, students are coming in from all over the world, with different kinds of knowledge, experiences, and goals. Unlike the more traditional academic settings, the community members of the Mathematics Discord Server are free to come and go as they desire, and participate as little or as much as they want. The complexities of how students acquire knowledge and how community members interact with each other necessitates a multifaceted theoretical approach.

I adopt core ideas from the theories of connectivism and CoPs as an integrative conceptual framework to analyze the dynamics of learning in this learning environment. Connectivism, as articulated by Siemens (2005), is a theory of learning which emphasizes the creation of digital learning networks, highlighting the importance of navigating diverse information sources in a rapidly evolving digital age. CoPs focus on members within a shared domain of interest (in this case, mathematics) engaging in discussions, helping each other, and sharing information (Lave and

Wenger, 1991; Wenger-Trayner and Wenger-Trayner, 2015). While is it essential to clarify that the MDS does not fully align with a CoP, the theory provides some helpful framing when talking about the mathematical interactions that are taking place in the learning environment. In the following sections, I provide brief primers on these theories and how they guide the analytical work I do in this chapter.

3.3.1 Connectivism

Connectivism, a learning theory coined by George Siemens to account for our transition into the digital era, positions learning as a process of creating connections and network building. Siemens (2005) notes that it is not merely the acquisition of a specialized set of information that epitomizes learning but the ability to see and create links between fields, ideas, and concepts is what is viewed as a key component to learning. This perspective shifts the focus from the content of 'what is known' to the process of 'knowing', placing a premium on the capability to continuously adapt and acquire new information in a rapidly changing environment. One of the foundational principles of connectivism is that learning and knowledge are deeply embedded in a diversity of opinions. It emphasizes that learning is a dynamic process of connecting specialized nodes or information sources, and that "learning may reside in non-human appliances" (Siemens, 2005, p.5). Another important aspect to this theory is that maintaining and nurturing these connections is essential for effective learning.

This is a particularly useful theory in the case of studying the mathematical interactions that take place in the MDS, as it recognizes the fluid and networked nature of knowledge in the digital age. Learners in the community are naturally going out of their way to take the content they are learning in school and talk about it and discuss it with others, and this community provides a space for members to do continuous learning, adapt what they are learning in 'school' mathematics by talking about it with others, and as they engage more in the discussions, use search functionality or be pointed to references of other conversations or references where people have worked on similar problems. By intertwining these elements with some of the the communal and reflective aspects of communities of practice as outlined by Wenger-Trayner and Wenger-Trayner (2015), it helps provide a lens of looking at the ways in which members are engaging and learning mathematics as a part of being a community member.

3.3.2 Communities of practice

While CoPs provide a foundational framework for understanding collective learning processes in shared human endeavors their traditional characteristics need to be carefully considered when applied to online platforms like the MDS (Smith et al., 2017). CoPs are typically characterized by groups of people who share a concern or a passion for something they do, and learn *collectively* through regular interaction Wenger (1998); Wenger-Trayner and Wenger-Trayner (2015). This framework becomes particularly relevant when examining how people come together to learn mathematics in the MDS, where participants are engaged in synchronous dialogue, exchange of knowledge and resources, and collaborative problem-solving, all in an effort to enhance understanding in mathematics teaching and learning. However, the nature of participation in the MDSs leans more towards individual knowledge construction, as the members are not coming together in pursuit of a common goal or collective professional development that is found in traditional CoPs.

According to Henri and Pudelko's (2003) classification, virtual communities are seen to exist on a continuum, ranging from 'communities of interest', where the focus is primarily on individual knowledge construction, to 'goal-oriented' and 'learners' communities', where there are more collective goals, activities proposed by an instructor, and ultimately to 'communities of practice', which embody the highest level of social bonding and collective professional development. Within this spectrum, the MDS aligns most closely to a *community of interest*, as learners primarily engage in 'knowledge construction for individual use' converging around their shared interest in mathematics to enhance personal understanding and problem-solving skills. This orientation towards self-directed learning, as opposed to the more communal and interdependent learning in traditional CoPs, highlights the unique positioning of the MDS on the continuum of virtual learning communities.

3.3.2.1 Domain, community, and practice in the MDS

The MDS, while not fitting neatly into the traditional CoP framework, exhibits elements of the domain, community, and practice elements described by Wenger-Trayner and Wenger-Trayner (2015). The MDS's domain revolves around mathematics, providing a shared space for members of varying expertise to engage in mutual knowledge sharing and problem-solving. This aligns with the assertion by Wenger-Trayner and Wenger-Trayner (2015) that CoPs are groups bound by a shared passion or concern, learning collectively as they interact regularly. As noted by Zhang and Watts (2008), online communities like this one play a significant role in knowledge dissemination and skill development. In the MDS, the sense of community is fostered through active participation, collaboration, and support amongst members. This sense of belonging and mutual respect is vital for sustained learning and engagement, as demonstrated in the interactions captured in the *MathConverse* dataset. The practice within the MDS is characterized by a shared repertoire of resources like problem sets, solution strategies, and other sources. This is one aspect where the MDS diverges from a traditional CoP, as the practices being developed and sustained are not being made explicit to the members of the group. While there are some rules and norms within the server that have developed over time to make sure that more effective help can take place (e.g., help channels with one conversation at a time, norms about how to ask for help and how to provide help in the form of rules, moderation in the group), this isn't the primarily aim for the community.

In summary, the analysis of the MDS through the lens of connectivism and CoPs highlights its effectiveness as a learning environment. It demonstrates the server's role in facilitating collective learning and knowledge creation, serving as a testament to the adaptability of the CoP framework in encompassing online platforms and its relevance in contemporary educational contexts.

3.4 Methods

3.4.1 Overview of methodological approach

Social science research is undergoing a transformative phase, driven by the integration of computational approaches and large-scale text analysis. In this section, I introduce the use of text-as-data methods, an approach that has revolutionized social science research by enabling the analysis of large collections of documents and text interactions (Grimmer et al., 2022). This methodology is centered on the analysis of text data to draw inferences about human behavior and social phenomena. Its growing popularity stems from the pivotal role language plays in social interactions, whether in the form of legislation, historical documentation, religious discourse, or everyday communication (Benoit, 2020). This approach is particularly apt for this study, given the extensive text interactions within the Mathematics Discord server. Utilizing advanced computational techniques, text-as-data methods allow for the processing and analysis of large-scale text collections, transforming copious amounts of conversational data into structured, analyzable information.

Historically, the use of text data in quantitative social science research was limited, primarily due to the challenges in processing and analyzing large amounts of text. The advent of methods from NLP, a subfield of machine learning, has made it feasible to process, analyze, generate, and make meaningful inferences from and about text data, something that had previously been impractical (Chowdhury, 2003; Jurafsky, 2009). We can see text-as-data methods in action through several commercial applications: our emails providing suggestions for how to complete a sentence, or our ability to ask Google how to translate an English phrase to French. Additionally, applications of NLP extend to academic research settings. For example, political scientists have used NLP to classify political affiliation from speech (Yu et al., 2008), and mathematics educators have used topic modeling to analyze five decades of articles from a prominent education journal (i.e., *Journal for Research in Mathematics Education*) to determine how the major areas of focus in mathematics education have shaped and shifted (Inglis and Foster, 2018). I argue that as education researchers, we should know more about these methodologies and the advantages they might offer to study how people learn and teach mathematics. By employing advanced computational techniques, this methodology allows for the dissection of complex online interactions, transforming vast amounts of unstructured conversational data into structured, analyzable formats, further enabling an in-depth exploration of the patterns of communication, knowledge sharing, and learning processes that take place in this online learning community.

3.4.2 Dataset description

3.4.2.1 Nature and source of the dataset

As outlined in Chapter 2, the *MathConverse* dataset, derived from the Mathematics Discord Community, offers a comprehensive view of online mathematical discourse. The *MathConverse* dataset is extensive, encompassing 6,384,642 messages across 205,885 unique conversations. This data was contributed by 41,275 unique users, reflecting a wide array of mathematical discussions and tutoring interactions. This dataset is ideal for this study due to its extensive volume, the diversity of interactions, and the structured yet dynamic nature of communication among participants. The rich text and conversational data provide an unparalleled opportunity to examine the nuances of learning and problem-solving in a digital educational setting. The selection of the *MathConverse* dataset is justified by its alignment with the theoretical frameworks of connectivism and communities of practice. The server's environment, where knowledge is collaboratively built and shared, mirrors the core principles of these frameworks. Furthermore, the dataset's volume and diversity allow for a robust analysis of patterns in tutoring strategies, student engagement, and problem-solving approaches, addressing the research questions effectively.

3.4.3 Characterization of engagements

When the learners are looking for help, they are met with the following screen as we see in Figure 3.1. Typically, I have seen that there is always a couple of help channels always available, but usually most of the 49 help channels are being occupied with a student asking for help, showing that at all times of the day, the server is lively and burgeoning with activity with students.

3.4.4 Methodological approaches

In this section, I describe the methodological approaches I use to analyze the qualitative content of conversations within *MathConverse*. First, I describe how I use a pre-trained machine learning classifier designed to distinguish questions from other sorts of messages. Next, I expand on the

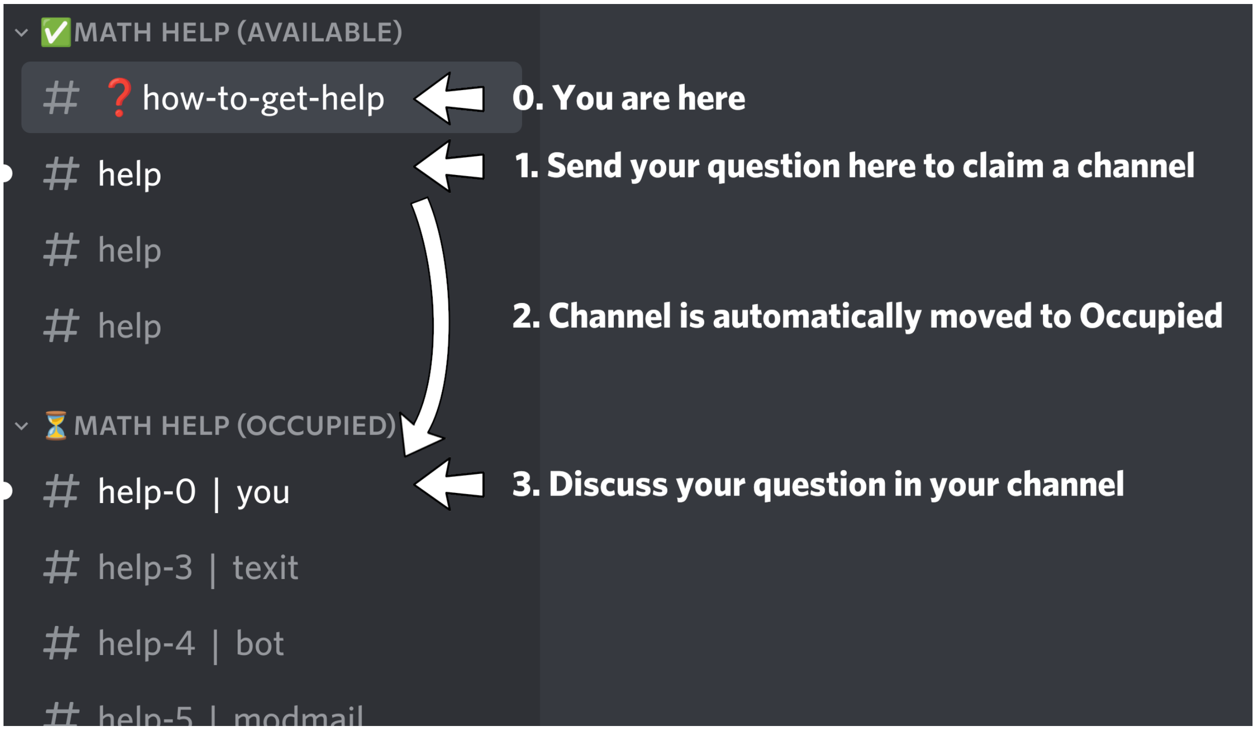


Figure 3.1: What the users see when they are looking for help in the help channels

classifier's application to to the messages of *MathConverse*, first elaborating on the criteria for question classification and and how I evaluate performance by comparing the automated outputs with my own set of manual annotations. Lastly, I discuss the use of topic modeling as an analytical tool to automatically extract latent themes from the conversations. This commonly used technique shows the most common words that cluster together in the documents, which can give some insight into the mathematical subjects and pedagogical dialogue present in conversations. From here, I show how I use this label the conversations for further analysis (e.g., the conversations with that are labeled with the 'derivative' topic were filtered and used as a dataset to study which conceptions of the derivative emerged in *MathConverse* in Chapter 4).

3.4.4.1 Using a pre-trained machine learning model to find questions

In this study, I use the Transformers library, a powerful and widely-used framework developed by Hugging Face, which provides access to a large collection of pre-trained machine learning models[1] which enable researchers to do inference on their own datasets. Within the context of this study, I use the "Keyword Statement vs. Question Classifier for Neural Search" model developed by the Haystack

[1]Pre-trained machine learning models, are equipped with billions of parameters (values that the model adjusts to make accurate predictions) that are refined during an initial learning period. This takes place using massive datasets to gain proficiency on a number of generic NLP tasks before the model is trained to perform well on specific tasks (fine-tuned).

team. This model excels in differentiating between keyword-based statements and natural language questions, a key capability for my task of finding the messages within *MathConverse* which contain questions. For example, in the dataset, a message might ask, 'Can someone help me understand how to find the zeros of this polynomial?' while another might simply say, 'Thanks, I got it.' Using this classifier as a preliminary step allowed me to efficiently filter through the conversations, distinguishing pertinent questions from other message types. This filtration is foundational to my methodological approach, as it enables a focused analysis on the nature and types of questions posed by participants, in aim of answering RQ 2a.

In sum, the classification model takes in each message as an individual data point, and processes it through the classification function in sequence. Subsequently, the outputs are reincorporated into the dataset, now paired with labels that state whether the message is a 'question' or a 'statement'. With the messages categorized, the dataset can be filtered for the next step of analysis to look at the types of questions posed in the conversations. The results of the classification are shown later in Figure 3.9, but briefly, around 20% (approximately 1 million) of the messages were questions.

3.4.4.2 Classifying questions by type

To address RQ 2a, which examines the types of questions present in the conversations in *MathConverse*, this part of the study builds on the foundational work of Graesser and Person (1994). Their seminal work on question asking during tutoring sessions provides a detailed framework for decomposing mathematics students' questions into detailed aspects of presupposition and focus, as elaborated in their prior work (Graesser et al., 1992). Their framework provides guidance on how to classify questions into 18 distinct types, capturing a range of both cognitive and linguistic diversity present while students ask questions about mathematics problems. The original taxonomy of 18 question types devised by Graesser et al. presents an extensive assortment, from ones that elicit short answers like verification (e.g., 'Is a fact true?'), disjunction (e.g., 'Is it X or Y?'), and quantitative questions (e.g., 'How many?'), to more elaborate ones that require open-ended longer answers, to those that are interpretational, and those that are procedural in nature. However, in examining the dialogues within *MathConverse*, I noticed a recurring theme: many questions often fit into several of these categories, blurring the lines of distinction set by the original framework. For example, there were often questions that would elicit an open-ended, yet procedural reply. To adapt Graesser et al.'s (1992) framework for this study, my goal was to build a more streamlined set of clearly defined categories that would ensure each category was not only distinct but also fully representative of the types of questions posed within the questions posed in *MathConverse*.

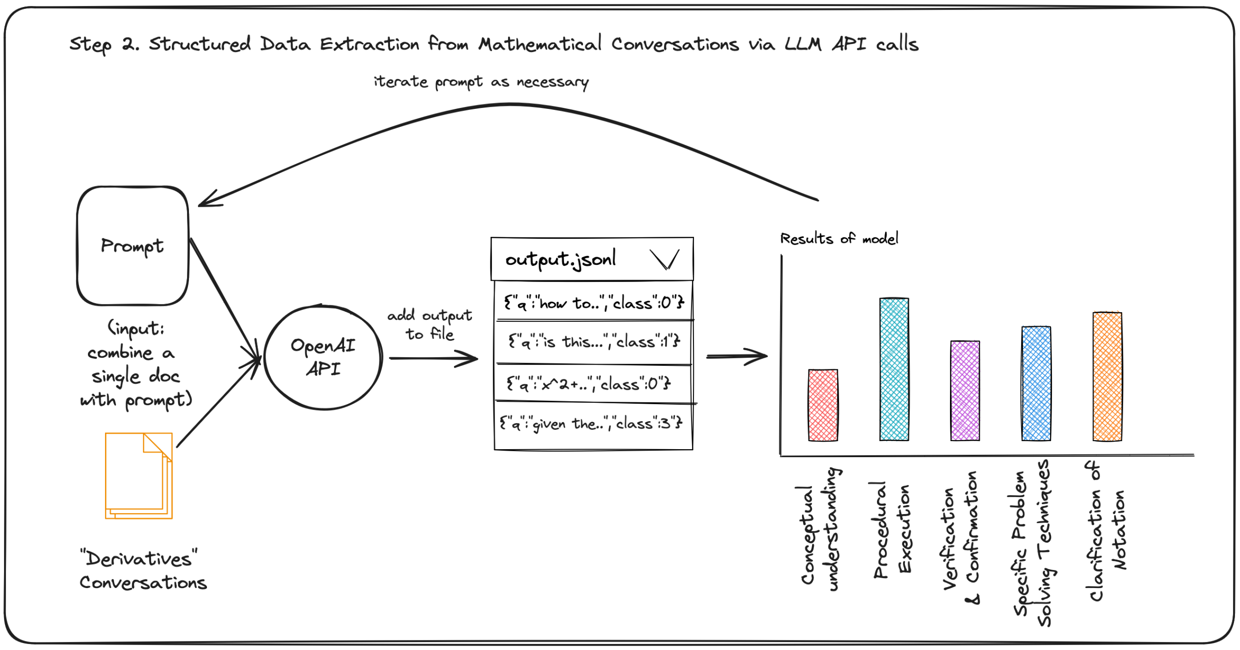


Figure 3.2: Structured data extraction with LLMs

Table 3.1: Question Type Coding Framework

Question Category	Definition	Example
Basic Inquiry	Seeks a simple, factual response or clarification of basic information.	What is the first step in solving a quadratic equation?
Procedural Reasoning	Concerns the steps or procedures taken to achieve a certain goal or solve a problem.	What are the steps involved in isolating a variable in an algebraic expression?
Context Inquiry	Questions that require understanding and applying context or scenarios.	How does the concept of elasticity apply to the demand for a product in a competitive market?
Exploratory Inquiry	Aims to explore concepts or definitions, often asking for elaboration or examples.	Can you explain the concept of a derivative in calculus?
Assertive Communication	Involves making a statement or claim, often with confidence or certainty, sometimes to persuade.	The theorem can be proved by applying the principle of mathematical induction.

The Question Type Coding Framework, as illustrated in Table 3.1, serves as the basis for the coding process. The categories are designed to capture the overarching nature of the questions asked in the conversations found in *MathConverse*, in some order of increasing abstraction. With the coding framework in place, the next phase of the study involves applying these categories to the dataset, both through manual annotation and the use of a LLM. The coding process is designed as follows:

Procedure for Analyzing *MathConverse* Dataset Questions

1. Apply the "Keyword Statement vs. Question Classifier" across the entire dataset of *MathConverse* to isolate messages that are likely to be questions.

2. Extract a randomized sample of 1000 (potential) questions to construct a corpus for manual analysis. I made this set larger than the sample I intended to label as I anticipated there might be some messages erroneously labeled by the previous classifer as questions that I would not want to label.

3. Methodically review each question within the sample, ensuring its classification as question and categorize it with the most suitable category from the Question Type Coding Framework.

4. Repeat this process to the first 500 messages in the sample labeled as questions.

5. Once this process has been finished, identify the labeled questions by ID, and run the prompt to automatically classify each question using GPT-3.5 Turbo. I use GPT-3.5 Turbo as it blends cost effectiveness (10x cheaper than GPT-4), performance, and ease of use. Additionally, the GPT models provided by OpenAI were shown to outperform the open-source models on similar tasks looking at inferring student errors in mathematics tutoring dialogues (Wang et al., 2023).

6. Compare the results, see if there are any places where I believe the model got the prediction correct and I was incorrect with the first application of coding. Correct these responses and set this group aside as the new 'gold standard' dataset.

7. Run evaluation metrics to evaluate performance. If performance is good, run the classification model over a larger set of questions from the sample.

This process of labeling a random sample of the data, then running a model against the same responses and comparing the results is going to come up again in Chapter 4 when I use GPT-3.5 Turbo to classify conversations for the presence of derivative conceptions. As the OpenAI models are decoder-based, these models are generally best for 'text-generation' tasks, and as both of these tasks are classification tasks, it is important to make sure that the output is constrained in some way that can be useful for me to use at scale. I use Pydantic, a Python library which uses Python type annotations to validate, serialize, and deserialize data as a way to structure and validate the outputs of the model. This is useful because I can run the script and produce structured output in JSONL format, which is suitable for further analysis without manual manipulation (Figure 3.2).

3.4.4.2.1 Prompting Unlike traditional machine learning models, pre-trained large language models offer the ability to produce quality, unique insights without necessitating the retraining of the model. *Prompts* are instructions provided to an LLM to enforce rules, streamline workflows, and guarantee particular attributes (both in terms of quality and quantity) of the produced output (White et al., 2023). *Prompt engineering* is a technique for directing an LLM's responses toward specific outcomes without altering the model's weights or parameters, relying solely on strategic in-context prompting. While the transformer architecture-based LLMs I implement in this study show remarkable linguistic capabilities, in the context of this project, creating prompts for the various tasks which use the LLM has been an iterative process that must take into consideration a number of limitations: (1) LLMs lack persistent memory, that is, I must assume that every LLM call will take in one new input, provide one output, and refresh; (2) identical prompts can produce variability in responses due to the probabilistic nature of the LLMs; (3) LLMs are trained on historical data, and unless prompted to connect to the internet, will not have any real-time awareness or updates; and (4) LLMs can (and will) generate plausible yet factually incorrect information, often called *hallucinations*.

For the purpose of this study, I developed a prompt construction framework using Python that categorizes questions into predefined types—Basic Inquiry, Procedural Reasoning, Context Inquiry, Exploratory Inquiry, and Assertive Communication——based on their content and intent. Each category is defined as follows: Basic Inquiry focuses on straightforward, factual responses; Procedural Reasoning on the steps to achieve goals or solve problems; Context Inquiry on understanding contexts or scenarios; Exploratory Inquiry on the questions that dive deeper into concepts or seeking further explanations; and Assertive Communication on making confident claims or persuasions.

```
from enum import Enum
from pydantic import BaseModel, Field

class QuestionType(Enum):
    BASIC_INQUIRY = "Basic Inquiry"
    PROCEDURAL_REASONING = "Procedural Reasoning"
    CONTEXTUAL_INQUIRY = "Context Inquiry"
    EXPLORATORY_INQUIRY = "Exploratory Inquiry"
    ASSERTIVE_COMMUNICATION = "Assertive Communication"

QUESTION_TYPE_DESCRIPTIONS = {
```

```python
    QuestionType.BASIC_INQUIRY: "Seeks a simple, factual response or clarification of basic information.",
    QuestionType.PROCEDURAL_REASONING: "Concerns the steps or procedures taken to achieve a certain goal or solve a problem.",
    QuestionType.CONTEXTUAL_INQUIRY: "Questions that require understanding and applying context or scenarios.",
    QuestionType.EXPLORATORY_INQUIRY: "Aims to explore concepts or definitions, often asking for elaboration or examples.",
    QuestionType.ASSERTIVE_COMMUNICATION: "Involves making a statement or claim, often with confidence or certainty, sometimes to persuade."
}

QUESTION_TYPE_EXAMPLES = {
    QuestionType.BASIC_INQUIRY: "What is the first step in solving a quadratic equation?",
    QuestionType.PROCEDURAL_REASONING: "What are the steps involved in isolating a variable in an algebraic expression?",
    QuestionType.CONTEXTUAL_INQUIRY: "How does the concept of elasticity apply to the demand for a product in a competitive market?",
    QuestionType.EXPLORATORY_INQUIRY: "Can you explain the concept of a derivative in calculus?",
    QuestionType.ASSERTIVE_COMMUNICATION: "The theorem can be proved by applying the principle of mathematical induction."
}

def create_prompt(question: str) -> str:
    categories = ', '.join([qt.value for qt in QuestionType])
    return (f"Given the question: '{question}', classify it into one of the following categories based on its content and intent: "
            f"{categories}.")

class QuestionClassification(BaseModel):
```

```
    classification: QuestionType = Field(
        description="A list of classifications for a question,
            indicating the types or categories the question belongs
             to based on predefined criteria. Each classification
            must be one of the enumerated types defined in `
            QuestionType`."
    )

# Example usage:
question = "What is the derivative of x^2?"
prompt = create_prompt(question)
print(prompt)
```

Listing 3.1: Question classification prompt

The code excerpt in Listing 3.1 presents an Enum class defining these categories, associated descriptions, and examples, which show some of the work that goes into building a prompt. The function `create_prompt` encapsulates this logic by accepting a question and classifying it into the appropriate category. This function ensures that the questions posed to the LLM are systematically categorized, thereby streamlining the process and enhancing the quality of the responses.

Next, I go over some of the details for my use of topic modeling in order to answer RQ 2b.

3.4.4.3 Topic modeling

Topic modeling is a popular Bayesian statistical model to categorize individual texts and identify topics or patterns across documents (Vayansky and Kumar, 2020). This method is particularly suitable for analyzing large amounts of conversation data, where underlying themes might not be immediately evident. The task at hand is one of dimensionality reduction; there are over 200,000 distinct conversations within *MathConverse*, but given the nature of the help channels' objectives, it is reasonable to presume that certain themes recur in what the learners discuss. Topic modeling is an effective methodology for pattern identification by detecting which words commonly occur together. This allows for the creation of 'topics', where each document is then characterized by a mixture of these topics. Manually analyzing and categorizing topics in large datasets can be not only impractical and time-consuming but also vulnerable to interpreter bias and inconsistency. Topic modeling addresses these challenges by automating the process, thereby enhancing efficiency, consistency, and objectivity. Moreover, it facilitates the involvement of subject-matter experts, who can interpret and contextize the results effectively at scale.

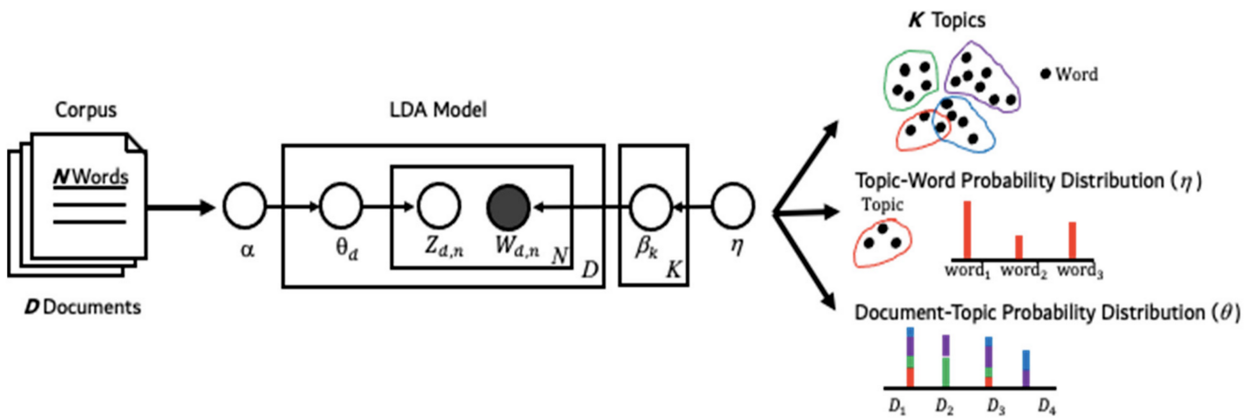


Figure 3.3: Graphical model representation of LDA. The boxes are "plates" representing replicates. The outer plate represents documents, while the inner plate represents the repeated choice of topics and words within a document. Diagram from Hwang and Cho (2021)

.

To implement the topic modeling process, I use MALLET (MAchine Learning for LanguagE Toolkit; McCallum, 2002). MALLET uses Latent Dirichlet Analysis (LDA) via an implementation of Gibbs sampling, a statistical technique meant to quickly construct a sample distribution, to create its topic models. As shown in Figure 3.3, the model provides three important outputs: (a) the word-topic distribution, η, which provides a list of words and associated weights for each topic; (b) the document-topic probability distribution θ, which provides for each document the probability of it being generated by words coming from each of the topics; and (c) topic distributions, $\beta_{1:k}$, where each β_k is a distribution over words. I settled on a 30-topic model after running several models; the choice of the number of topics for an LDA model is a hyperparameter to be set by the researcher a priori, and one that is determined based on subjective evaluation from the researcher after looking at the results (Zhao et al., 2015).

3.5 Findings

3.5.1 RQ 1: What does activity at the participant- and conversation-level look like in the MDS?

3.5.1.1 Activity patterns of the MDS

Figure 3.4 illustrates the frequency of user participation over the observed period from November 2021 to January 2023. I provide this chart to guide the inquiry of RQ 1a which asks what user participation looks like in the MDS across time, in terms of months, days, and times of day by going over not only the general trend of participation over time but also the variability in daily message counts.

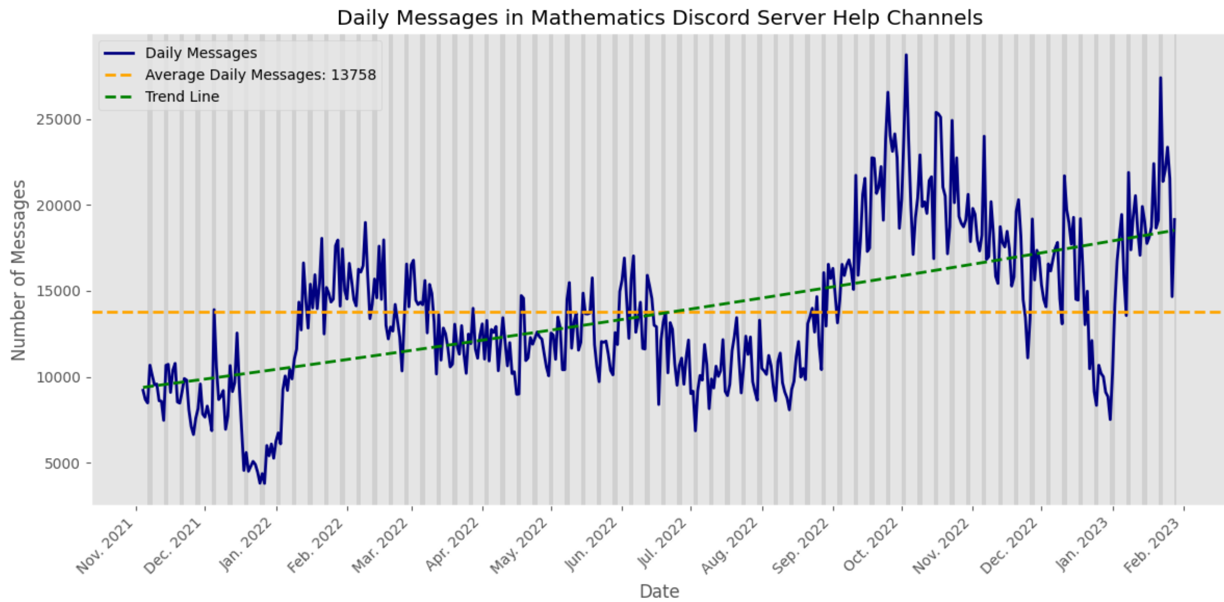


Figure 3.4: Number of daily messages in the help channels from November 2021 to January 2023

The fluctuations revealed by the blue line in Figure 3.4 provide a detailed view of user engagement within the MDS. A cyclical pattern emerges, which indicates both weekly rhythms and seasonal variations in activity. For instance, the visible dips corresponding with the weekends—represented by shaded vertical strips—suggest a regular decrease in activity during these periods. Similarly, the reduced frequency of messages during typical holiday seasons emphasizes the seasonal characteristic of user participation. Additionally, the trend line, shown in green, is particularly revealing. It not only confirms a general growth in user participation, as evidenced by the approximate 500 message per month average increase. This rate of growth over time in the help channels represents significant and sustained engagement by users, providing quantifiable evidence of increasing user interaction over time within the MDS.

Referring to Figure 3.5, where the columns represent each day in the week, the rows represent each hour of the day, and the darker the cell the more activity we see in that hour on the day, we see that between the times of 10PM - 2PM Eastern time on any day of the week is when the server is most active. While the server is still very active on the weekend, the most active time is in the center of the plot, with the most active day of the week being Wednesday. This is a surprising result, as this is when we could expect high school and university students to either be in class or attending office hours. This prompts further investigations into why students might turn to these online learning environments rather than traditional forms of help they could receive while present in school or

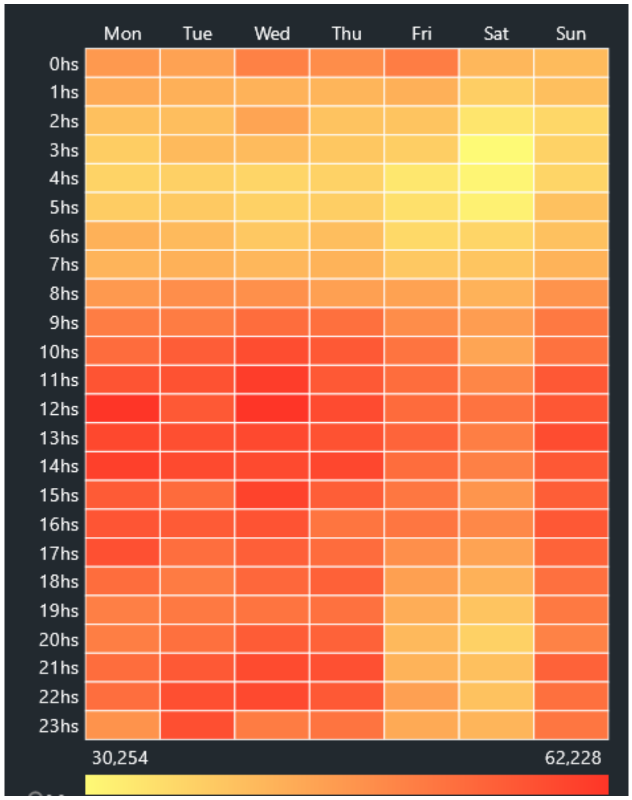


Figure 3.5: Heatmap observing most popular times to send messages in the help channels. Times are in Eastern Standard Time (EST)

university. In the next section, I look to some quantitative measures that can provide some metrics of the quality of the exchanges that happen between the students and the helpers in these conversations.

3.5.1.2 How students and helpers engage with each other

This section aims to answer RQ 1b by examining how students and helpers converse on MDS in terms of time- and message-based involvement. Specifically, I investigate three key metrics: interchange interval, the length of conversations, and the number of turns of talk within each conversation. To better align with the goals of the research question, it is important to make some adjustments to the data and be clear about the definitions. First, *interchange interval* is defined to capture only the intervals between messages when there is a shift in the sender, focusing on exchanges between different participants. This distinction is necessary for accurately measuring the responsiveness of participants in the MDS, as it excludes consecutive messages from the same individual, a common practice that exists in this platform. Similarly, *turns of talk* is measured based on the number of exchanges that happen between different participants. By focusing on intervals where speakers change—either from student to helper or vice versa—the goal is to capture genuine conversational exchanges, as if I just measured time between messages I would often be measuring time between messages sent by the same person.[2] In each analysis I provide a visual through the use of a histogram and provide some initial interpretations of each metric, and to finish the section, I provide a table of summary statistics (Table 3.2) to synthesize the findings.

3.5.1.2.1 Interchange interval The interchange interval," representing the time span between consecutive messages involving a shift in the sender's role, is visualized in the histogram provided in Figure 3.6. The histogram illustrates the distribution of message intervals across all dialogues within the MDS, showcasing the platform's capacity to facilitate quick exchanges. An initial analysis yielded an average interchange interval of 1.81 minutes (1 minute and 49 seconds), with half of the back-and-forth exchanges happening within the first 0.30 mins (18 seconds). This is a surprising result, as it shows that nearly half of the replies tend to be nearly synchronous between the students and helpers.

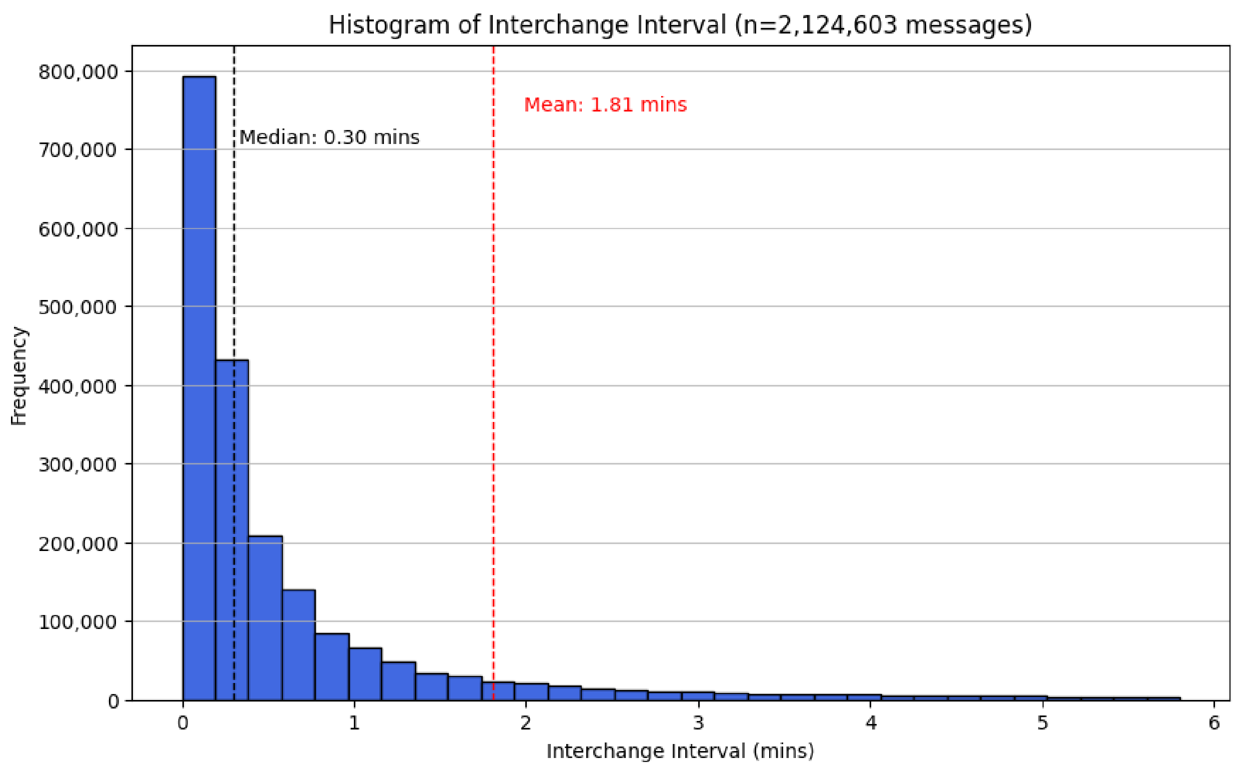


Figure 3.6: Histogram with interchange interval values for the $n = 2,124,603$ exchanges between participants

[2]Perhaps a strength as a communication tool, but a weakness as an analytical tool, but often when people are using chat communication platforms they will send several messages in succession rather than simply sending one message and waiting for a reply.

3.5.1.2.2 Duration of conversations As the interchange interval analysis suggests promptness in exchanges, the duration of conversations provides some insights into the extent to which students and helpers engage with mathematics problems. With an average duration nearing 24 minutes and a median indicating that half of the discussions are concluded in just over 11.28 minutes (11 minutes 17 seconds), the findings signal that conversations can be both quick and can lead to more prolonged, dialogues. This *duration of conversations* metric provides another measure to examine the multifaceted nature of engagement in the help channels of the MDS, as it helps gauge how long a helper will stick around to help a student or how long a student is willing to get help on a problem.

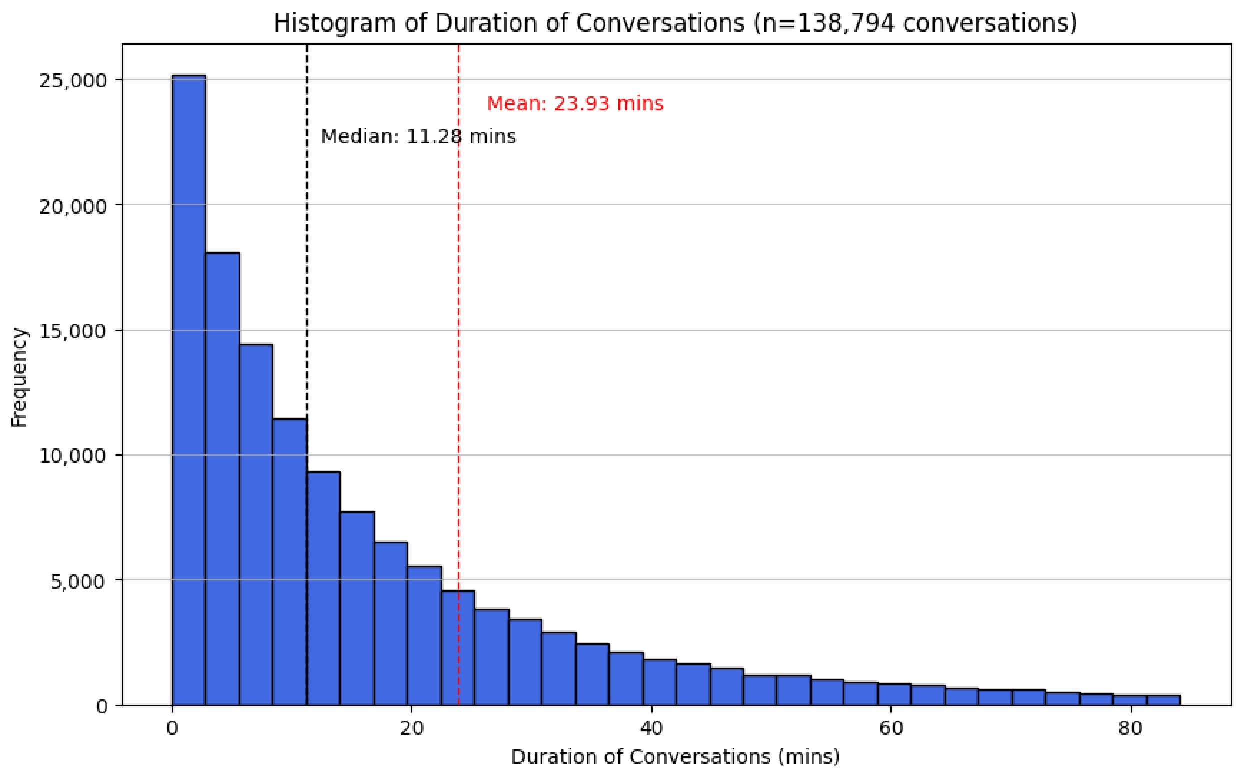


Figure 3.7: Histogram with duration of conversation values for the $n = 138,794$ conversations in the analytic sample

The histogram in Figure 3.7 shows a right-skewed distribution, indicating that a majority of the interactions transpire over a shorter time-span than the mean, suggesting that while half of the conversations are shorter than 11 min 17 seconds, there seems to be a diversity in the range of duration that is worth investigating. From my years as an observer of the learning space, this parallels what I have seen, as some students come looking for quick clarifications while others aim to have more lengthy, conceptual discussions. In this context, comparing the average length of conversations to the interchange interval helps shed some light on how the MDS serves as a learning space. The quick turnaround time for individual messages combined with the overall lengthy conversations exemplifies a learning environment that can accommodate a wide range of mathematical inquiries.

The choice to examine turns of talk was a natural progression from here, as it helps tie together the duality of these two findings. The number of turns signifies not only the back-and-forth nature of the conversations but also provides a measure of how the helpers in the community have made a commitment to not just ‘give an answer’ when a student comes in and asks a question, and that students are willing to continue to engage once they know they will have to do more than just copy down an answer. These three metrics together—interchange interval, duration, and turns of talk—can help provide a comprehensive picture of how students and helpers engage with one another on the platform.

3.5.1.2.3 Turns of talk The histogram in Figure 3.8 showcases a distribution where most conversations involve a moderate number of exchanges. By measuring the number of times the sender changes within each conversation, the histogram reveals an average of 9.35 turns per conversation, with a median of 6 exchanges, suggesting that while there is a tendency for conversations to be relatively concise, there is still substantial room for extended dialogue when necessary. This indicates an environment where, typically, questions are not only answered but also discussed to a certain extent, pointing towards an engaged community that values thorough understanding over quick fixes.

What was found by the interchange interval and duration of conversations measures matches what is presented here in turns of talk. A lower median relative to the mean indicates that while most conversations might only require a few exchanges to resolve simpler questions, there are instances of longer exchanges, where more elaborate, iterative discussions are happening. These findings inspire the research question of what types of questions are being asked, which comes up in the next section as I describe the work of using pre-trained machine learning models to help classify questions by type. Doing this work can help better understand why students come and ask questions in the help channels of the MDS, and whether they come to ask simple clarifying questions that require brief clarifications or ask more complex problems that require more extensive back-and-forth discussion.

3.5.1.2.4 Final synthesis In Table 3.2, I present the summary statistics for the three analyses. In addressing RQ 1b, the findings from this analysis reveals participant engagement that aligns neatly with elements from theories of connectivism and CoP. The MDS serves as a prototype for how some learners have gravitated towards online spaces for help on homework problems, and of a type of space where the rapid interchange of ideas and information mirrors the connectivist principle of learning as a network-forming process. The empirical data, drawn from a meticulous analysis of interchange intervals, the length of conversations, and turns of talk, aligns with Siemens’ (2005)

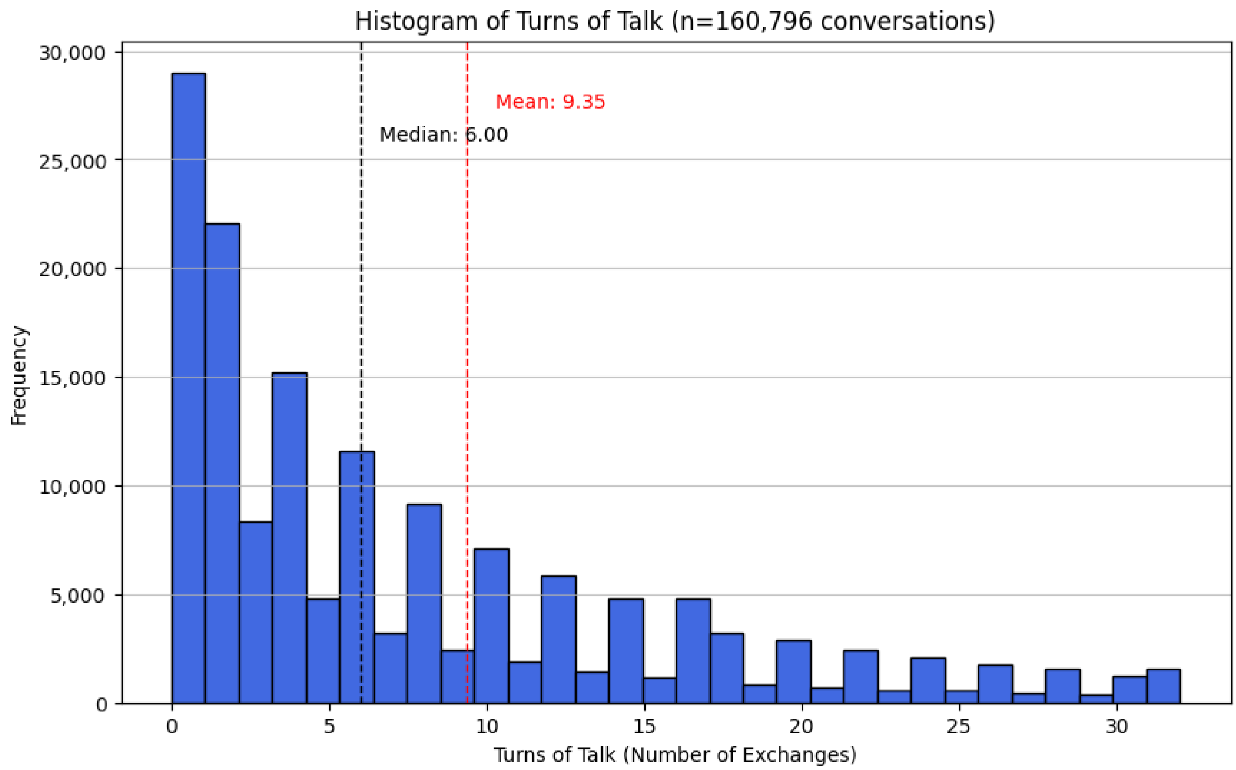


Figure 3.8: Histogram with number of turns of talk for the $n = 160,796$ conversations in the analytic sample

characterization of knowledge as the construction of connections across a diverse informational network. This environment, as reflected in the duration of conversations metric, accommodates a wide array of engagement from rapid, transactional interactions to more sustained and engaged discussions. With conversations ranging from a few seconds to several hours, the MDS embodies the connectivist principle that learning flourishes on a diversity of opinions, approaches, and time commitments. This adaptability is required for online learning communities to thrive, as it aligns with learners' needs to have the capacity to navigate and adapt in an ever-evolving landscape of knowledge. By integrating the conceptual underpinnings of connectivism with some of the community-based principles of CoPs, the findings in this section help highlight the MDS as a type of online space that provides a new mode of learning—one that is networked, communal, yet individually tailored.

Table 3.2: Revised Summary Statistics for Communication Dynamics and Turns of Talk

Statistic	Interchange Interval[1]	Duration of Conversations	Turns of Talk (Messages)
Count	2,124,603 messages	138,794 conversations	160,796 conversations
Mean	1 min 48 secs	23 min 56 secs	9.35
Std Dev	10 min 35 secs	42 min 59 secs	10.24
Min	2 secs	2 secs	1
25%	7 secs	4 min 10 secs	2
50%	18 secs	11 min 17 secs	6
75%	49 secs	26 min 8 secs	13
Max	7 hr 19 min 59 secs	9 hr 53 min 2 secs	50

[1] "Interchange Interval" refers to the interval between consecutive messages sent by different participants (student to helper or vice versa), aiming to capture genuine interactive exchanges.
Note: Summary statistics are presented with outliers, zero-duration conversations, and conversations with important missing data excluded.

3.5.2 RQ 2: What do students and helpers discuss in the MDS?

3.5.2.1 Question types

In order to understand what types of questions are being asked in the *MathConverse* dataset, the first step is to distinguish between messages that pose questions and those that do not. This analysis serves a dual purpose: it refines the dataset for further in-depth examination of question-based messages (RQ 2a) and provides a quantitative measure of the proportion of questions versus statements within the dataset.

3.5.2.1.1 Statement vs. Question Classifier Results As presented in Figure 3.9, the application of the "Keyword Statement vs. Question Classifier" shows that approximately 20% of the messages, amounting to nearly 1 million, are classified as questions. This ratio is significant as it shows that within the conversations, there are many follow-up questions. The presence of such a large number of questions helps validate the notion that students feel empowered to ask questions in the MDS, aligning with the findings of Graesser and Person (1994) who found in their study that student questions were approximately 240 times as frequent in tutoring settings than in classroom settings.

The next set of findings aims to bridge the quantitative metrics of engagement with the qualitative content of the discourse. By using the framework outlined in Table 3.4.4.2, I aim to address RQ 2a by classifying the types of questions addressed in the conversations.

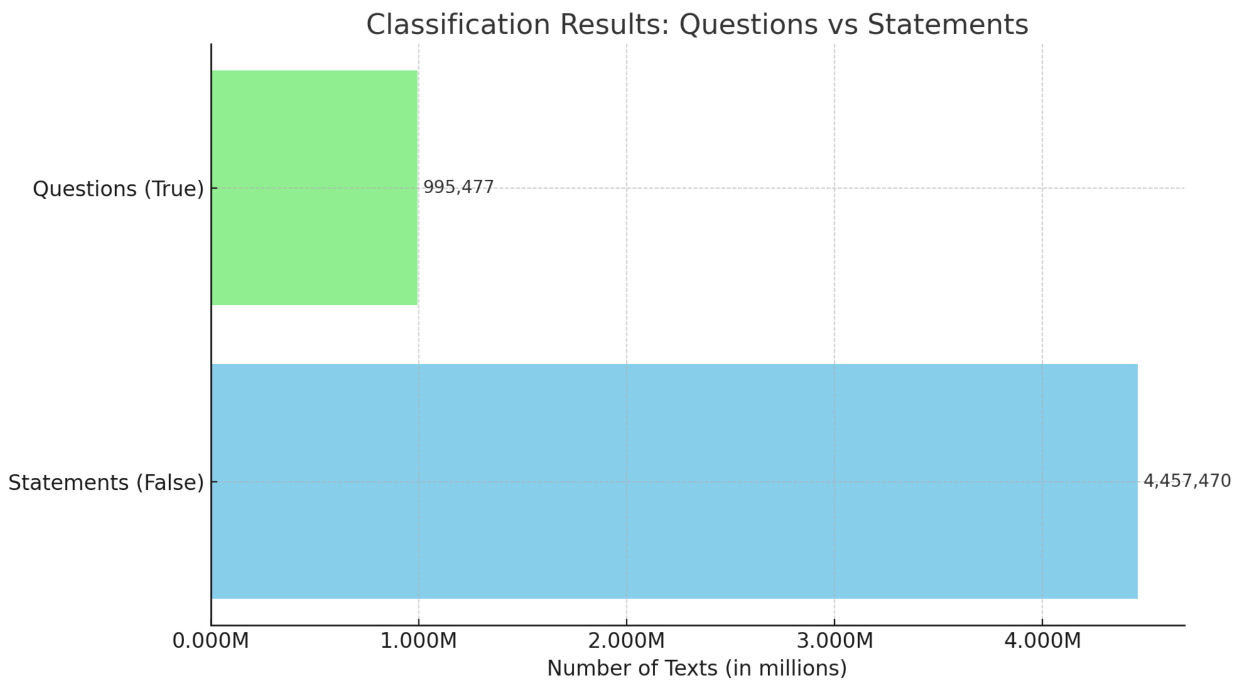


Figure 3.9: Results from the question vs. statement classifier model, highlighting the proportion of queries among the discourse in the MDS.

3.5.2.1.2 Initial findings on random sample In order to assess how well the model performs on this task, I first needed to hand label a set of questions myself. I took a random sample of 1000 questions from the prior work done classifying questions from statements, and used Label Studio, an open-source data labeling platform to label the data. To my delight, the prior classifier had done a very good job deciphering questions from statements, and it was rare that I had to skip a data point and move on to the next one, as nearly every message I ran into was a question. In Figure 3.10, I provide a bar chart of counts of the questions types of the random sample that became the validation set to compare the model results to.

As shown in the bar chart, the most common questions were 'Basic Inquiry' ($\frac{199}{503} = 40.0\%$) and 'Procedural Reasoning' ($\frac{163}{503} = 32.4\%$). Basic Inquiry being the most common result aligns with my observations and what educators might expect, as many of the questions in the conversations started off with a question that can be answered quickly (e.g., 'Is $sin(x) + cos(x)$ defined for the entire domain?' or 'Does this reasoning look right?'), or finished off with some sort of clarifying question (e.g., 'Is that it?', 'What answer did you get?', or 'Would I use spherical coordinates here too?'), whether or not the question came from the student or the helper. The result that was surprising for me was what the model was helpful for. Typically, qualitative coding is done by two or more researchers, as it is important to be able to check the work by establishing inter-rater reliability.

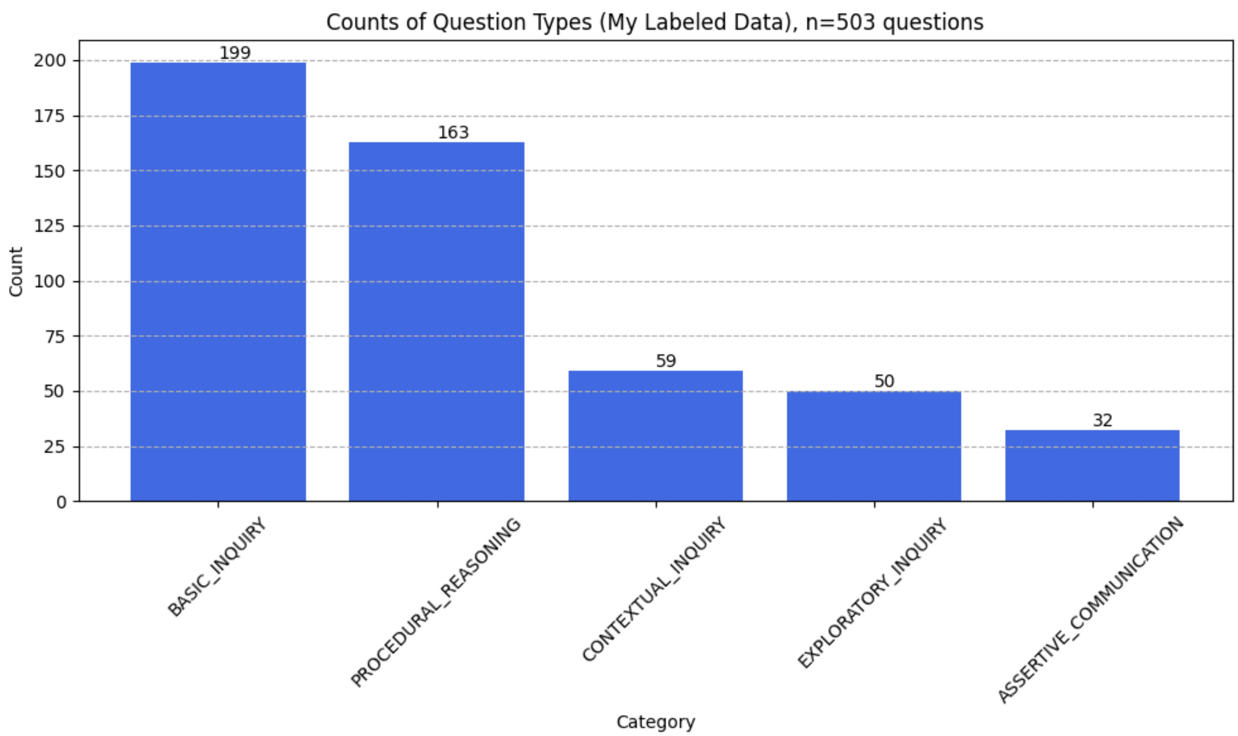


Figure 3.10: Counts of each question type on random sample of $n = 503$ questions, hand-labeled examples.

After labeling the examples and running the messages through the model, I assessed each predicted label against mine and found that *I needed to change many of labels in order to be consistent with the definitions I had in my framework*. Specifically, the model was doing a much better job at identifying instances of procedural reasoning where I was erroneously identifying the messages as exploratory inquiry. What I present above are the hand-labeled counts, after doing some calibration, drawing another random sample, re-labeling, and running the model again, which is typical of the cyclical process necessary in this type of work.

In Figure 3.11, I present a 5×5 confusion matrix, a specific type of visualization used in machine learning and statistics to evaluate the performance of classification models. The matrix is helpful in helping assess the performance of classification algorithms, as we can see how well it performed on individual categories. In this case, the matrix is structured as a 5×5 grid, where both the rows and columns represent the five classes predicted by the myself (labeled as 'true' labels) and the model, respectively. Each cell in the matrix shows the number of observations known to be in a given class and predicted by the model to be in a certain class. The diagonal cells from the top-left to the bottom-right represent correct predictions (true positives for each class), whereas the off-diagonal cells indicate incorrect predictions (false positives and false negatives). When evaluating performance with a multi-class (more than two category) classification problem, it is typical to compute statistics for each class, such as precision (the proportion of positive identifications that were actually correct), recall (the proportion of actual positives that were identified correctly), and accuracy.

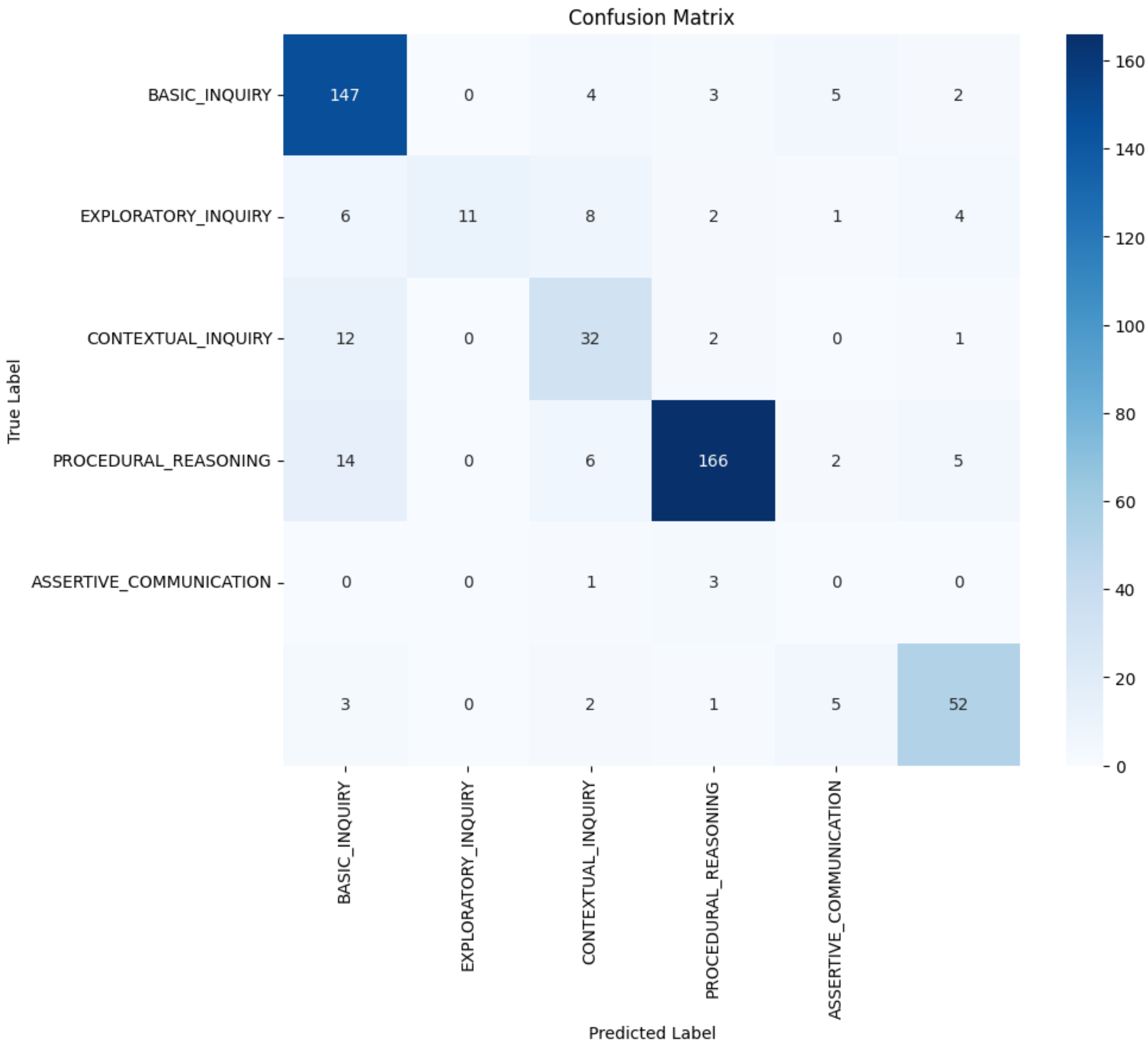


Figure 3.11: Confusion matrix representing the alignment between my labeling of the question types and the model's (GPT-3.5 Turbo)

3.5.2.1.3 Evaluation In establishing a foundational baseline for comparison, I utilized the ZeroR (Zero Rate) Classifier. This simple model operates on a straightforward principle: it predicts the outcome belonging to the most frequently occurring class in the dataset. When dealing with a binary classification task, the ZeroR approach could theoretically exceed a 50% accuracy rate simply by aligning its predictions with the most common outcome. For example, suppose we had a dataset of photographs of dogs and cats where 80% of the photos are of dogs. The ZeroR classifier would predict 'dog' every time, and be correct 80% of the time. In the case of this example, where we have 5 question type categories with unbalanced classes, in order for the model to be considered effective, it must surpass the performance benchmark set by the ZeroR Classifier. This comparison is critical, as it ensures that any observed predictive accuracy is not due to chance, but rather is indicative of the model's capability to distinguish between the different types of questions in the data.

Precision and recall are two valuable metrics used in the evaluation of classification models, offering detailed insights into the accuracy and completeness of a model's predictions for each category. Precision measures the proportion of correct positive identifications made by the model, reflecting its exactness. High precision indicates that the model reliably discerns the category in question with minimal false positives. This statistic matters most when the cost of a false positive is high, where a higher precision score indicates that the model is returning more relevant results than irrelevant ones. The formula for precision is as follows:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

Recall, on the other hand, measures the model's ability to correctly identify all of the actual positives in the data. Also known as sensitivity, it is defined as the number of true positives divided by the sum of true positives and false negatives. Recall is a measure of a classifier's completeness. A higher recall score indicates that the model returns most of the relevant results.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

While precision and recall provide insight into the performance of a classification model, evaluating a model in a multi-class classification scenario necessitates a more holistic approach. This is where the F1 score, and more specifically, the Macro Average F1 Score, become pertinent. The F1 score is the harmonic mean of precision and recall, offering a balance between the two by accounting for both false positives and false negatives. It is defined as:

$$F1\ score = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \tag{3.1}$$

In the case of a multi-class classification problem like the one in this study, where the classes are unbalanced, and the model's ability to correctly predict each class is equally important, I utilize the Macro Average F1 Score. It computes the F1 Score for each class independently and then takes the average of these scores. This averaging method does not take the class frequencies into account, which is particularly useful in our context, where some question types are more infrequent than others. Thus, the Macro Average F1 Score is calculated as follows:

$$\text{Macro Average F1 Score} = \frac{1}{K}\sum_{k=1}^{K} F1\ score\ for\ class\ k \tag{3.2}$$

The Macro Average F1 Score is important in this analysis as it provides a singular measure to assess the overall performance of our classification model across all question types. It is especially indicative of the model's efficacy in distinguishing between different types of questions, ensuring that performance is not biased towards the more prevalent classes.

Table 3.3: Category-wise Evaluation Metrics

	ZeroR		GPT-3.5 Turbo	
Category	Precision	Recall	Precision	Recall
Basic Inquiry	.395	1.0	.945	.869
Procedural Reasoning	0	0	.804	.933
Context Inquiry	0	0	.813	.881
Exploratory Inquiry	0	0	1.0	.375
Assertive Communication	0	0	.643	.706

For each category in the dataset, precision and recall have been calculated based on the model's predictions. Table 4.2 outlines these metrics for the five question types coded for in the data. In interpreting the classification efficacy of the GPT-3.5 Turbo model compared to the ZeroR baseline, a distinct improvement in the precision and recall metrics across all question categories is evident. The baseline model, adhering to the principle of majority class prediction, demonstrates a precision of 0.395 in the category of Basic Inquiry, coupled with a recall of 1.0. While this indicates a detection of all Basic Inquiries, it does so at the expense of a high false positive rate, as reflected by its lower precision. This inherent limitation of the ZeroR classifier shows its utility as a benchmark rather than a practical solution. Meanwhile, the GPT-3.5 Turbo model shows high precision of 0.945 in the Basic Inquiry category, suggesting that the vast majority of its predictions are accurate, while a recall of 0.869 points to a high rate of true positive identifications. This balance between precision and recall is indicative of the model's nuanced understanding of this question type. For Procedural Reasoning and Context Inquiry, the GPT-3.5 Turbo model again outperforms the baseline with impressive recall scores of 0.933 and 0.881, respectively. These figures suggest an astute recognition of relevant queries, albeit with a slight compromise in precision for Procedural Reasoning, potentially indicative of occasional misclassifications among similar question types. The Exploratory Inquiry category presents an interesting anomaly; the model achieves a perfect precision score, thus every instance classified as Exploratory Inquiry is correct. However, the recall of 0.375 reveals a shortfall in the model's sensitivity to this question type, possibly hinting at a more complex or less well-defined set of characteristics that govern this category, which the model has yet to fully learn. Assertive Communication sees a moderate performance from the GPT-3.5 Turbo model with a precision of 0.643 and recall of 0.706. These metrics suggest that while the model is reasonably adept at identifying Assertive Communications, there exists an avenue for refinement, particularly in reducing the false positive rate.

3.5.2.1.4 Overall Model Performance In the evaluation of classification models, it is beneficial to synthesize the performance metrics into a single, comprehensive measure that can be used to compare models. To this end, the Macro Average F1 Score is particularly advantageous in the context of multiclass classification problems with imbalanced classes. Table 3.4 presents the Macro-Precision, Macro-Recall, Accuracy, and Macro F1 for both the baseline ZeroR model and the more advanced GPT-3.5 Turbo model. The Macro Average F1 Score for the ZeroR model is approximately 0.113, which, given its methodology of predicting only the most frequent class, underscores the model's limited applicability in a multiclass setting. Conversely, the GPT-3.5 Turbo model achieves a Macro Average F1 Score of approximately 0.767, indicating a substantial improvement in overall classification performance. These scores are calculated by averaging the individual F1 scores for each category, which themselves are the harmonic mean of precision and recall. The Macro F1 thus provides an aggregate measure of each model's ability to correctly and consistently predict across all categories, offering a holistic indicator of performance.

Overall, the main takeaway from these these metrics is that the GPT-3.5 Turbo model not only outperformed the baseline on the predominant class but also was effectively able to distinguish between the less common categories (Table 3.4). The performance on this task was great, and in line with an expert human labeler, affirming its use as valuable analytical tool for being able to distinguish between question types.

Table 3.4: Model-level Evaluation Metrics

Model	Macro-Precision	Macro-Recall	Accuracy	Macro F1
ZeroR	.079	.200	.395	.113
GPT-3.5 Turbo	.841	.753	.843	.767

3.5.2.1.5 Classification results on a large sample of questions Given the reliability of GPT-3.5 Turbo to predict the types of questions being asked in the *MathConverse* dataset, in this section, I leverage the prompt to classify the question types on a much larger sample of the data. Shown below in Figure 3.12 are the results of the model prediction, which as we can compare with Figure 3.10, provides a very similar distribution when applied to the larger dataset. This provides some evidence of the ability for the model to perform well on this fairly low-inference task that can serve well in answering the next set of research questions, looking at patterns of how question types change over time (and by time of day) by using the large amounts of data I have been able to gather here.

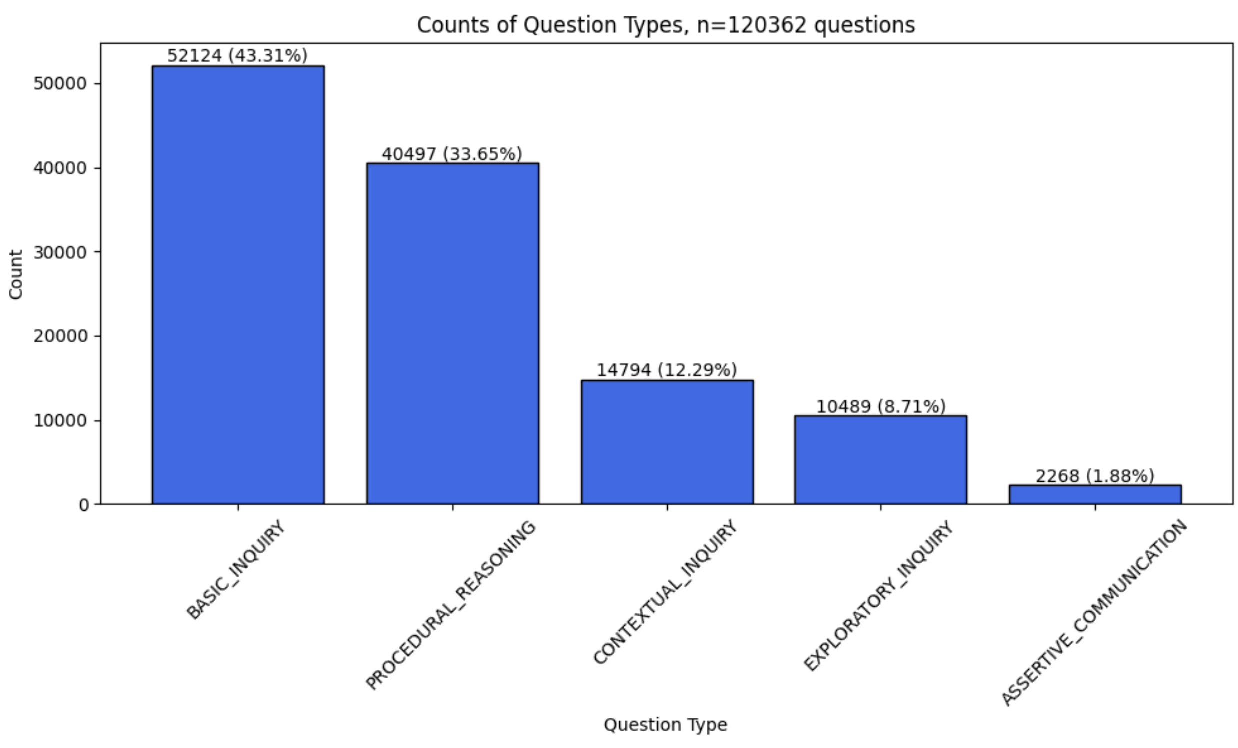


Figure 3.12: Model predicted question types on a large random sample of $n = 120,362$ questions

3.5.2.2 Topics and problem types addressed

In this section, I present the findings of the topic model produced by MALLET by identifying the labels I associated with each of the topics, the proportions of the documents with the associated labels, as well as the words with the highest weights for each topic. The topic model is a 30-topic model, so I show the top 15 topics here, and post the full 30-topic model results in the appendix.

I ran a number of topic models, and for this analysis, I show a topic model that I ran on only the language of the students (the one's asking the questions). Figure 3.13 shows the steps of this work, the main detail being that in each conversation, it is the student that initially starts the conversation, so the algorithm is able to pick out that the person with this ID is the 'student', while the other people in the conversation serve as helpers.

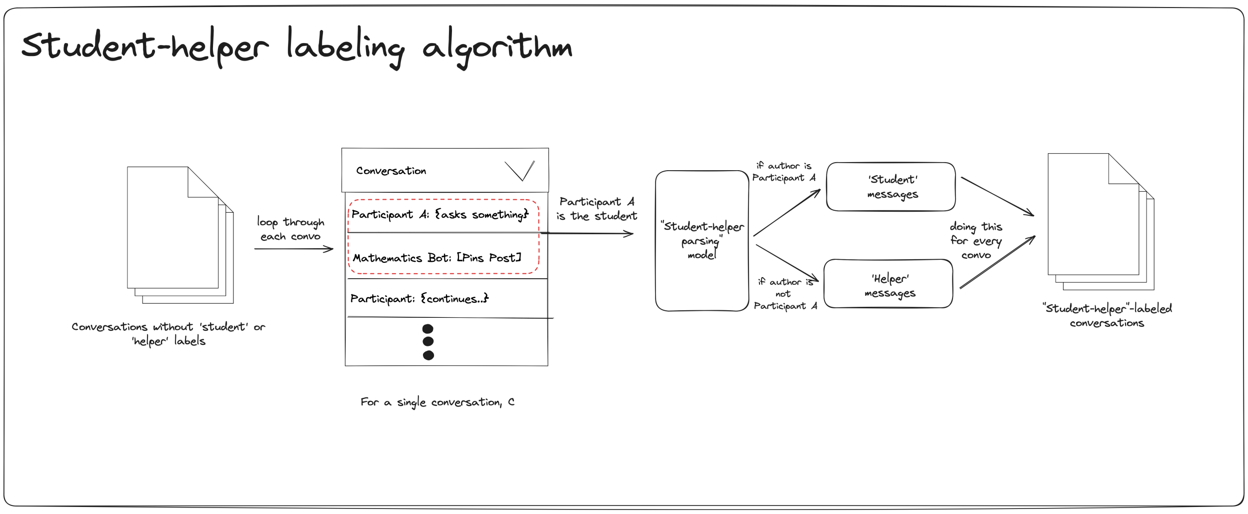


Figure 3.13: A schematic showing an overview of how I labeled the students and helpers in the dataset

3.5.2.3 Common topics

Topic Label	Proportion	Top Characteristic Words
General Inquiry	0.29455	help, close, helpers, someone, need, please, question, anyone, thanks, explain
Problem Clarification	0.28811	like, would, yeah, think, see, sense, right, makes, one, something
Answer Checking	0.22963	answer, wrong, got, close, correct, question, right, get, thanks, sure
Help and Formality	0.21241	thank, close, yes, okay, thanks, sorry, much, understand, get, help
Academic Assistance	0.19026	dont, know, like, get, understand, yes, thats, cant, idk, think
School Terms	0.13771	math, like, know, help, need, time, teacher, questions, school, good
Calculating	0.13564	get, right, multiply, would, like, side, denominator, first, factor, left
Algebraic	0.1114	equation, solve, find, solution, equations, value, quadratic, close, form, solutions
Casual Conversation	0.11079	wait, yes, lol, like, idk, bro, yea, right, got, lemme
Graph Analysis	0.08483	line, point, graph, points, find, equation, axis, slope, function, would
Function Behavior	0.08095	function, value, limit, infinity, domain, range, negative, find, positive, inf
Integrals and Derivatives	0.07118	integral, derivative, rule, function, use, integrate, chain, close, integration

Table 3.5: Summary of Top 12 Topics with Proportions from Topic Modeling

Table 3.5 shows that the 12 topics with the highest proportions relate to language that is agnostic of the mathematics content. This makes sense as when we look at a topics like 'General Inquiry', ‘Answer Checking’, ‘Help and Formality’ as we know that students will have a lot of language in their dialogue relating to asking questions (e.g., ‘can *someone help* me’), checking answers (‘does this *seem right*?’) and thanking the helper (e.g., ‘*thanks* for the *help*’, ‘.close’).

Looking to the subject-specific results, we see that the most common topics that emerge seem to be calculus-related. While I labeled the top two math topics as ‘Calculating’ and ‘Algebraic’, these are topics that I am not surprised show up as the top topics, as these techniques show up across mathematics domains. Students are expected to find *slopes*, work with *denominators* and *equations* from Algebra up through senior-level mathematics courses. The next two topics, which I label as ‘Function Behavior’ and ‘Integral and Derivatives’, have terms that show up more specifically in the Calculus I and Calculus II course (e.g., *derivative, integrate, chain, chain (rule), infinity*). The results from these can help filter out the conversations that pertain to the derivative, which can be useful for analysis in Chapter 4 in which I aim to look at the types of conceptions about the derivative that emerge when asking about derivative problems.

3.6 Discussion

In this chapter, I provided the first of two case studies using the *MathConverse* dataset constructed in Chapter 2 to characterize how participants engage on the MDS platform by looking into the nature of participant activity, the content of their interactions, and the connection between these two. The first set of findings reveal the platform to be very active and growing each month, with tens of thousands of messages being exchanged in the help channels each day as well as a noticable cyclical pattern that aligns with academic schedules and seasons, which suggest its connection as a supplementary educational resource for students. When looking at the findings related to how participants interact with one another, the wide-ranging interchange intervals, conversation lengths, and turns of talk further illustrate the MDS as a dynamic community of interest (Henri and Pudelko, 2003) that caters to a wide range of learner needs, from quick clarifications to deep, exploratory discussions. The observed patterns of engagement resonate elements of with Siemens’ (2005) of connectivism and Wenger-Trayner and Wenger-Trayner’s (2015) CoP frameworks, which show signs that the communities such as these can acts as spaces where learners can come in as they need to, connect with resources and learn how to work on problems regularly. This kind of learning ecosystem can help foster an adaptive kind of learning, important in this modern, digital age where knowledge is fluid and access to diverse perspectives enriches the learning experience.

CHAPTER 4

Studying Concceptions of the Derivative in MathConverse Using Large Language Models

In this chapter, I continue my work with the *MathConverse* dataset; however, I choose to zoom in on a specific mathematical concept that is discussed within the student-tutor dialogues. By focusing on the subset of conversations where the concept of the derivative comes up, I provide an illustration of how large language models and other various techniques from machine learning can be used to analyze conversations between students and tutors where the researcher chooses a concept of interest. The *MathConverse* dataset and the methodologies I use in this study represent a significant departure from the traditional approaches that educational researchers use to collect and analyze data to understand students' mathematical conceptions. Unlike research on traditional classroom data, which oftentimes grapples with challenges related to size, cost, and validity, *MathConverse* presents itself as a scalable, adaptable, and practical alternative. The work I describe here circumvents many of the limitations that typically hinder education research, thereby providing new ways to study student knowledge at scale.

Historically, mathematics education research has heavily relied on data from classroom recordings and transcripts. While these methods are valuable for extracting rich insights for how our students and educators think about mathematics, these sources of data often come with their challenges when it comes to examining them for research purposes; in terms of labor-intensiveness and privacy concerns, which in turn restrict their availability and the ability to share insights broadly. As a result, the progression and replication of research in mathematics education with respect to what we can do with conversational data has been notably constrained. The progression of research in machine learning, markedly, the ability to run and apply large language models on consumer hardware, marks a turning point in educational research. Methods from *natural language processing*, a subfield of machine learning concerned with giving computers the ability to 'understand' and analyze text (Hirschberg and Manning, 2015; Jurafsky, 2009) have enabled researchers across many academic disciplines to analyze the texts pertinent to their fields. in of student-tutor interactions, shedding light on how mathematical concepts, like derivatives, are comprehended and communicated in an online educational setting. The combination of using these 'text as data' methods alongside the dataset I have constructed from the MDS, *MathConverse*, offers a solution by harnessing the power of the innovations in the fields of machine learning and natural language processing intersecting with the transition to how students are learning mathematics in the current era to provide a more accessible and adaptable way to research the teaching and learning of mathematics. This study aims to leverage these methodological tools to uncover the nuances of learning and teaching derivatives, offering valuable insights that could significantly enhance educational strategies and student understanding in mathematics.

This chapter is structured to first provide a comprehensive literature review of how researchers in mathematics education have examined *what* students know about various concepts in mathematics, laying the groundwork for understanding the current state of research in both mathematics education and machine learning applications. Following this, the methodological framework employed in this study will be elaborately presented. The chapter will then dive deeper into the specific facets of derivative concepts that this research aims to explore, employing the use of large language models to analyze the conversational data from *MathConverse*. With this approach, the study aims to contribute significantly to the field of mathematics education, offering new perspectives and methodologies for understanding how students engage with mathematical concepts.

4.0.1 Research question

Given a set of conversations where students bring homework questions about derivative problems:

1. How do conversations about derivatives on MathConverse reflect or differ from Zandieh's (2000) theory of student conceptions of the derivative?

4.1 Background

4.1.1 A historical review of examining students' knowledge of mathematical concepts

In this historical overview, I first identify and differentiate theories addressing what students know from those that address how students know. Second, I highlight the various methodologies that mathematics educators have used to study conceptions and how these methods have contributed to the research on mathematical conceptions. The term "conception" will be defined more formally later in the paper, however, in this section, it is useful to think about a conception as a dynamic understanding of a mathematical concept that a learner forms through interaction with their learning environment. This understanding, which is shaped by the learner's actions and feedback, can vary in different contexts and tasks, reflecting the multifaceted nature of mathematical knowledge. I conclude this section by explaining how my proposed study fits and contributes to this area of research methodologically.

There have been two main perspectives used by mathematics education researchers in studying students' knowledge of mathematical concepts. Some researchers have chosen to focus on the meanings that can be ascribed to actions or behaviors that students engage in while completing mathematical tasks, such as examining responses to word problems, to investigate students' understanding of specific mathematical concepts (e.g., White and Mitchelmore, 1996). These studies possess an epistemological nature, as they implore into what students know. However, some researchers, inspired by the cognitive revolution in psychology during the early 1950s, have shifted their focused their focus to mental processes such as thinking, problem-solving, decision-making, and memory (e.g., Schoenfeld, 1983). Despite this, since direct access to one's mind is unattainable, these researchers also monitor students' actions and behaviors, albeit with a different objective. Adopting a more psychological perspective, these studies aim to understand how students come to know what they know, rather than what students know. To rephrase this distinction, epistemological research aims to understand the meanings (the 'what') attributed to students' processes (behavioral or mental), while cognitive research attempts to explain the processes a subject engages to construct meaning (the 'how'). These perspectives are not contradictory; instead, they represent two facets of the same coin, enabling a more holistic understanding of student conceptions. In this study, my objective is to explore 'what aspects of derivatives can one learn?' rather than 'how does one acquire this knowledge?'

Mathematics education researchers have primarily used four methods to inspect what students know about mathematical concepts: interviews, classroom observation studies, analysis of student work on problems, and textbook analysis, with interviews being the primary method of inquiry (Bingolbali and Monaghan, 2008; Clement et al., 1981). A notable example is Lamon's (1993) study of 24 sixth graders' conceptions of proportion and ratio. In this study, students were asked to think aloud while working on problems designed to help the researchers understand what the students know about proportion and ratio prior to instruction in these topics. The interview questions were developed using a framework comprising problems of four distinct semantic types. Students' strategies were then classified into two overarching categories: non-constructive and constructive. Each of these categories encompassed subcategories describing students' chosen approaches to the given problems (e.g., problem avoidance, pattern building, quantitative proportional reasoning). By identifying the strategies employed by students in each semantic type, the study provided insights into the various conceptions of ratio and proportion students may engage with while solving problems. For example, when comparing the cost of two items, the conception of unit rates (i.e., a ratio in which the denominator is 1) may emerge to help students in determining which item offers better value. Conversely, in problems with similar figures, setting up proportional expressions is key for identifying missing lengths or angles. Although interview studies like Lamon's yield rich data for understanding the different ways students can understand a mathematical concept, they are often resource-intensive and challenging to implement on a large scale.

Researchers have also studied conceptions by analyzing students' collaborative work through observations of classroom settings via collecting fieldnotes, video, or audio recordings with the goal of these studies is to understand how students develop conceptions as a collective. In one such study, Noble et al. (2006) recorded a teaching experiment involving a bilingual 11th and 12th grade algebra class. Their aim was to investigate how students used graphical tools during problem solving. The lesson specifically involved the use of drawing machines, designed to enable groups of students to generate shapes by controlling two parametric functions. Noble and colleague's analysis of classroom video revealed that the students' actions demonstrated a conception of circle as a set of parametric equations. This study exemplifies how conceptions (e.g., a conception of circle) can be discerned by outside observers in a collaborative activity between students, even when the students themselves might not be consciously aware of any conception of a circle beyond the shape they use to control their individual action. By examining conversations and collaborative work, classroom observation studies such as Noble et al.'s demonstrate that student conceptions can be identified in environments where students actively and collectively problem-solve. However, similar to interviews, these types of studies are labor intensive and challenging to execute on a large scale. This limits the extent to which the findings can be generalized or the range of situations one could observe the development of mathematical understandings in larger groups. The need for such large-scale observations arises from the desire to understand how mathematical conceptions develop not just at the individual level but also within larger social contexts.

Mathematics education researchers have also explored conceptions through analysis of students' written work to carefully crafted mathematics problems. This approach offers an efficient and cost-effective method for gathering students' responses to both open-ended and multiple-choice mathematical tasks. For instance, Ketterlin-Geller and Yovanoff (2009) used cognitive diagnostic assessments to explore students' conceptions of and operations on fractions. They suggested that well-validated cognitive diagnostic items can offer significant insights into which conceptions students have fully grasped or have yet to master. In another study that employed open-ended tasks, Davis and Vinner (1986) investigated students' conceptions of limits. Students who had completed an introductory calculus course were asked to write everything they knew about the limit of a sequence. The study identified nine distinct misconceptions of limit, each of which the researchers believed had a rational basis. Studies that analyze student work, such as these, have the advantage of being more scalable than interview-based and classroom observation methods. To ensure validity, researchers can replicate the administration of these assessments in different settings to confirm the results. However, this approach lacks the in-depth insights provided by methods like interviews and classroom observations, which allow for direct follow-up with students and more insight into what they are thinking about.

Although less direct than interviewing, observing, or testing students, textbook analysis offers another avenue for understanding student conceptions, as textbooks serve as "environment[s] for construction of knowledge" (Herbst, 1995, p.3) for students. Textbook analysis studies aim to understand what conceptions of a given concept (or set of concepts) are represented by certain elements of the textbook (e.g., solutions, exercises). This is often accomplished by collecting a sample of textbooks that meet specific criteria (i.e., country or region of interest, grade level, or subject of interest), compiling sets of textbook tasks (problems or examples) and their associated solutions, and coding them based on a concept framework. One such example is Mesa's (2004) examination of 35 secondary mathematics textbooks from 18 countries, which were chosen from the Third International Mathematics and Science Study (TIMSS) database. This analysis revealed the practices and contexts related to the concept of functions in these textbooks. Textbook analysis studies like Mesa's have provided a valuable framework for studying conceptions, as they are they can reveal student access to conceptions by providing insight into the curriculum and how conceptions are presented to students through their textbooks. Furthermore, large-scale analysis is possible, given the computational capacity to detect patterns in text with automated methods. However, textbook analysis is limited to what is available in the text and cannot provide a full picture of the dynamic nature of conceptions as students reason through problems.

As this dissertation study focuses on a tutoring environment as a site to study conceptions, it is important to compare this study to research done in these spaces. While there has been an ample amount of research devoted to understanding what goes on in these environments (e.g., Bloom, 1984; Fuchs et al., 1997), this line of research has historically been used in cognitive studies concentrated on program design and investigating what interventions work (i.e., examining the effectiveness of the learning intervention) rather than why they work (what are the conceptions that emerge in the conversations; see the literature review by Roscoe and Chi, 2007). The proposed study aims to contribute to the body of research on student conceptions (focusing on conceptions of the derivative) by investigating conversations about mathematics problems that take place between students and peer tutors in an online collaboration platform. While studies exist that examine the effectiveness of peer-tutoring on student understanding of mathematical concepts (see Roscoe & Chi, 2016), no study of peer-tutoring has specifically aimed to distinguish and classify on the content of the conversations between tutors and tutees. This study aims to fill a gap in the existing literature by examining conceptions of the derivative in online tutoring contexts.

4.1.2 Where this study fits methodologically

Like traditional classroom settings where researchers record interactions among students, teachers, mathematical tasks, and mathematical environments to uncover student conceptions at play, ample opportunities exist to observe how such conceptions arise when students have conversations about mathematical tasks within tutoring spaces with more-knowledgeable peers, who are closer in age and status to the students than their teachers. In this chapter, I aim to meticulously examine a particular mathematical concept, harnessing the potential of LLMs on a randomly selected collection of dialogues from an online mathematics learning forum. The primary task includes identifying and categorizing the interactions stemming from problem-solving situations within these dialogues. Instead of following traditional analysis methods, this dissertation adopts an approach rooted in 'one-shot prompting'. This strategy effectively uses a small set of examples (the 'prompts') to guide the LLM towards recognizing and identifying key concepts. By doing so, it greatly simplifies the process of analyzing an extensively large dataset. Hence, the core intent of this dissertation is to demonstrate the considerable potential of utilizing LLMs and carefully crafted prompts in understanding and dissecting students' comprehension of mathematical concepts, with an aim to show that these models can provide structured, reliable output. Overall, this study makes a unique contribution to the field by combining the study of student conceptions with large-scale data analysis, offering valuable insights into the various ways that students can come to know a complex mathematical concept.

I chose to focus on a concept from calculus (derivatives) for two main reasons. First, calculus is often the first college mathematics course that university students are required to take, which coincides with first year college students' adjustments to other aspects of college life, including less time spent in class and more time devoted to independent study (Moreno and Muller, 1999). In the U.S., K-12 students typically have more daily contact hours with their teachers than they do in college settings, leading to a shift in the student-teacher dynamic in college. First-year college students may be less aware of the resources available to them, such as office hours or in-person tutoring centers and may feel hesitant to talk to their instructors about the difficulties they are having with the course material. Online mathematics learning communities can provide these students virtual spaces to ask for help immediately after encountering a question and receive prompt feedback, which has been shown to encourage revision and improve performance (Roth et al., 2008). Calculus is often a course required by many first-year undergraduate students, and these online spaces are likely a place for these students to go to when they need help. My second reason for studying a concept from calculus is because of the ease of access to an immense number of conversations about calculus in online learning communities, as calculus is often the most active subject in them. In the next section, I review the research on student understanding of the specific calculus topic my study is concerned with—derivatives.

4.1.3 Research on student understanding of derivatives

Calculus is commonly viewed as the mathematics of change, measuring change via the two important concepts: the derivative and the integral (Stewart, 2012). These concepts are so fundamental to the study of calculus that the first two semesters of the university course are commonly referred to as differential and integral calculus, respectively. The concepts preceding differentiation—limits, continuity, and finding slopes of tangent lines—serve as stepping-stones for differentiation. Furthermore, the derivative concept has important implications in the real world; whenever we study any type of relationship, we want to know: how can we describe the change in one variable in reference to the change in another, related variable (Strogatz, 2019)?

This section analyzes the research on students' comprehension of the derivative, dividing it into two groups: investigations on the foundational concepts from prior coursework required for understanding the derivative and examinations of the derivative comprehension across different representations contexts.

4.1.3.1 Foundational concepts for derivative understanding

Early research in this area focused on identifying the difficulties students face when solving derivative problems and understanding the reasons behind their errors. In an interview study, Orton (1983) investigated the reactions of 110 students to problems involving differentiation and rate of change. Alongside algebraic errors, he found that students also struggled with rate of change and limit ideas. The main finding of the study is that the concept of ratio underlies the notion of rate of change, and to gain an understanding of the derivative, one must develop an understanding of slope, which is based on an understanding of ratio. Numerous studies that followed Orton's found comparable results; for instance, Ferrini-Mundy and Graham (1994) investigated student errors on differentiation tasks and discovered that students performed well when computing derivatives using formulas but struggled with understanding the core concepts of the derivative (i.e., ratio, limit, and function), as well as the ability to move between representations of the derivative.

4.1.3.2 Students' understanding of derivative across representations

The derivative is typically taught through four interconnected representations: graphical, verbal, symbolic, and physical. However, many students struggle to recognize or effectively utilize the connections between these different representations (Asiala et al., 1997; Hershkowitz et al., 2001). Among the four representations, mathematics education researchers have primarily concentrated their investigations on students' perceptions of the derivative as a slope. Research in this area has examined the factors that contribute to students having a stronger understanding of the derivative through graphical representations and has sought to establish theoretical frameworks that explain how students' understanding of the derivative develops through graphical representations. Asiala et al. (1997) investigated calculus students' understandings of a function and its derivative through a graphical perspective. The authors provide a theoretical account of the cognitive constructions necessary to develop graphical understanding of the derivative in terms of actions, processes, objects, and schemas (APOS), and designed a treatment with calculus students to elicit the formation of the mental constructions. Other studies that examined students' graphical understandings of the derivative include Baker et al. (2000) who also used elements of APOS theory to analyze students' comprehension of a challenging calculus graphing problem, Aspinwall et al. (1997) who conducted a case of a student to demonstrate how vivid and dynamic imagery invoked by calculus graphs can create unexpected obstacles to students' understandings, and most recently Vincent et al. (2015)'s interview study asking students to verbally describe a tangent line, sketch tangent lines for multiple curves, and apply tangent lines to multiple curves.

Other forms of representation and the connections between them have also been areas of research. Hähkiöniemi (2006) examined the types of representation students acquire when learning the derivative concept for the first time and found that students initially form perceptual representations, which allow them to understand the derivative as an object and apply differentiation rules to it. The study revealed that students experience difficulties in understanding the derivative in graphical representation and linking the limit of the difference quotient to other forms of representation.

4.1.3.3 Zandieh's framework of conceptions of derivative

Zandieh (2000) theoretical framework for examining students' understandings of the derivative has been frequently used by mathematics education researchers to structure studies pertaining to the derivative concept (e.g., Carlson et al., 2002; Feudel and Biehler, 2021; Hähkiöniemi, 2006; Likwambe and Christiansen, 2008; Roundy et al., 2015; Zandieh and Knapp, 2006). This model seeks to describe the mathematical community's concept image of the derivative. Zandieh developed the framework by considering how the derivative is used by various stakeholders in the mathematical community: mathematics textbooks, mathematicians, and mathematics graduate students.

	Contexts				
	Graphical	Verbal	Paradigmatic Physical	Symbolic	Other
Process-object layer	Slope	Rate	Velocity	Difference Quotient	
Ratio					
Limit					
Function					

Figure 4.1: Outline of the framework for the concept of derivative from Zandieh (2000)

The framework (Figure 4.1) is a matrix with columns representing contexts (i.e., representations) in which one may think about the concept of the derivative: (1) graphically, as a slope of a tangent line; (2) verbally, as rate of change; (3) physically, as a velocity in kinematic situations; (4) symbolically, as the limit of the difference quotient; and (5) other, as less commonly used contexts, such as numerical or other physical measurements than velocity.

The framework utilizes Sfard (1992) work on process-object layers, where processes are actions performed on pre-existing objects, and each process can be transformed, or reified into an object that can be acted upon by other processes. In other words, each layer describes the duality in which ratio, limit, and function play as either as dynamic process or as static objects. Although this framework does not provide any predictions on which understandings will emerge in what order and how, it is flexible enough to be used in a diverse number of settings and can organize a wide range of student understandings of derivatives. The rows of the matrix represent three process-object layers—ratio, limit, and function, which are viewed both as dynamic processes and as static objects, and "are linked in a chain" (Zandieh, 2000, p.107). The columns utilize the method of introducing new mathematical concepts by building one abstract object using one or multiple other abstract objects.

Within any of the contexts or columns, the ratio as a process takes two objects (e.g., distance and time, two lengths) and acts on them with division. When the ratio is reified into an object, it gets used in the next layer's process—the limit, where the limiting process involves passing through an infinite number of ratios to approach the value of the limit at one point. So if f is a function and $x = a$ is a value in the function's domain, then we can define the derivative of f at a point $x = a$, denoted $f'(a)$, by taking the limit of a ratio, $f'(a) = \lim_{h\to 0} \dfrac{f(a+h) - f(a)}{h}$, provided the limit exists. The limit is then also reified into an object as a quantity that gets used in the last layer's process (function) that involves going through (potentially) an infinite number of input values (x is in the domain of f), and for each one, determining an output value that is given by the limit of the difference quotient at that point. Finally, the derivative function itself ($f'(x) = \lim_{h\to 0} \dfrac{f(x+h) - f(x)}{h}$, for all x in f') can be reified as an object like any other function. This definition of the derivative, (function layer) is given in nearly every calculus textbook and involves a ratio, a limit, and a function, the three object-layer processes in the framework.

In summary, research on students' understandings of the derivative typically involve small-scale studies that aim to examine how students connect their understanding of the derivative to concepts of ratio, limit, and function as they perform derivative computations, as well as how students develop different understandings across representational contexts.

4.2 Theoretical Framework

In this section, I provide details of Balacheff and Gaudin (2009) conception model framework. After providing an account of how to define and model conceptions, I connect back to the derivative framework by Zandieh (2000), presenting a preliminary working theory of a way to study student conceptions of the derivative at scale.

4.2.1 Balacheff and Gaudin's conception model

In their paper describing their conception model, Balacheff and Gaudin (2002) begin with the epistemological problem of coherence in studying student knowledge. They note that Bourdieu (1990)'s concept of a sphere of practice can be a way to explain the phenomenon of a student holding a knowing of a mathematical concept that seems rational given their experiences and prior practice, yet contradictory from an observer's perspective. For Bourdieu (1990), individuals and groups have different social positions within a field, determined by their access to different forms of economic, social, and cultural capital. These positions in turn shape their experiences, perceptions, and actions within the field, and define the limits of their spheres of practice. A sphere of practice, in this sense, refers to the range of actions and opportunities that are available to an individual or group, based on their position within a field. It also includes the set of rules and expectations that govern their behavior within that field. Time is a vital component of spheres of practice, as the social positions, opportunities, and situations available to individuals and groups may be different over time. Additionally, the way that people understand and act within their sphere of practice can also change over time. This means that something that may have been considered rational and acceptable behavior within a sphere of practice at one point in time, may not be considered as such at another point in time. Given that the sphere of practice is shaped by the position of the individual or group within a specific field, and that position and opportunities are in a constant state of flux, it is possible for an observer to perceive something as contradictory that the actor or set of actors perceives as rational.

As an example, two distinct conceptions of the area of a rectangle can be considered within the context of students' understanding while problem-solving: (1) computational conception: the area as the product of the length (l) and the width (w), and (2) geometric conception: the area as the quantity of one unit by one unit squares enclosed by the rectangle with a length of l units and a width of w units. Note that specific problems are more effectively tackled using the first conception (for instance, calculating the area of a rectangle with dimensions 35 by 50) whereas others are better approached using the second conception (for example, determining the area of a blue region within a 5 by 3 rectangular grid, in which each unit square is either red or blue, with 8 red squares and the remaining squares as blue). This perspective highlights the importance of understanding and characterizing the various conceptions of fundamental mathematical concepts that may emerge in problem-solving situations across mathematics courses.

4.2.1.1 A definition of conception

According to Balacheff and Gaudin (2002), the term "conception" has been widely utilized as a tool in mathematics education research, but its implicit definition has made it challenging for researchers to effectively examine students' conceptions as objects of study (Artigue, 1989; Vinner, 1983). Different mathematical tasks, such as defining a term, describing a function, or simplifying an expression via symbolic manipulation, can reveal numerous ways of understanding a concept. For example, a student may understand the concept of the derivative well enough to use it in a problem-solving situation yet struggle to communicate it verbally or in writing. Similarly, a student might know the derivative as an instantaneous rate of change, but this knowing may or may not be sufficient to solve a problem requiring them to find the slope of a tangent line to a curve at a point. Recognizing that learners may have different models-in-action to mobilize what an outside observer might consider as the same piece of knowledge, Balacheff defined a conception as "the state of dynamic equilibrium of an action/feedback loop between a subject and a milieu under proscriptive constraints of viability" (Balacheff and Gaudin, 2009, p. 11). Considering Brousseau's (1997) theory of didactical situations, Balacheff and Gaudin opted for terms "subject" and "milieu" instead of "student" and "environment" in their definition. The subject denotes an individual, group, or entity in relation to a specific piece of knowledge, while the student encompasses both cognitive and non-cognitive aspects. The milieu refers to the relevant subset of the environment for learning

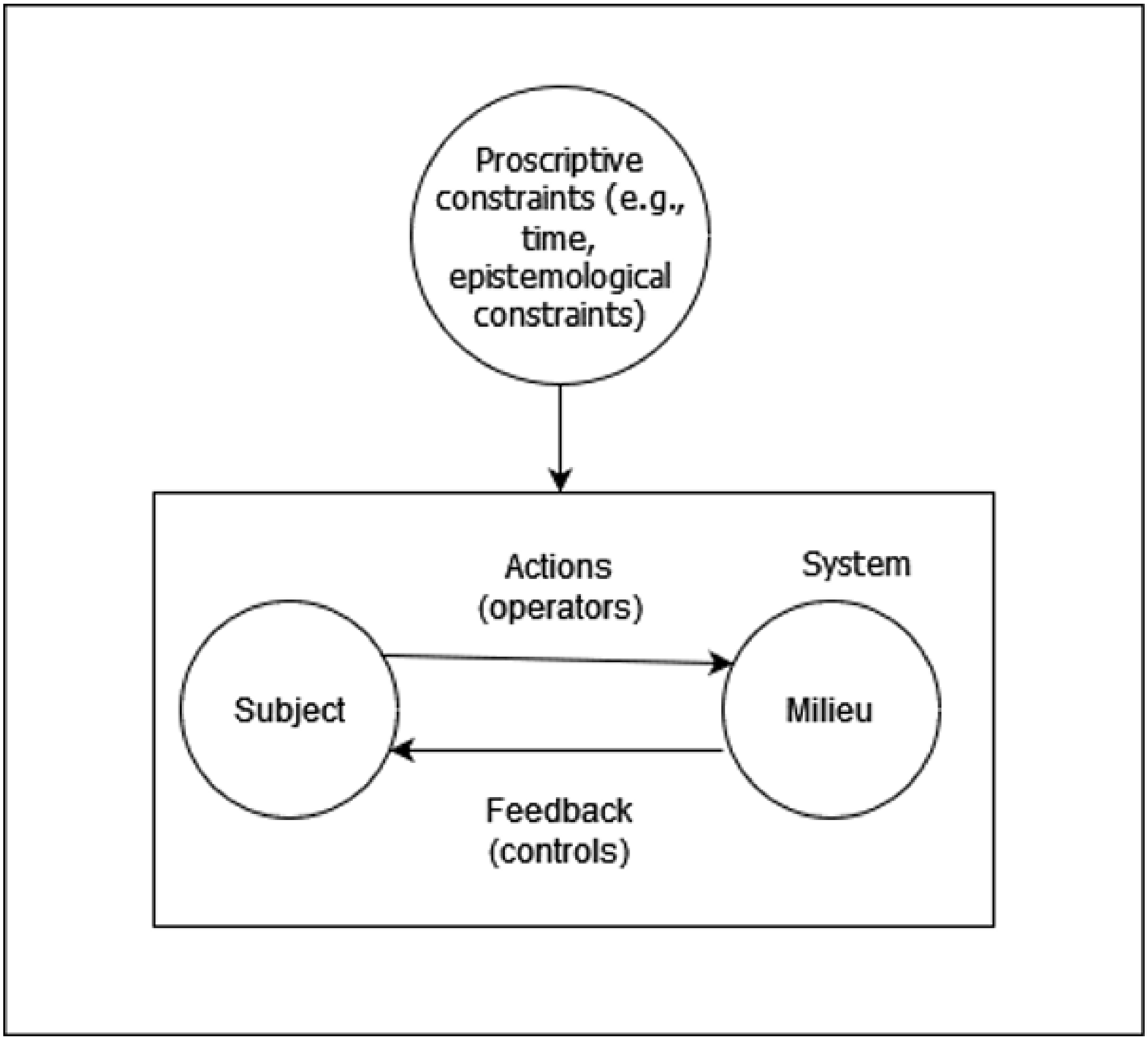


Figure 4.2: Subject-milieu system. Adapted from Balacheff and Gaudin (2003).

a particular concept, including both physical and symbolic interactions, while the environment represents the student's overall surroundings.

The subject-milieu system, adapted from Balacheff et al. (2003), illustrates the interaction between the subject and the milieu during problem-solving tasks. The subject's understanding of a concept is revealed through their actions (operators) and the feedback from the milieu (controls). This system emphasizes the need to consider the specific states of equilibrium achieved under proscriptive constraints of viability. Proscriptive constraints outline the necessary conditions for maintaining the system's viability by defining its boundaries. These limitations or restrictions don't provide direct instructions on maintaining equilibrium. Although the constraints are not exhaustively known, Balacheff and Gaudin identified two that are specific to didactical situations: time (e.g., school year organization, lesson planning) and epistemological constraints (i.e., existence of underlying reference knowledge for the content being taught). To effectively investigate conceptions as objects of inquiry, a conception should be regarded as a subject's local understanding of a concept within a specific situation. This understanding is in a state of continuous change, influenced by the subject's actions and the feedback obtained from the milieu. By examining conceptions, we can perceive problems as disruptions to the subject-milieu system's equilibrium and recognize the presence of a knowing through its manifestation as a problem-solving tool. Consequently, a knowing is not exclusively attributed to the subject or to the milieu but emerges as result from the interaction between the two. Learning can be understood as the process of restoring balance to the subject-milieu system's equilibrium following a perturbation that interrupts, changes, or disrupts this balance. A key indicator of this process is the discrepancy between a person's expectations and what an observer might view in the environment. An error exemplifies this phenomenon, when the subject fails to recognize the gap, but an external observer can identify it. The primary goal of studying conceptions is not to understand the thought processes of individual subject but to provide an account of the subject-milieu system.

4.2.2 From a definition to a model

To operationalize the conception definition in my study, I use Balacheff's conception quadruplet (P, R, L, Σ) of concept features (problems, operators, representations, and controls) as inspiration to help discern distinct derivative conceptions in dialogues between students and peer tutors addressing derivative problems (Balacheff and Gaudin, 2009; Balacheff, 2013). These features alongside Zandieh's (2000) can help understand how conceptions of a derivative are enacted in practice:

- P refers to a set of problems (or sphere of practice), with each problem $p \in P$ requires the use of a concept. Two solutions have been proposed to characterize P: (1) include all problems for which the conception provides efficient tools (Vergnaud, 1981, p.145) and (2) consider a finite set of problems from which other problems will derive (Brousseau, 1997). For the derivative concept, the first option to write out the set of all problems for which the derivative conception provides efficient tools is unfeasible, as there are potentially innumerable problems from every conception type which could meet the criteria. The second solution, coming up with a finite set of solutions, is more reasonable, yet comes with a flaw of not knowing rigorously if that set of problems can serve as a basis to generate all problems for that conception. A more pragmatic solution is to define a description of the set P to define the set of problems, refining this description as needed through the analytic process.
- R refers to a set of operators that students use to solve problems in P. Operators, as tools

of action, are the means to change the relationship between the subject and the milieu. In practice, the operators will be evident in the steps of the students' work. Pragmatically, identifying distinct conceptions of the derivative at scale involves deriving a description of the set of operators unique to solving $p \in P$, contingent on the duality of the representational context and the role one of ratio, limit, or function plays as a dynamic process (operator) on a pre-existing static object (control).

- L refers to the linguistic, graphical, or symbolic means (a.k.a., representations) that support the interaction between the subject and the milieu through actions, feedback, and the final answer or outcome. In this context, the representational nature of the problems (i.e., whether the problem elicits students to talk about the derivative in graphical, verbal, physical, or symbolic means) will act in duality with the problems to distinguish between the models of the conceptions.

- Σ is a control (or regulatory) structure, consisting of all the means required for the subject to use operators and the milieu to receive them, determine their adequacy and validity of the used operator, and select and provide feedback to the subject, which would determine whether the problem has been solved. The control also ensures that the conception C is not contradicted. This process is catalyzed by problems as tools to diagnose, reinforce a previously identified diagnosis of, question, destabilize, or reinforce conceptions. As the goal of the study is to model the conceptions of the derivative in a way that can: (1) be distinguishable across levels of knowing (as marked by the process-object layers), and (2) be concise enough to build a machine learning model that can distinguish conceptions via differences in the way students talk about problems solving $p \in P$ that is contingent on the duality of the representational context, I establish a pragmatic description of the set of controls that satisfies the following conditions. Finding distinct conceptions of the derivative at scale involves deriving a description of the set of controls unique to solving $p \in P$ that is contingent on the duality of the representational context, (1) if the operators are in the form of limit or function, then these controls are in the form of ratio and limit, respectively and serve to validate the operators as described above; (2) if the operator is in the form of ratio, then the controls are in the form of using one of: (a) the rise and run of a slope (graphical context; derivative as a slope); (b) the rate and time interval (verbal context; derivative as a rate of change); (c) the velocity and time interval (the physical context; derivative as a velocity), or (d) the change in y and the change in x (symbolic context; derivative as a difference quotient).

One conception falls outside Zandieh's framework (using rules to take a derivative), in which tables of derivative rules serve as a control here. In this study, I use these two frameworks alongside LLM prompting to identify distinct conceptions of the derivative as they arise in conversations collected between students and peer tutors discussing derivative problems. The goal is to initially identify the representations (the columns), and then prompt again for the process-object layer. In the next section, I show how I use this framework in combination with an adaptation of Zandieh's (2000) to model distinct conceptions the derivative based on whether students instantiate their knowings of these concepts of ratio, limit, and function as operators (i.e., what they appear to do in the steps of their work) or controls (i.e., what they take to be true while they do their work).

4.2.3 Conception models of the derivative

Table 2 presents the model for conception of derivative as instantaneous rate of change, describing each element in the quadruple. Next, I present the models of conceptions of derivative with ratio as operator, limit as operator and ratio as control, and function as operator and limit as control, in Figure 4.3, Figure 4.4, and Figure 4.5 respectively. In each table, a column represents a model of a conception of derivative. For example, in Figure 4.4, the second column represents the IRoC conception. This conception arises when limit is used as an operator and the problems are in the verbal representational context (i.e., the derivative is conceptualized verbally as a rate of change).

Table 4.1: Model of the Instantaneous Rate of Change (IRoC) Conception: limit as operator and verbal representation

Elements (notation)	**Description**
Problems (P_{IRoC})	Find the instantaneous rate of change of a function f at a point $x = a$.
Operators (R_{IRoC})	Determine the instantaneous rate of change of the function f at $x = a$ by considering the limit of the average rates of change as the interval $[a, b]$ shrinks to a single point.
Representation (L_{IRoC})	Verbal
Control Structure (Σ_{IRoC})	Use the average rate of change (computed using the ratio of the change in y-values to the change in x-values) to estimate the instantaneous rate of change at $x = a$.

4.3 Methodological Framework

In this chapter, my goal is to analyze a subset of the conversations where the students are asking questions about the derivative. To begin, I use the topic-modeling work done in chapter 3 to extract

Concept features \ Reps	Slope of Secant Line (SoSL)	Average Rate of Change (ARoC)	Average Velocity (AV)	Difference Quotient (DQ)
	Graphical (as a slope)	Verbal (as a rate of change)	Physical (as a velocity)	Symbolical (as a difference quotient)
Problems	Given the graph of a function *f*, find the slope of the secant line between the points corresponding to the interval [*a*, *b*]	Find the average rate of change of a function *f* over the interval [*a*, *b*].	Given a position function *s*(*t*) describing the movement of an object, find the average velocity of the object over the time interval [*a*, *b*].	Given a function *f*, find the difference quotient (*f*(*b*) - *f*(*a*)) / (*b* - *a*) to determine the average rate of change of the function over the interval [*a*, *b*].
Operators	Compute the slope of the secant line by dividing the change in the *y*-values by the change in the *x*-values, using the points on the graph that correspond to the interval [*a*, *b*].	Calculate the average rate of change of the function *f* over the interval [*a*, *b*] by finding the difference quotient (*f*(*b*) - *f*(*a*)) / (*b* - *a*).	Determine the average velocity of the object over the time interval [*a*, *b*] by computing the change in position divided by the change in time using the position function *s*(*t*).	Calculate the difference quotient (*f*(*b*) - *f*(*a*)) / (*b* - *a*) to find the average rate of change of the function *f* over the interval [*a*, *b*].
Controls	The coordinates of the points on the graph that correspond to the interval [*a*, *b*] are used to compute the change in y-values (Δ*y*) and the change in *x*-values (Δ*x*).	Use the values of *f*(*a*) and *f*(*b*) to compute the change in *y*-values (Δ*y* = *f*(*b*) - *f*(*a*)) and the change in *x*-values (Δ*x* = *b* - *a*).	The change in position (Δ*s* = *s*(*b*) - *s*(*a*)) and the change in time (Δ*t* = *b* - *a*) are used to find the position function s(t) and the given time interval [*a*, *b*].	The function *f* and the interval [*a*, *b*] are used to compute the change in *y*-values (Δ*y* = *f*(*b*) - *f*(*a*)) and the change in x-values (Δ*x* = *b* - *a*).

Figure 4.3: Four conceptions of the derivative when ratio acts as operator: Layer 1

Concept features \ Reps	Slope of Tangent Line (SoTL)	Instantaneous Rate of Change (IRoC)	Instantaneous Velocity (IV)	Limit of the Difference Quotient (LoDQ)
	Graphically (as a slope)	Verbally (as a rate of change)	Physically (as a velocity)	Symbolically (as a difference quotient)
Problems	Find the instantaneous rate of change of a function *f*, at a point *x*=*a*, given the graph of the function.	Find the instantaneous rate of change of a function *f* at a point *x*=*a*.	Given a position function *s*(*t*), find the instantaneous velocity of the object at time *t*=*a*.	Given a function *f*, find the derivative $f'(x)$ at *x*=*a* using the limit definition of the derivative.
Operators	Estimate the slope of the tangent line at the point *x*=*a* on the graph of the function *f* by considering the limit of the slopes of the secant lines as the interval [*a*, *b*] shrinks to a single point.	Determine the instantaneous rate of change of the function *f* at *x*=*a* by considering the limit of the average rates of change as the interval [*a*, *b*] shrinks to a single point.	Compute the instantaneous velocity of the object at time *t*=*a* by considering the limit of the average velocities as the time interval [*a*, *b*] shrinks to a single point.	Calculate the derivative $f(x)$ at *x*=*a* using the limit definition of the derivative: $\lim_{h \to 0} \frac{f(a+h)-f(a)}{h}$
Controls	Slopes of the secant lines (computed as the ratio of the change in *y*-values to the change in *x*-values) is used to estimate the slope of the tangent line.	Average rate of change (computed using the ratio of the change in *y*-values to the change in *x*-values) is used to estimate the instantaneous rate of change at *x*=*a*.	Average velocities (computed using the ratio of the change in position to the change in time) is used to estimate the instantaneous velocity of the object at time *t*=*a*.	The difference quotient (*f*(*a*+*h*) - *f*(*a*))/*h*, computed as the ratio of the change in *y*-values to the change in *x*-values is used to evaluate the limit definition of the derivative at *x*=*a*.

Figure 4.4: Four conceptions of the derivative when limit acts as operator and ratio acts as control: Layer 2

Reps / Concept features	Graph of Derivative Function (GoDF)	Rate of Change of Function (RoCF)	Velocity as a Function of Time (VoFoT)	Derivative as a Function (DaF)
	Graphically (as a slope)	Verbally (as a rate of change)	Physically (as a velocity)	Symbolically (as a difference quotient)
Problems	Given the graph of a function f and its derivative $f'(x)$, analyze the behavior of f in terms of its instantaneous rate of change.	Describe the behavior of a function f based on its instantaneous rate of change, given its derivative $f'(x)$.	Given a position function $s(t)$, find the velocity function $v(t)$ by finding its instantaneous rate of change using the derivative $s'(t)$.	Using the limit definition, find the derivative of a function f
Operators	Use the graph of the derivative $f'(x)$ to identify the behavior of the function f, such as increasing or decreasing intervals, critical points, and inflection points.	Interpret the derivative $f'(x)$ in terms of instantaneous rate of change and describe the behavior of f, including increasing/decreasing intervals and critical points.	Analyze the velocity function $v(t)=s'(t)$, to describe the movement of the object in terms of its instantaneous velocity, including changes in direction or acceleration.	Extend limit definition at a point to any arbitrary point x. $f'(x) = \lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$, given the limit exists.
Controls	The limit of the slopes of the secant lines approaching $x=a$ is the slope of the tangent line at a, which can be computed as $f'(a)$	The instantaneous rate of change at a point $x=a$ is computed as $f'(a)$.	The instantaneous velocity at a point $x=a$ is computed as $s'(a)$.	The derivative of a function f at a point a can be found by taking the limit as h approaches 0 of the difference quotient $f'(a) = \lim_{h\to 0} \frac{f(a+h)-f(a)}{h}$

Figure 4.5: Four conceptions of the derivative when function acts as operator and limit acts as control: Layer 3

these conversations ($N = 2088$). Figure 4.6 provides a schematic of this work. Once this is done, the next step is to see whether the conceptions described by Zandieh's (2000) derivative framework show up in the conversations.

4.3.1 Classification schema

To answer RQ 1, which asks whether the conversations about derivatives on MathConverse reflect or differ from Zandieh's (2000) theory of student conceptions of the derivative, I use the Figure 4.3, Figure 4.4, and Figure 4.5 as a guide so that for each conversation, I will first do some pre-processing of the conversations that will do some 'feature extraction', where I will use the language written in these figures to extract key phrases from the conversations and put them in a separate column if they exist. If they do, these key phrases get added onto the conversation with an indicator in the prompt. To do this, I use *regular expressions*, (typically shortened as regex), a tool used to specify a match pattern in text.

This coding process follows a very similar technique to Table 3.1. The difference here is that instead of classifying messages, I am instead classifying *conversations*, by concatenating all of the messages in a conversation by their conversation ids, as created by the algorithm in Chapter 2. Before running the messages through the model, I label a random sample of 500 of the conversations for the presence of the 12 derivative conceptions found in the columns of Figure 4.3, Figure 4.4, and

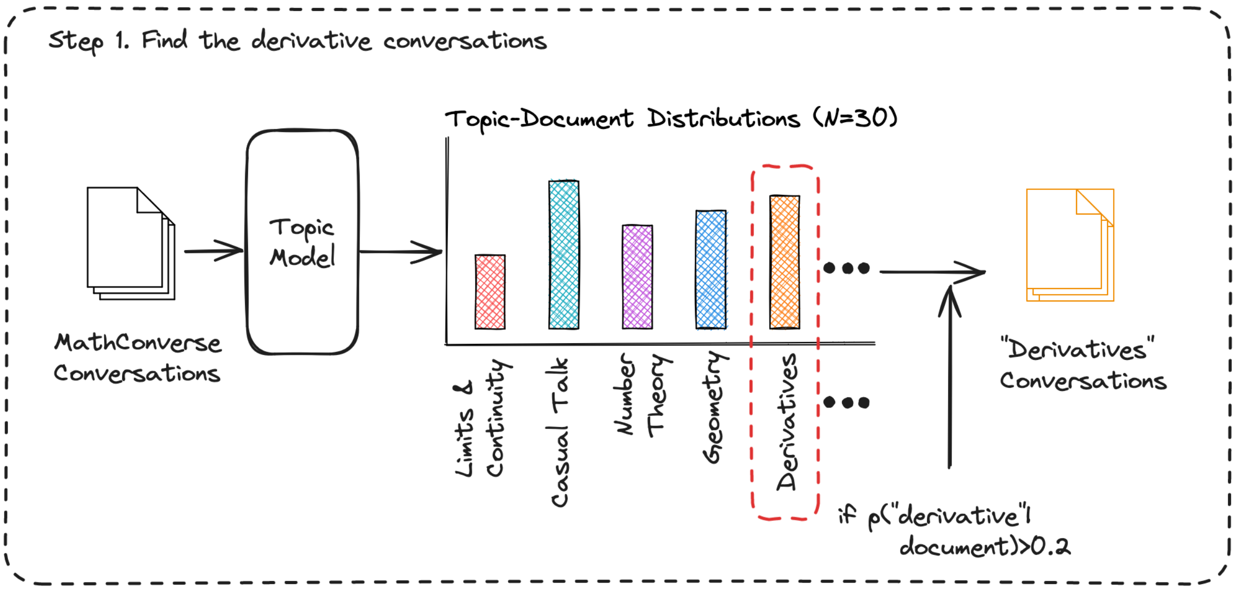


Figure 4.6: Identifying derivative-based conversations

Figure 4.5, with the addition of the 13th conception below:

Applying Rules to take a Derivative The process of finding the derivative of a function by applying standard rules and formulas. Using a table of derivative rules to find the derivative of a given function f without resorting to the limit definition.

The Applying Rules to take a Derivative (ARttaD) conception was also noted by Zandieh (2000) in her interview study as the most common student conception with her students, but said that it did not fit into the object-process framework, and that this form of instrumental thinking, while very prevalent, is not at the root of the type of conceptual thinking we are trying to get our students to understand about the derivative from our courses. As procedural questions was a question type in Chapter 3, and through my experiences in the platform and as a calculus educator is the type of calculus knowledge students tend to walk away with, my goal with this work is to try to see if in the conversations we see any of the other 12 derivative conceptions emerging.

After hand-labeling the conceptions the derivative conversations are processed through the OpenAI API using gpt-3.5-turbo. In the work by Wang et al. (2023), their team used a number of LLMs to benchmark three step-by-step mathematics tasks. Given a piece of student work with an error, their goal was to use the models to: (1) infer the type of student error, (2) determine the strategy to address the error, and (3) generate a response that incorporates that information. Using the benchmarks, they found that the OpenAI gpt-3.5 and gpt-4 models were far superior to the

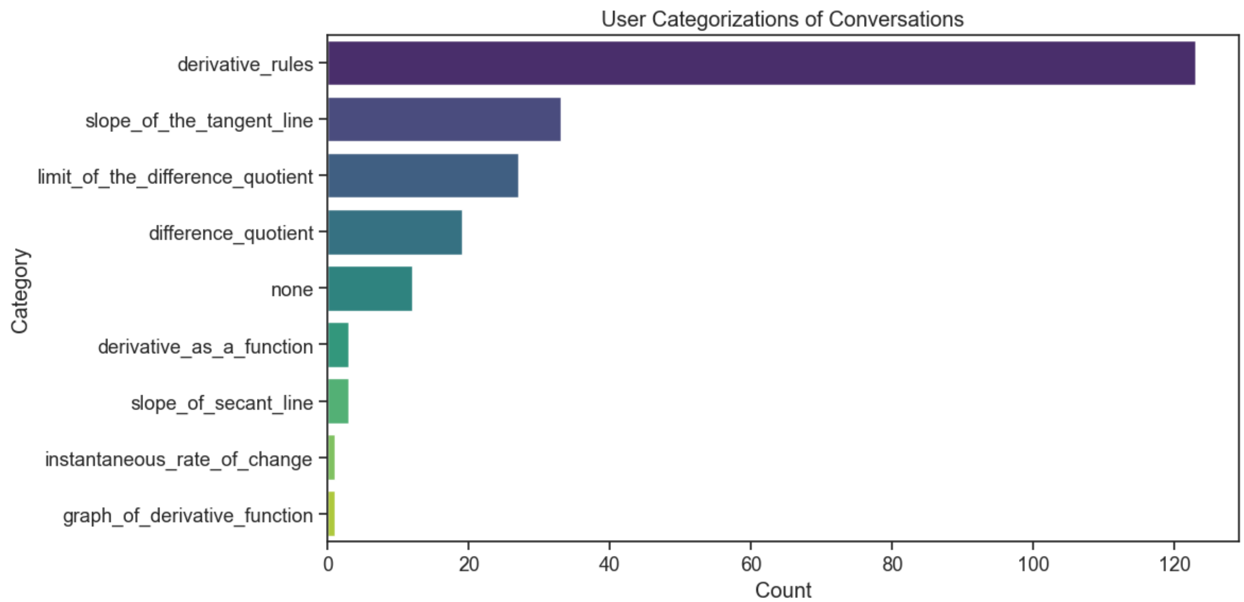


Figure 4.7: Hand-labeled categorizations of presence of conceptions in student messages in the conversations. $n = 238$ conversations

fine-tuned open source models, so I use these findings in their similar research and follow their lead to use the state of the art models. I choose to use gpt-3.5 as it is much more cost-effective (at the time of writing, more than 10x cheaper to run the models). As I am using a pre-trained model, the goal is to write a prompt that can 'teach' the model what it needs to do to perform well on the task each time it sees an input, and attach this prompt to the relevant information from the conversation (Figure 3.2). The model is hosted on an API, so in this case, I am not doing any training of a machine learning model, rather, I am using a very powerful model for model inference. For the conversation data, I am using the student-extracted data from the conversation (see Figure 3.13 in the previous chapter for more details) for this task.

4.4 Findings

4.4.1 RQ 1: What are the conceptions of the derivative that emerge, and how do they compare to Zandieh's (2000) framework?

4.4.1.0.1 Hand-labeled results Similar to the work of Chapter 3, in order to assess how well the model performs on this task, I first needed to hand label a random sample of 250 conversations having to do with the derivative by the topic modeling analysis done in subsubsection 3.5.2.2 filtered by the messages sent by students used Label Studio to label these for the presence of the conceptions. Unlike the work of the fine-tuned question classifier, I am not sure how representative these conversations are of how students talk about the derivative as a whole in the server. This was the label that I assigned to this topic, having to do with the fact that words like ‘chain rule’, ‘derivative’, and ‘derive’ were more likely to come up in these conversation than others, but as I was reading through these messages, I became more aware of the fact that I might be looking at a somewhat biased sample based on that technique of filtering that I did. With that in mind, in **??**, I provide a bar chart of counts of the questions types of the random sample that became the validation set to compare the model results to.

As shown in this chart, using the coding framework I described in subsection 4.2.3, I found that a substantial proportion of the students’ messages in these conversations had language describing the conception that I describe as applying rules to take a derivative. While this result is not entirely surprising, as it is common in the literature and in our coursework that a substantial portion of our calculus coursework details students working through derivative calculations, it was a little surprising to see just how many of the conversations were so similar in this way. The second most common problem type (and conception) was the slope of the tangent line, which aligns with a graphical representation of thinking about slope. In these problems, students typically were asked to find a tangent line or would even bring up specifically that they thought about the derivative as a tangent line.

Recall from Chapter 3, a confusion matrix provides a visual representation of the performance of a classification model, where each row represents the categories I have assigned, while the columns represent the categories assigned by the model, in this case, GPT-3.5-Turbo. Looking at Figure 4.8, there is fairly strong agreement on the derivative rules category, where out of the 93 times the model labeled an input text as ‘derivative rules’, 82 of those I agreed with, and 11 I marked ‘none’, meaning I did not see any evidence of conceptions of the derivative showing up in the message. While the slope of the tangent line was the second most prevalent category it was also the one with the most disagreement between myself and the model, where the model labeled 17 examples that I saw as ‘derivative rules’ as ‘slope of tangent line’. I looked more carefully at all of these and saw no language of slope of tangent line anywhere in any of these examples, so it is hard to know why the model made these predictions. In the next section, I go over some evaluation metrics.

4.4.1.0.2 Evaluation In order to set a benchmark for comparison, I again used the ZeroR (Zero Rate) Classifier, and precision and recall to establish the accuracy of the model against my manually-labeled results (see paragraph 3.5.2.1.3 for expanded discussion on the ZeroR model, as well as precision and recall definitions).

Table 4.2 presents the precision and recall figures for each category coded within this sample of

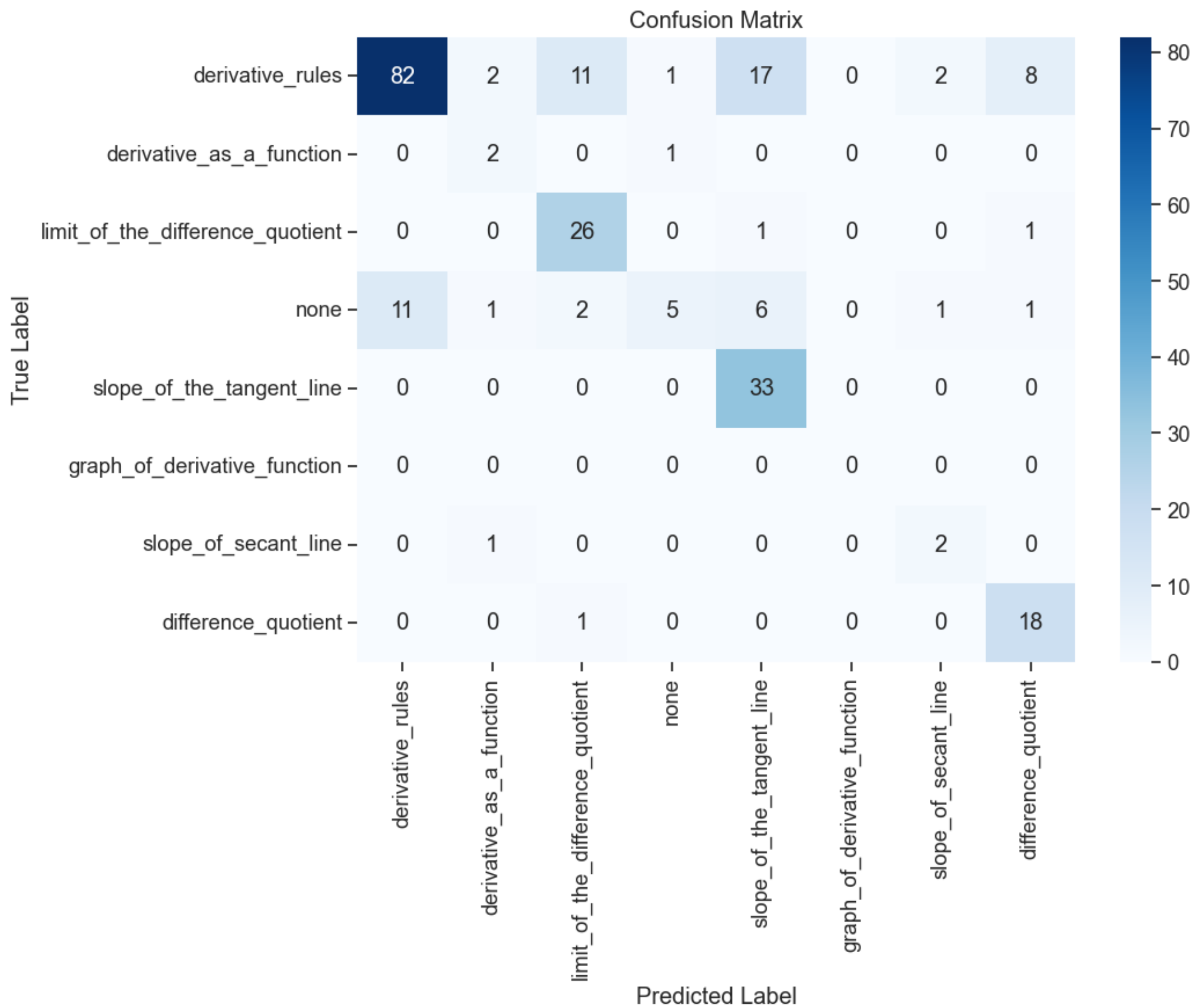


Figure 4.8: Confusion matrix representing the alignment between my labeling of the conceptions and the model's (GPT-3.5 Turbo)

Table 4.2: Category-wise Evaluation Metrics

Category	Count	ZeroR		GPT-3.5 Turbo	
		Precision	Recall	Precision	Recall
Derivative Rules	123	0.519	1.000	0.882	0.667
Derivative As A Function	3	0.000	0.000	0.333	0.667
Limit Of The Difference Quotient	28	0.000	0.000	0.650	0.929
None	27	0.000	0.000	0.714	0.185
Slope Of The Tangent Line	33	0.000	0.000	0.579	1.000
Graph Of Derivative Function	1	0.000	0.000	0.000	0.000
Slope Of Secant Line	3	0.000	0.000	0.400	0.667
Difference Quotient	19	0.000	0.000	0.643	0.947

the dataset. When comparing the GPT-3.5 Turbo model's performance to the ZeroR baseline, it is important to consider the trade-offs between precision and recall. The ZeroR classifier, which predicts based on the majority class, in this case, Derivative Rules, shows a precision of 0.519 with a recall of 1.0, which denotes being able to capture all instances within this category at the cost of a large number of false positives, a byproduct of its design. In contrast, the GPT-3.5 Turbo model shows high precision of 0.882 in the Derivative Rules and a recall of 0.667 which reflects that it can accuratly classify a majority of the instances of this case, but also capture of substantial proportion of the true positives. In cases where there are very low counts, it is hard to provide any conclusions about performance here, although as I'll discuss in the discussion, I would want to change the input and run more examples to see how well it performs on these categories.

Categories such as Limit of the Difference Quotient and Difference Quotient have high recall values of 0.929 and 0.947 respectively, demontrating the models ability in identifying relevant instances. However, the precision scores of 0.650 and 0.643 suggest that there are some misclassifications happening that might be caused by some similarity in question types or not being able to distinguish between some of the defining features of the conceptions. Further, the 'none' category is one that stands out with a very low recall of 0.185, indicative of the model struggling to correctly identify the instances that do not fall into any of the predefined categories. My hypothesis here is that my prompt might not have been clear enough that 'none' was an option, as looking at the results of Figure 4.8 show that it only chose this option 2.9% of the time which is not representative from what I saw in this sample.

4.5 Discussion

Building upon the analysis presented in Chapter 3, this chapter presented another case study of analysis using the *MathConverse* dataset, with a specific focus on the mathematical concept of the derivative within student-tutor dialogues. I presented a novel exploration on the use of LLMs for investigating mathematical conversations, highlighting the benefits of using text-as-data methods to understand mathematical dialogues. This chapter pivoted in its approach, moving from a more general look at patterns of engagement and question-asking, to a concept-specific task, looking at how students are engaging with the derivative in the help channels. Through the use of state-of-the-art language models such as OpenAI's GPT-3.5 Turbo, I provided an example of a data analysis effort that side-stepped sample-size, cost, and replicability issues that often challenge more traditional classroom data analysis.

My findings show the prevalence of procedural knowledge in the form of applying rules to take derivatives, which from my experiences as a calculus educator, resonates with the experiences many students have with the course. However, about one-third of the random sample of 238 responses were jointly labeled by myself and the model with three of the conceptions found in Zandieh's (2000) model: Limit of the Difference Quotient (26), Slope of the Tangent Line (33), and Difference Quotient (18). This provides some evidence that: (1) there are these other conceptions of the derivative emerging in the conversations, and (2) the LLM is able to pick up on these somewhat reliably. However, if the choice is to use an off-the-shelf encoder-based generative LLM like GPT-3.5, there is much work to be done on what data should be fed in, and what prompt should be used in order to make sure the model chooses the right category.

The use of LLMs in this study can help start conversations about their future role in educational research, and where we as education researchers fit in the picture of how they are being used in mathematics classrooms and to build intelligent tutoring systems. Much of this work is being done in rooms without those with training in education, and I believe it is important for those who study the teaching and learning of mathematics to be involved with this work, and understand the strengths and limitations when it comes to what these models can do.

CHAPTER 5

Conclusion

5.1 Introduction

The aim of my dissertation work is to provide the research community insights into the kinds of activity that takes places in online communities dedicated to the subjects that engage us as academics and educators, whether through our research or our teaching. Moreover, I argue that in order to better understand the phenomena at the heart of our studies as social science researchers, methods from machine learning and NLP can effectively unite the most valuable elements from the quantitative and qualitative research methodologies that often are used in isolation in our field (Grimmer et al., 2022; Shaffer, 2017). In this chapter, I revisit the overarching research questions introduced in Chapter 1, reflecting on the key findings and lessons gleaned throughout the course of the study. To conclude, I discuss the limitations and offer recommendations for future research.

5.2 Revisiting the Research Questions

When writing the three overarching research questions presented in Chapter 1, my goal was to designate one to correspond to the focus of the respective chapters, while also allowing for more deeper, focused research questions to evolve in each chapter. Chapter 2 serves more as a foundation for the work done in chapters three and four and is less of an empirical study, so the question I pose for this chapter is broad and exploratory. In contrast, the other two questions link closer to the empirical findings sections found in their respective chapters.

5.2.1 Transforming conversations into data

1. What processes are involved in converting large amounts of mathematical conversations into a structured dataset for analysis using text-as-data methods?

In Chapter 1, I posed this question as I way to encompass the efforts implemented across chapters two through four. While the writing in these chapters was able to provide some of these processes generally involved in this work, and how they were implemented in my particular use case with the conversations from MDS, I would also like to provide some more takeaways that can extend to studies beyond this project.

5.2.1.0.1 Data collection and initial processing The initial step in my project involved using the Discord Chat Exporter (Holub, 2023) to gather conversations from the MDS. This extraction process happened over multiple iterations, with a final download in January 2023. These messages were exported in JSON format, which offer a hierarchically-structured representation of the conversations. This format provides easy access to the metadata for each message, including message IDs, timestamps, message content, attachments, amongst other relevant information. In building research studies like this one, it is useful to assess the availability of APIs to access data at scale. Additionally, it is important to note how the data is formatted when you download it, formats available for download, and ease of access for cleaning and analysis, as well as finding ways to store the data that can collaborative access and version control can help ensure data integrity, reproducibility, and save on time and resources.

5.2.1.0.2 Building datasets for public use is an active field of research Before getting involved in the fields of computational social science, machine learning, and NLP, I was largely unaware of the extensive effort, time, and resources dedicated to the creation, development, and dissemination of datasets for public use. As described in subsection 2.2.3, the process of putting together high-quality datasets is critical not only for advancing research and practice in mathematics education but also for driving progress and innovation across a wide array of domains beyond education.

In some ways, these datasets can help augment the work we do in our more traditional forms of doing research on teaching and learning; that is, via video, audio, and transcript recordings of classroom activities. However, as noted by Major and Watson (2018) and Kim et al. (2023), these sources of data come with some limitations regarding replicability, privacy concerns, and practicality of data sharing. In response to some of these challenges, projects like the "Million Tutor Moves Observatories Project" (Reich et al., 2023), the work of Demszky et al. (2021) on measuring teachers' conversational uptake, and Suresh et al.'s (2022) work on building the TalkMoves dataset have shown promise in this area in their efforts to collect large-scale machine-readable datasets that could in turn be used to train intelligent tutoring systems.

The transformation of the large number of exchanges from the help channels from the MDS into the *MathConverse* dataset, as detailed in Chapter 2, represents a significant contribution in our field. It provides a ready to use dataset for a number of questions related to engagement patterns, the dynamics of peer and mentor-student interactions, and the development of how students are seeking guidance in this non-traditional online context.

5.2.2 Understanding engagement in an online learning community

2. What are the characteristics of participant engagement and conversational dynamics within the MDS?

5.2.2.0.1 Defining what it means to engage Before getting into the key findings of this chapter, I would like to discuss a few analyses that, while not directly included in this study, were important to the ways in which I thought about the community and aided in my development of key definitions. Over the past few years, a typical day of research for me would begin with visiting the MDS and observing tutoring sessions as they happened live. One aspect I noted from the onset of my research was the presence of a group of community members who were consistently active, almost round-the-clock, in various server discussions, and these individuals had been granted 'helper' roles—a badge of recognition within the community. The array of helper roles ranged from 'very helpful' to 'extremely helpful', with the most active participants bearing titles that reflected their frequent and substantial contributions to the help channels. These roles were not only symbolic but served as a motivating factor for engagement within the community.

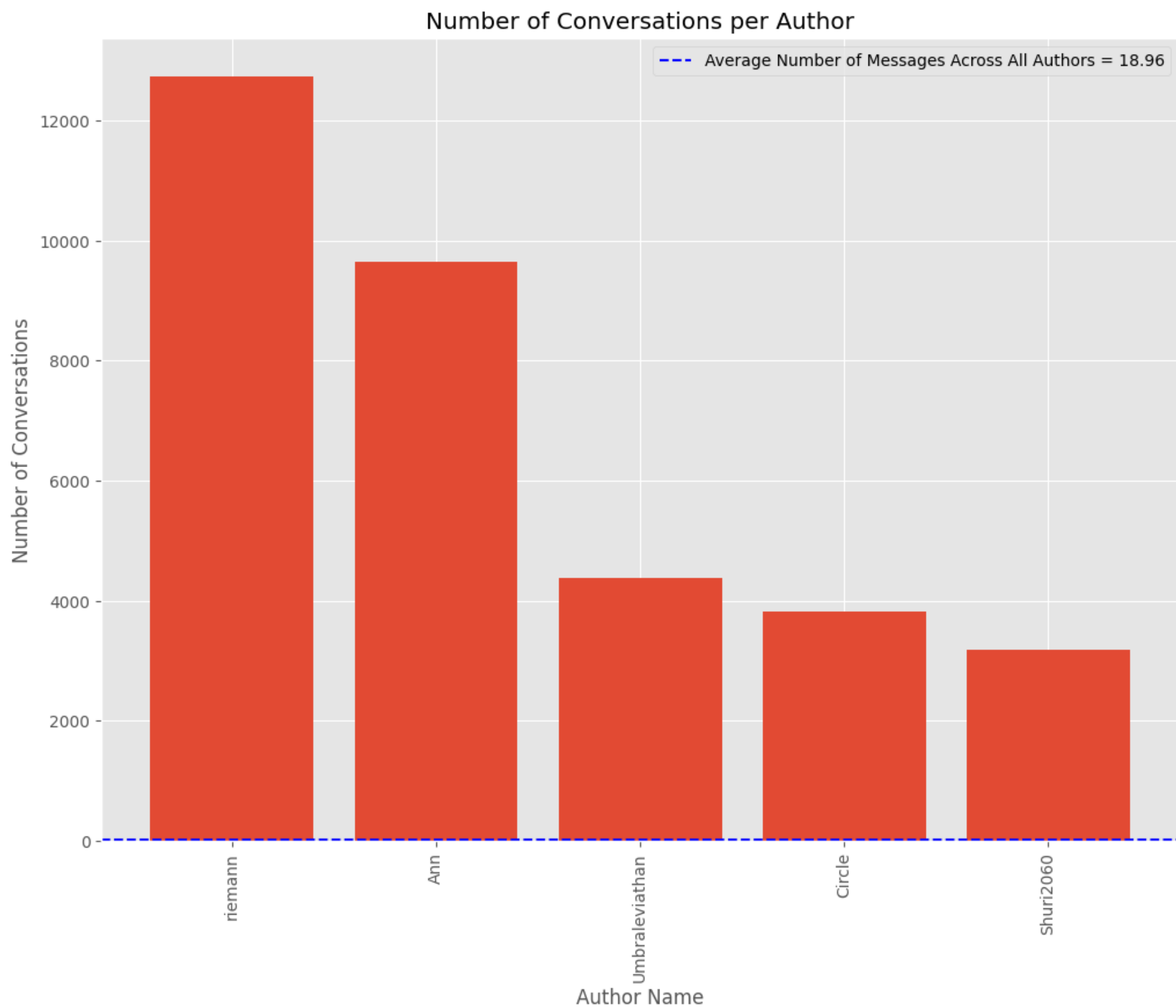


Figure 5.1: Top 5 most active participants in the help channels from November 2021 to January 2023, measured by number of conversations

The data from Figure 5.1, which compares the participation of the top five active users in the MDS against the overall average, reveal a significant concentration among these leading contributors. Their level of engagement, as measured by the number of conversations they have participated in, substantially exceed the community average of 18.96, indicating that these few individuals are central to the support received within the platform. This observation led me to question the nature of engagement within the community and to ponder the underlying factors that drive such sustained and prolific participation. It also influenced the development of criteria to define what constitutes meaningful engagement in the context of the MDS, going beyond mere message counts to understanding the quality, responsiveness, and impact of these interactions on the learners they support. Specifically, in subsubsection 3.5.1.2 where I talk about how students engage, I had to redefine what it meant to have an exchange (or turn of talk) by making sure that it was defined as the author switching and not just a new message occurring.

This particular finding in connection to my observations watching the most active helpers interact with learners everyday has me interested in a few questions related to their engagement: (1) What are the effects of someone getting help from a ‘very active’ tutor compared to one with less experience (measured by number of conversations) on the platform?; (2) What gravitates people to volunteer several hours of their time on platforms such as these?

5.2.3 Conceptions of the derivative in *MathConverse*

What conceptions of the derivative emerge in the MDS, and how do these discussions reflect broader trends in students’ understanding of calculus concepts?

In this chapter, I used the conversations labeled by the topic model from Chapter 3 as pertaining the derivative to build a random sample to label for the conceptions of the derivative using a theoretical framework that combines Zandieh’s (2000) derivative conceptions with Balacheff’s (2013) theory of conception models. I then wrote a prompt that described each of these models and wrote a script that would concatenate this prompt with the students’ contributions from each conversation, and one-by-one, these would get sent to an LLM (GPT-3.5-Turbo) hosted by OpenAI, and the output would come back as one of the categories that I requested. Additionally in this script, I wrote some code that verifies that the model doesn’t say anything else, try to teach me as the user what to do, just explicitly ‘categorize’, something of a workaround for this type of model that was originally trained for text-generation (or to predict the next word).

The findings in Chapter 4 revealed a significant proportion of the conversations labeled as containing the derivative by the topic model to show notions of procedural knowledge in the form of students asking for help in applying rules for finding derivatives. This observation is consistent experiences found by many in their calculus coursework; nevertheless, three conceptions from Zandieh’s (2000) emerged with some prevalence in the sample: Limit of the Difference Quotient, Slope of the Tangent Line, and Difference Quotient. These findings show promise that there are these multifaceted conceptions showing up in the conversations, and that it is possible to prompt pre-trained models to find these conceptions in conversations.

5.3 Limitations

5.3.1 Conversation disentanglement is a challenging, well-studied problem

Conversation disentanglement (covered in subsection 2.2.4), the task of identifying separate threads in conversations, can be an important part of the data cleaning process for researchers working with chat data where the unit of analysis is the conversation rather than the message (Chatterjee et al., 2020; Elsner and Charniak, 2010; Yu and Joty, 2020; Zhu et al., 2021). Initially, my focus on this project was analyzing a different portion of the MDS data—specifically, the channel devoted to calculus learning. In this channel, there were no indicators is no of where conversations started and stopped, so a major part of my work involved training custom machine learning models that could decipher these boundaries. While I was able to make some progress, the model accuracy did not get to a satisfactory level to be able to apply the model at scale and use the results the analyses for the rest of my study. It was in late 2021 when the moderators of the MDS implemented an automated bot system in the help channels that made the conversation disentanglement task more feasible. The policy change changed the norms on how students and helpers interacted with one another with giving and receiving help, so that only one person would ask for help on a problem at a time, which naturally structured the chat logs with clear beginning and and end-points for the conversations.

I share these insights to highlight how important it has been as a researcher who studies text interactions at scale to know the difficulties of this problem of conversation disentanglement, as well as how important being able to identify threads in chat data. I argue that for those who starting research projects involving learners collaboratively engaging with text, choosing a system that balances between the user-friendliness and synchronicity chat platforms with the capability to collect structured data effectively. This balance can ensure the platform is accessible and engaging for the learners while at the same time allow researchers to extract and analyze meaningful datasets from it. As an example, in one learning community I participate in (Uplimit) which utilize Large Language Models (LLMs) to study learner interactions that take place on Slack, there is an emphasis on encouraging participants to use threaded replies in the chat platform. This approach not only facilitates smoother communication among users but also significantly simplifies the research process by organizing conversations more coherently.

5.3.2 Costs and privacy concerns associated with using closed-sourced models

In this project, I opted to use OpenAI's GPT models for many of my analyses, because as of the time of my writing, they currently provide models that are state-of-the-art on a number of foundational tasks, provide an API that makes it easy to run queries from any machine, and several software developers have written libraries that integrate well with the models.

However, using these models does require sending data to this company and signing their user agreement that this data can be used in training their models in the future. Furthermore, it costs money to use these models, and the $1.50/million tokens for the GPT-3.5 Turbo model can add up quickly. In Table 5.1, I provide some statistics for one of my (successful) model runs and the costs associated with it.

Run Count	Total Tokens	Cost (USD)
120,687	8,325,116	$12.49

Table 5.1: Number of runs, total tokens, and associated cost for the question classification task on $n = 120,687$ questions

Therefore, if possible, it can be useful to leverage open-sourced models (e.g., Meta's LLAMA models, Mixtral AI's Mixtral models, or Google's Gemma models). Ollama provides an easy-to-use interface for getting open-source models up-and-running on consumer hardware or one can rent a machine from a cloud service (e.g., Paperspace or Hugging Face). I chose to use GPT

5.4 Future Research Directions

To finish the dissertation, I highlight two findings that inspire several branching studies I see coming off of this work.

5.4.1 Looking at the dual roles participants play as students and helpers

In Figure 5.2, the histogram shows the role distribution of 16,045 participants who acted as a 'student' and as a 'helper' within the help channels at least once. The x-axis represents the percentage of conversations in which participants were students, and the y-axis counts the number of participants corresponding to those percentages. Note a higher concentration of participants on the right side of the plot, which shows that of these participants that have taken on both roles, a majority of them are those who ask questions in the help channels.

Taking into consideration that many participants likely might have acted as a 'helper' and 'student' one time, this histogram depicts the distribution of participant involvement in the student conversations from the *MathConverse* dataset, encompassing 16,045 participants and a total of 463,092 messages. The x-axis represents the percentage of messages that participants contributed as students within conversations, while the y-axis shows the total number of messages corresponding

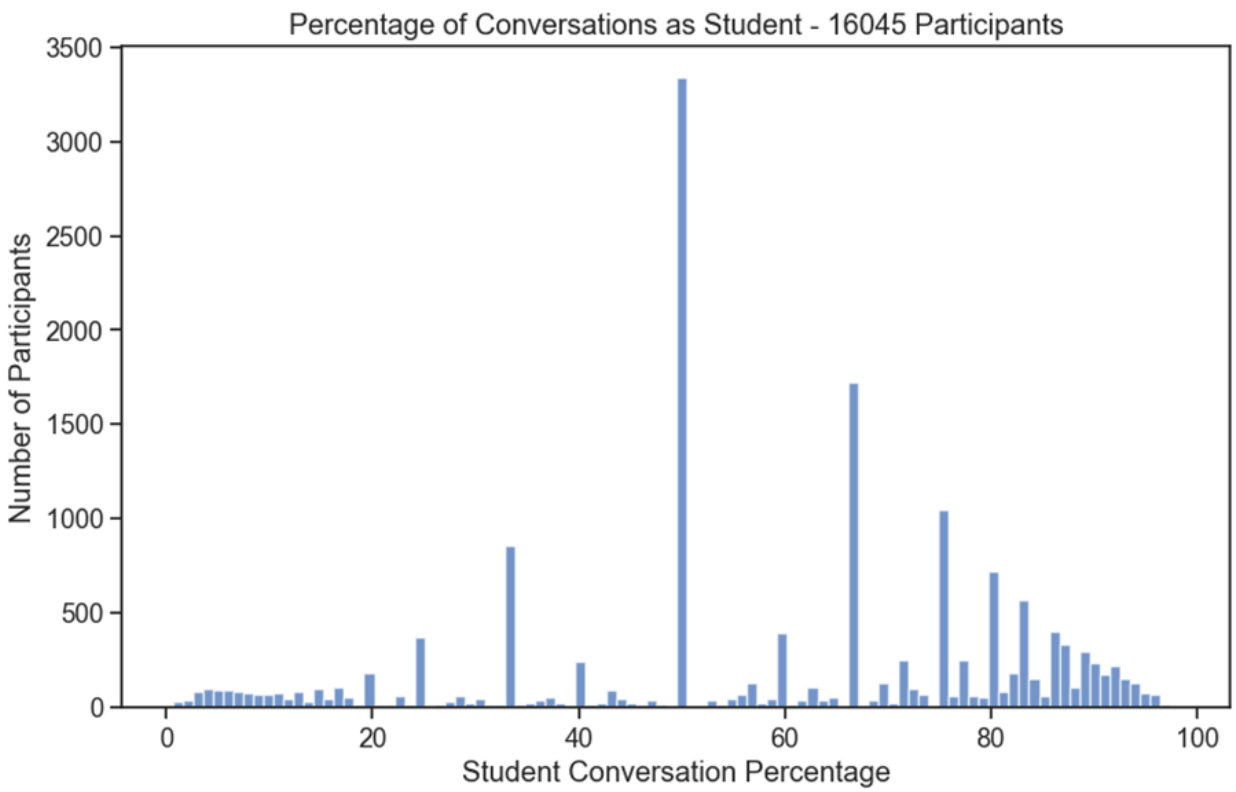


Figure 5.2: Role distribution of participants as students and helpers

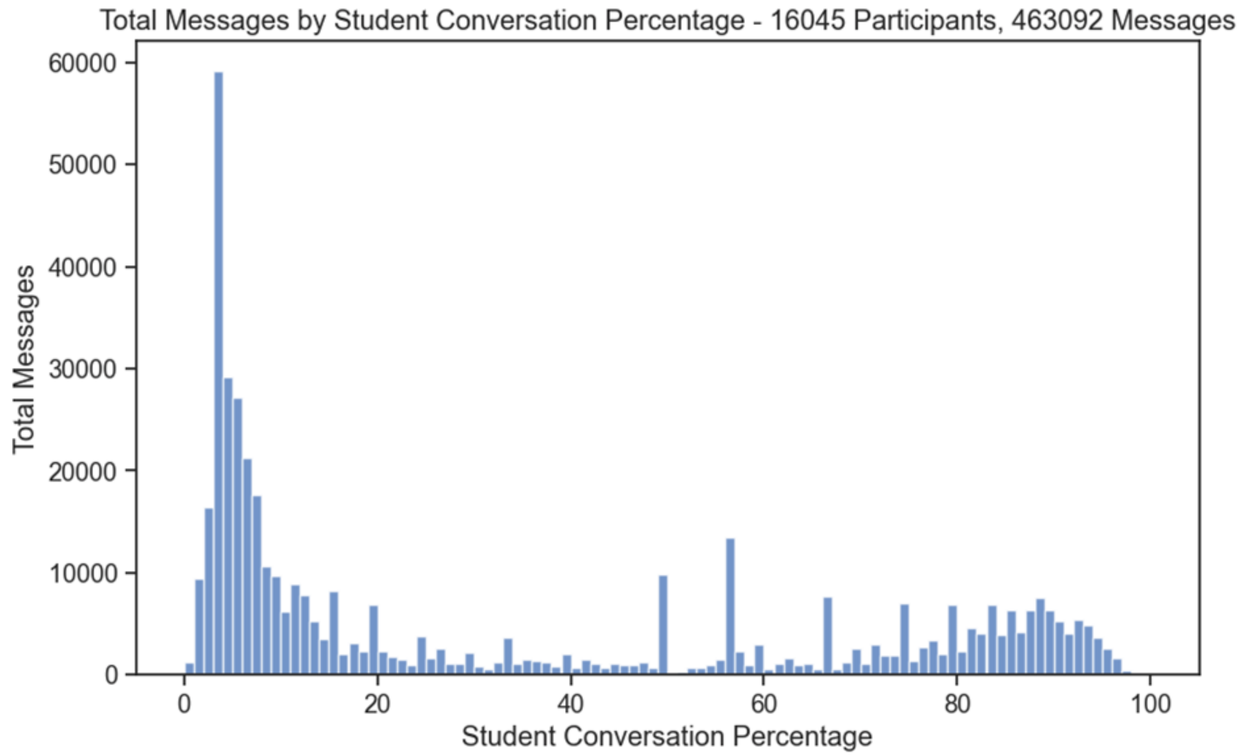


Figure 5.3: Distribution of message participation ratios in conversations

to those percentages. A noteworthy observation from the data is the prominent peak at the lower end of the spectrum, indicating a large number of participants who predominantly contribute as students.

From these visualizations, it shows that a good proportion of the community that use these help channels, 37.1% of the 43,249 total participants, assume dual roles throughout their engagement in the platform. This inspires future research questions on the relationship between the quantity, quality, and types of help students receive in the MDS and the effect these have on helping others with problems (or other forms of being active in the community).

5.4.2 Study of what makes good online tutoring

Having observed the growth and evolution of the MDS for over five years, I have had the privilege to witness some of the most expert helpers guide countless students and watch the ways in which they navigate getting them to work out the concept on their own, even if the student initially was trying to just get someone to answer the problem for them. Although for my dissertation there was considerable value in analyzing the dataset as a whole or through or through the subsets I created, I am drawn towards exploring these segments of the dataset that I believe offer valuable insights into the valuable knowledge gained by these helpers. This includes them trying to assess where the student is at in their mathematical trajectory, asking what they have done so far on the problem, and providing just the right amount of scaffolding to get them started. The knowledge and skill required to facilitate this work that's done over the internet, with anonymous students from around the world is remarkable. I also believe that this work fits outside what text-as-data methods would be able to capture effectively, and would require more careful attention that traditional qualitative methods offer to truly appreciate the subtleties from the exchanges.

APPENDIX A

Selected Python Scripts

A.0.1 Disentanglement Model

```
import json
import os
from datetime import datetime
from pytz import timezone
from dateutil.parser import parse as parse_date
import csv
import re

eastern = timezone('US/Eastern')

def replace_mentions_with_pseudonyms(text,
   author_name_to_pseudonym):
    def replace_mention(match):
        mention = match.group(0)
        author_name = mention[1:]  # Remove "@" symbol
        pseudonym = author_name_to_pseudonym.get(author_name,
           author_name)
        return "@" + pseudonym

    # Replace mentions with pseudonyms
    text = re.sub(r"@\w+", replace_mention, text)
    # Remove newline characters
    text = text.replace('\n\n', ' ')
    return text

def process_file(file, pseudonyms, output_directory):
    with open(file, 'r', encoding='utf-8') as f:
```

```python
        data = json.load(f)

    conversation_id = 1
    conversations = {}
    help_number = data['channel']['name'].split("-")[-1]
    author_pseudonyms_mapping = {}
    author_name_to_pseudonym = {}  # Create a mapping between author names and pseudonyms
    pseudonym_index = 0

    # Make sure the output directory exists
    if not os.path.exists(output_directory):
        os.makedirs(output_directory)

    for message in data['messages']:
        timestamp = parse_date(message['timestamp']).astimezone(eastern)
        message['timestamp'] = timestamp.strftime('%m/%d/%Y %H:%M:%S')

        if timestamp >= datetime(2021, 10, 29, tzinfo=eastern):
            # Extract author name and check if it's "Mathematics Bot"
            author_name = message['author']['name']
            author_id = message['author']['id']
            if author_name == 'Mathematics Bot':
                author_pseudonym = author_name
                author_name_to_pseudonym[author_name] = author_pseudonym
            else:
                if author_id not in author_pseudonyms_mapping:
                    author_pseudonym = pseudonyms[pseudonym_index]
                    author_pseudonyms_mapping[author_id] = author_pseudonym
                    pseudonym_index += 1
                    # Update author_name_to_pseudonym mapping
                    author_name = message['author']['name']
```

```
                    author_name_to_pseudonym[author_name] = author_pseudonym
                else:
                    author_pseudonym = author_pseudonyms_mapping[author_id]

            conversation_key = f"help-{help_number}-{conversation_id}"
            if conversation_key not in conversations:
                conversations[conversation_key] = []

            text = message['content']


            attachments = message.get('attachments', [])
            attachment_url = attachments[0]['url'] if attachments else None

            if author_name == 'Mathematics Bot' and not text:
                if message['embeds']:
                    if 'author' in message['embeds'][0] and 'name' in message['embeds'][0]['author']:
                        text = message['embeds'][0]['author']['name']
                    elif 'description' in message['embeds'][0]:
                        text = message['embeds'][0]['description']

        # Replace mentions with pseudonyms in the text
        text = replace_mentions_with_pseudonyms(text, author_name_to_pseudonym)


        conversation = {
            "author": f"{author_pseudonym}:",
            "text": text,
            "timestamp": message['timestamp'],
            "url": attachment_url
        }
```

```python

        conversation = {key: value for key, value in conversation.items() if value is not None}

        conversations[conversation_key].append(conversation)

        if len(message['embeds']) > 0 and message['embeds'][0]['title'] == 'Channel closed':
            conversation_id += 1

        for conversation_key in conversations:


            # Construct the absolute path for the output file
            output_file = os.path.join(output_directory, f'{help_number}-{conversation_id}.json')
            with open(output_file, 'w', encoding='utf-8') as f:
                formatted_data = {"dialogue": conversations[conversation_key]}
                json.dump(formatted_data, f, ensure_ascii=False, indent=2)

def generate_pseudonyms():
    # Get the psuedonyms from ./disentangle-help-channels/scripts/pseudonyms.csv using os.path.join
    pseudonyms = []
    # Need to use encoding='utf-8' to read the csv file.
    with open(os.path.join(os.path.dirname(__file__), 'pseudonyms.csv'), 'r', encoding='utf-8') as f:
        reader = csv.reader(f)
        for row in reader:
            pseudonyms.append(row[0])
    return pseudonyms

def extract_help_number(file_name):
    # Define a regex pattern to match "help-" followed by one or more digits
    pattern = r"help-\d+"
```

```python

    # Search for the pattern in the file_name and extract the
        match
    match = re.search(pattern, file_name)

    # Return the matched value if found, otherwise return None
    return match.group(0) if match else None

if __name__ == '__main__':

    user_input = input('Enter the file name or directory: ')

    help_number = extract_help_number(user_input)
    print(f'help_number: {help_number}')

    # Get the directory path of the user_input file
    file_directory = os.path.dirname(user_input)

    # Create an output directory in the same directory as the
        input JSON file
    output_directory = os.path.join(os.path.dirname(os.path.
        abspath(user_input)), f'output_{help_number}')



    if not os.path.exists(output_directory):
        os.makedirs(output_directory)

    pseudonyms = generate_pseudonyms()

    if os.path.isfile(user_input):
        process_file(user_input, pseudonyms, output_directory)
    elif os.path.isdir(user_input):
        for file in os.listdir(user_input):
            if file.endswith('.json'):
                process_file(os.path.join(user_input, file),
                    pseudonyms, output_directory)
    else:
```

```python
        print('Invalid input. Please enter a valid file or directory.')
```

Listing A.1: Disentanglement Model

A.0.2 Classification Models

```python
from enum import Enum
from openai import AsyncOpenAI
import instructor
from openai import OpenAI
import json
import asyncio
import time

client = instructor.patch(AsyncOpenAI(), mode=instructor.Mode.TOOLS)
sem = asyncio.Semaphore(5)

\begin{lstlisting}[language=Python, caption=Question Classification Prompt, label={lst:question_classification_prompt}]
from enum import Enum
from pydantic import BaseModel, Field

class QuestionType(Enum):
    BASIC_INQUIRY = "Basic Inquiry"
    PROCEDURAL_REASONING = "Procedural Reasoning"
    CONTEXTUAL_INQUIRY = "Context Inquiry"
    EXPLORATORY_INQUIRY = "Exploratory Inquiry"
    ASSERTIVE_COMMUNICATION = "Assertive Communication"

QUESTION_TYPE_DESCRIPTIONS = {
    QuestionType.BASIC_INQUIRY: "Seeks a simple, factual response or clarification of basic information.",
    QuestionType.PROCEDURAL_REASONING: "Concerns the steps or procedures taken to achieve a certain goal or solve a problem.",
    QuestionType.CONTEXTUAL_INQUIRY: "Questions that require understanding and applying context or scenarios.",
    QuestionType.EXPLORATORY_INQUIRY: "Aims to explore concepts or definitions, often asking for elaboration or examples.",
    QuestionType.ASSERTIVE_COMMUNICATION: "Involves making a statement or claim, often with confidence or certainty,
```

```
        sometimes to persuade."
}

QUESTION_TYPE_EXAMPLES = {
    QuestionType.BASIC_INQUIRY: "What is the first step in solving a quadratic equation?",
    QuestionType.PROCEDURAL_REASONING: "What are the steps involved in isolating a variable in an algebraic expression?",
    QuestionType.CONTEXTUAL_INQUIRY: "How does the concept of elasticity apply to the demand for a product in a competitive market?",
    QuestionType.EXPLORATORY_INQUIRY: "Can you explain the concept of a derivative in calculus?",
    QuestionType.ASSERTIVE_COMMUNICATION: "The theorem can be proved by applying the principle of mathematical induction."
}

def create_prompt(question: str) -> str:
    categories = ', '.join([qt.value for qt in QuestionType])
    return (f"Given the question: '{question}', classify it into one of the following categories based on its content and intent: "
            f"{categories}.")

class QuestionClassification(BaseModel):
    classification: QuestionType = Field(
        description="A list of classifications for a question, indicating the types or categories the question belongs to based on predefined criteria. Each classification must be one of the enumerated types defined in `QuestionType`."
    )

# Example usage:
question = "What is the derivative of x^2?"
prompt = create_prompt(question)
```

```python
print(prompt)

    # If you want only one classification, just change it to
    #   `classification: QuestionType` rather than `classifications: List[QuestionType]``
    classification: List[QuestionType] = Field(
        description=f"An accuracy and correct prediction predicted class of question. Only allowed types: {ALLOWED_TYPES}, should be used",
    )

    @field_validator("classification", mode="before")
    def validate_classification(cls, v):
        # sometimes the API returns a single value, just make sure it's a list
        if not isinstance(v, list):
            v = [v]
        return v



async def main(
    questions: List[str], *, path_to_jsonl: str = None
) -> List[QuestionClassification]:
    tasks = [classify(question) for question in questions]
    for task in asyncio.as_completed(tasks):
        question, label = await task
        resp = {
            "question": question,
            "classification": [c.value for c in label.classification],
        }
        print(resp)
        if path_to_jsonl:
            with open(path_to_jsonl, "a") as f:
                json_dump = json.dumps(resp)
                f.write(json_dump + "\n")

```

```python

# Modify the classify function
async def classify(data: str) -> QuestionClassification:
    async with sem:  # some simple rate limiting
        sleep_time = calculate_sleep_time_based_on_rate_limit()
        await asyncio.sleep(sleep_time)
        return data, await client.chat.completions.create(
            model="gpt-3.5-turbo-1106",
            response_model=QuestionClassification,
            max_retries=2,
            messages=[
                {
                    "role": "user",
                    "content": f"Classify the following question:
                      {data}",
                },
            ],
        )

def calculate_sleep_time_based_on_rate_limit():
    # Updated to 250,000 tokens per minute limit
    tokens_per_minute = 250000
    # Calculate requests per second for the average request size
    requests_per_second = tokens_per_minute / (1000 * 60)
    # Calculate sleep time in seconds to stay within the limit
    sleep_time = 1 / requests_per_second
    return sleep_time

import nest_asyncio
nest_asyncio.apply()

path = "./student_derivative_final.jsonl"

await main(student_docs, path_to_jsonl=path)
```

Listing A.2: Classification Model

```python

from enum import Enum
from pydantic import BaseModel, Field, ValidationError, validator
from langsmith.wrappers import wrap_openai


class DerivativeType(Enum):
    DERIVATIVE_RULES = "derivative_rules"
    SLOPE_OF_SECANT_LINE = "slope_of_secant_line"
    AVERAGE_RATE_OF_CHANGE = "average_rate_of_change"
    AVERAGE_VELOCITY = "average_velocity"
    DIFFERENCE_QUOTIENT = "difference_quotient"
    SLOPE_OF_THE_TANGENT_LINE = "slope_of_the_tangent_line"
    INSTANTANEOUS_RATE_OF_CHANGE = "instantaneous_rate_of_change"
    INSTANTANEOUS_VELOCITY = "instantaneous_velocity"
    LIMIT_OF_THE_DIFFERENCE_QUOTIENT = "
       limit_of_the_difference_quotient"
    GRAPH_OF_DERIVATIVE_FUNCTION = "graph_of_derivative_function"
    RATE_OF_CHANGE_OF_FUNCTION = "rate_of_change_of_function"
    VELOCITY_AS_FUNCTION_OF_TIME = "velocity_as_function_of_time"
    DERIVATIVE_AS_A_FUNCTION = "derivative_as_a_function"
    NONE = "None"

DERIVATIVE_DESCRIPTIONS = {
# LAYER 0
DerivativeType.DERIVATIVE_RULES : "The focus is on identifying the
     use of standard rules and formulas without resorting to the
    limit definition to find the derivative of a function. It
    includes the use of power rule, product rule, quotient rule,
    and chain rule.",
# LAYER 1
DerivativeType.SLOPE_OF_SECANT_LINE : "The focus is on finding the
     slope of a secant line on a graph and how it represents an
    average rate of change, offering a graphical understanding of
    the concept",
DerivativeType.AVERAGE_RATE_OF_CHANGE : "The focus is on
    calculating the average rate of change over an interval,
    verbally discussing the change in function values over the
```

```
    interval and what this indicates about the function's behavior.
    ",
DerivativeType.AVERAGE_VELOCITY : "The focus is on how to
    determine the average velocity of an object over time,
    physically interpreting the derivative as a measure of velocity
    .",
DerivativeType.DIFFERENCE_QUOTIENT : "The focus is on using the
    difference quotient to find the slope of a secant line,
    emphasizing the concept of the derivative as the limit of the
    difference quotient.",
# LAYER 2
DerivativeType.SLOPE_OF_THE_TANGENT_LINE : "The focus is on
    estimating the slope of the tangent line at a specific point on
     a function's graph, utilizing the limit of slopes of secant
    lines as they approach the point of interest.",
DerivativeType.INSTANTANEOUS_RATE_OF_CHANGE : "The focus is on the
     derivative as an instantaneous rate of change of a function at
     a point, verbalizing the concept by examining the limit of the
     average rates of change as the interval shrinks to a point.",
DerivativeType.INSTANTANEOUS_VELOCITY : "The focus is on using a
    position function to find the instantaneous velocity of an
    object at a specific time, emphasizing the concept of the
    derivative as the limit of the average velocity over an
    interval.",
DerivativeType.LIMIT_OF_THE_DIFFERENCE_QUOTIENT : "The focus is on
     using the limit definition to find the derivative of a
    function at a specific point, emphasizing the concept of the
    derivative as the limit of the difference quotient.",
# LAYER 3
DerivativeType.GRAPH_OF_DERIVATIVE_FUNCTION : "The focus is on
    analyzing the graph of a function to understand the behavior of
     its derivative in terms of the instantaneous rate of change,
    specifically how the graph illustrates increasing or decreasing
     behavior, critical points, and inflection points.",
DerivativeType.RATE_OF_CHANGE_OF_FUNCTION : "The focus is on
    describing the behavior of a function based on its
    instantaneous rate of change. It explores how the derivative
    explains the function's increasing or decreasing trends and
```

```
    critical intervals.",
DerivativeType.VELOCITY_AS_FUNCTION_OF_TIME : "The focus is on using a position function to find the velocity function, emphasizing how the derivative is used to describe the object's movement over time, including changes in direction or acceleration.",
DerivativeType.DERIVATIVE_AS_A_FUNCTION : "The focus is on using the limit definition to find the derivative of a function at any point, extending beyond just the slope at a point. It explores how the derivative is a function that describes the rate of change of the original function."
}


ALLOWED_CONCEPTS = [e.value for e in DerivativeType]

class DerivativeConception(BaseModel):

    concept: DerivativeType = Field(
        description=f"A list of classifications for a conversation, indicating the categories of derivative that emerged in the conversation. Each classification must be one of the enumerated types defined in `QuestionType', which includes: {ALLOWED_CONCEPTS}",
    )

def create_prompt(conversation: str) -> str:
    categories = ', '.join([qt.value for qt in DerivativeType])
    return (f" Here are descriptions of each of the derivative types: {DERIVATIVE_DESCRIPTIONS}. Here is a conversation that takes place between a student and a helper: {conversation}. Using these descriptions to inform your decision making, classify the conversation as pertaining to one of the following derivative conceptions: {categories}, only output the classification that best describes the conversation. If none of the classifications are appropriate, output 'None'.")

```

```python
def validate_classification(classification_str: str) -> str:
    try:
        # Convert the classification string to an enum and validate it
        classification_enum = DerivativeType(classification_str)  # Convert string to enum
        classification = DerivativeConception(classification=classification_enum)
        return classification.classification.value
    except ValidationError as e:
        print(f"Validation error: {e}")
        return None  # Handle invalid classifications as needed



client = wrap_openai(openai.Client())

def classify_conversation_with_gpt35instruct(conversation: str) -> (str, int, int):
    prompt = create_prompt(conversation)
    try:
        response = client.completions.create(
            model="gpt-3.5-turbo-instruct",
            prompt=prompt,
            temperature=0,
            max_tokens=20,
        )
        classification_str = response.choices[0].text.strip()
        print(classification_str)
        # Validate the classification using Pydantic
        classification = validate_classification(classification_str)

        input_tokens = len(prompt.split())
        total_tokens  = response.usage.total_tokens if response.usage else 0

        output_tokens = max(0, total_tokens - input_tokens)
```

```python
        return classification, input_tokens, output_tokens
    except Exception as e:
        print(f"Error: {e}")
        return None, 0, 0

def process_conversations_gpt3(df_conversations: pd.DataFrame) ->
   pd.DataFrame:

    # Classify each conversation
    classifications = []
    for conversation in df_conversations['content']:
        classification = classify_conversation_with_gpt35instruct(
           conversation)
        classifications.append(classification)
    # Update DataFrame with classifications
    df_conversations['predicted_category'] = classifications

    return df_conversations
```

Listing A.3: Derivative conception classification using GPT-3.5 Turbo into structured JSONL file with Langsmith logging

A.0.3 Parsing derivative conversations

```
    # Define the threshold for topic significance
threshold = 0.20

# This will hold the document numbers where topic 18 is above the
    threshold
significant_docs = []

# Open the file with topic distributions
with open('./student_mallet_output/mallet.topic_distributions.30',
     'r') as file:
    for line in file:
        # Split the line into a list of floats
        values = line.split()

        # The first value is the document number, so it's
            converted to an integer
        doc_number = int(values[0])

        # The rest are the topic distribution probabilities,
            converted to floats
        topic_distributions = [float(value) for value in values
            [1:]]

        # Check if the value for topic 18 (at index 18) is above
            the threshold
        if topic_distributions[19] >= threshold:
            # If so, add the document number to the list
            significant_docs.append(doc_number)

# Print out the document numbers
print("Documents significantly related to topic 18:",
    significant_docs)
```

Listing A.4: Parsing derivative conversations

A.0.4 Plotting

```
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib.dates as mdates
import numpy as np
import os


figures_dir = './figures'
os.makedirs(figures_dir, exist_ok=True)

# Load DataFrame
# df10 = pd.read_csv('your_data.csv')  # Uncomment and modify if
    your data is in a CSV file


df10['timestamp'] = pd.to_datetime(df10['timestamp'])

# Extracting date for daily analysis
df10['date'] = df10['timestamp'].dt.date

# Daily Analysis
# Count messages per day
daily_messages = df10.groupby('date').size()

# Calculate the average messages per day, rounded to the nearest
    integer
average_messages = round(daily_messages.mean())

# Fit a non-linear polynomial for the trend line
z = np.polyfit(mdates.date2num(daily_messages.index),
    daily_messages, 2)
p = np.poly1d(z)


# Plotting the daily messages
plt.figure(figsize=(12, 6))
```

```python
plt.plot(daily_messages.index, daily_messages, color='skyblue', label='Daily Messages')
plt.axhline(y=average_messages, color='r', linestyle='--', label=f'Average Daily Messages: {average_messages}')

# Non-linear trend line
plt.plot(daily_messages.index, p(mdates.date2num(daily_messages.index)), "g--", label='Trend Line')

# Highlight weekends
for i in range(len(daily_messages.index)):
    if daily_messages.index[i].weekday() >= 5:  # 5 = Saturday, 6 = Sunday
        plt.axvspan(daily_messages.index[i], daily_messages.index[i] + pd.Timedelta(days=1), facecolor='gray', alpha=0.2)

# Month separators and format x-axis labels
plt.gca().xaxis.set_major_formatter(mdates.DateFormatter('%b. %Y'))
plt.gca().xaxis.set_major_locator(mdates.MonthLocator())

# Rotate x-axis labels for better visibility
plt.xticks(rotation=45)

plt.title('Daily Messages in Mathematics Discord Server Help Channels')
plt.xlabel('Date')
plt.ylabel('Number of Messages')
plt.legend()
plt.grid(True)

# Save the plot
plt.tight_layout()  # Adjust layout to fit x-axis labels
plt.savefig(os.path.join(figures_dir, 'daily_messages_enhanced_nonlinear.png'))
plt.close()  # Close the plot to free memory
```

Listing A.5: Plot for daily messages

A.0.5 Parse students and helpers

```python

# Sort the DataFrame by conversation_id and then by timestamp, in
    place
cleaned_df.sort_values(by=['conversation_id', 'timestamp'],
    inplace=True)

# Function to apply to each group to determine student and helper
def label_student_helper(group):
    # Identify the student as the first author in the conversation
    student = group['author_name'].iloc[0]
    # Create 'student' and 'helper' columns
    group['student'] = (group['author_name'] == student).astype(
        int)
    group['helper'] = (group['author_name'] != student).astype(int
        )
    return group

# Apply the function to each conversation group directly on
    cleaned_df now
df_sorted = cleaned_df.groupby('conversation_id').apply(
    label_student_helper)

# Reset index
df_sorted.reset_index(drop=True, inplace=True)
```

Listing A.6: Parsing students and helpers from conversations

BIBLIOGRAPHY

Artigue, M. (1989). Épistémologie et didactique. In de Paris, I., editor, *Recherches en didactique des mathématiques*, volume 2-3 of *Cahier de DIDIREM*, pages 241–286. IREM de Paris.

Asiala, M., Cottrill, J., Dubinsky, E., and Schwingendorf, K. E. (1997). The development of students' graphical understanding of the derivative. *The Journal of Mathematical Behavior*, 16(4):399–431.

Aspinwall, L., Shaw, K. L., and Presmeg, N. C. (1997). Uncontrollable mental imagery: graphical connections between a function and its derivative. *Educational Studies in Mathematics*, 33:301–317.

Baker, B., Cooley, L., and Trigueros, M. (2000). A calculus graphing schema. *Journal for Research in Mathematics Education*, 31(8):557–578. tex.mag_id: 2317299888 tex.pmcid: null.

Balacheff, N. (2013). Ck¢, a model to reason on learners' conceptions. In *Proceedings of the 35th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education.*, pages 2–15, Chicago, IL: University of Illinois at Chicago.

Balacheff, N., Caferra, R., Cerulli, M., Gaudin, N., Maracci, M., Mariotti, M. A., Müller, J.-P., Nicaud, J.-F., Occello, M., Olivero, F., Peltier, N., Pesty, S., Soury-Lavergne, S., Sutherland, R., Trgalova, J., and Webber, C. (2003). Baghera Assessment Project, designing an hybrid and emergent educational society.

Balacheff, N. and Gaudin, N. (2002). Students conceptions: an introduction to a formal characterization. *Le Cahiers du Laboratoire Leibniz*, 65:22.

Balacheff, N. and Gaudin, N. (2009). Modeling students' conceptions: the case of function. *CBMS Issues in Mathematics Education*, 16:207 – 234. ZSCC: 0000003 tex.ids= balacheff2009a, balacheff2010 tex.mag_id: 2995504303 tex.pmcid: null.

Benoit, K. (2020). Text as data: an overview. In *The SAGE Handbook of Research Methods in Political Science and International Relations*, pages 461–497. SAGE Publications Ltd, 1 Oliver's Yard, 55 City Road London EC1Y 1SP.

Bettinger, E. P. and Long, B. T. (2009). Addressing the needs of underprepared students in higher education: does college remediation work? *Journal of Human Resources*, 44(3):736–771.

Bingolbali, E. and Monaghan, J. (2008). Concept image revisited. *Educational Studies in Mathematics*, 68(1):19–35.

Bloom, B. S. (1984). The 2 sigma problem: the search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*.

Borba, M. C., Askar, P., Engelbrecht, J., Gadanidis, G., Llinares, S., and Aguilar, M. S. (2016). Blended learning, e-learning and mobile learning in mathematics education. *ZDM*, 48(5):589–610.

Borba, M. C., Chiari, A. S. D. S., and De Almeida, H. R. F. L. (2018). Interactions in virtual learning environments: new roles for digital technology. *Educational Studies in Mathematics*, 98(3):269–286.

Borup, J., Graham, C. R., West, R. E., Archambault, L., and Spring, K. J. (2020). Academic communities of engagement: an expansive lens for examining support structures in blended and online learning. *Educational Technology Research and Development*, 68(2):807–832.

Bourdieu, P. (1990). *The logic of practice*. Stanford university press.

Brousseau, G. (1997). *Theory of didactical situations in mathematics: didactique des mathématiques, 1970–1990*. Springer Netherlands.

Carlson, M., Jacobs, S., Coe, E., Larsen, S., and Hsu, E. (2002). Applying covariational reasoning while modeling dynamic events: a framework and a study. *Journal for Research in Mathematics Education*, 33(5):352. 00709.

Chapman, L., Masters, J., and Pedulla, J. (2010). Do digital divisions still persist in schools? Access to technology and technical skills of teachers in high needs schools in the United States of America. *Journal of Education for Teaching*, 36(2):239–249.

Chatterjee, P., Damevski, K., Kraft, N. A., and Pollock, L. (2020). Software-related slack chats with disentangled conversations. In *Proceedings of the 17th International Conference on Mining Software Repositories*, pages 588–592, Seoul Republic of Korea. ACM.

Chowdhury, G. G. (2003). Natural language processing. *Annual Review of Information Science and Technology*, 37:51–89.

Cleary, P. F., Pierce, G., and Trauth, E. M. (2006). Closing the digital divide: understanding racial, ethnic, social class, gender and geographic disparities in Internet use among school age children in the United States. *Universal Access in the Information Society*, 4(4):354–373.

Clement, J., Lochhead, J., and Monk, G. S. (1981). Translation difficulties in learning mathematics. *The American Mathematical Monthly*, 88(4):286–290. Publisher: Taylor & Francis.

Cunningham, A. W., Dias, O., and Angulo, N. (2011). Math is not a spectator sport: the effect of online homework-completion tutoring on community college remedial mathematics performance. *Journal of Mathematics Education at Teachers College*, 2(2):59–65. MAG ID: 2281096940 S2ID: 54988f77f1b9daca9d4b96108fcda704dce95845.

Davis, R. B. and Vinner, S. (1986). The notion of limit: Some seemingly unavoidable misconception stages. *The Journal of Mathematical Behavior*, 5(3):281–303. tex.ids= davis1986a tex.mag_id: 1736753123 tex.pmcid: null place: Netherlands publisher: Elsevier Science.

Demszky, D., Liu, J., Mancenido, Z., Cohen, J., Hill, H., Jurafsky, D., and Hashimoto, T. (2021). Measuring conversational uptake: a case study on student-teacher interactions. In Zong, C., Xia, F., Li, W., and Navigli, R., editors, *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1638–1653, Online. Association for Computational Linguistics.

Duff, P. A. (2010). Language socialization into academic discourse communities. *Annual Review of Applied Linguistics*, 30:169–192.

Eisenstein, J. (2019). *Introduction to natural language processing*. MIT Press.

Elbaum, B., Vaughn, S., Tejero Hughes, M., and Watson Moody, S. (2000). How effective are one-to-one tutoring programs in reading for elementary students at risk for reading failure? A meta-analysis of the intervention research. *Journal of Educational Psychology*, 92(4):605–619.

Elsner, M. and Charniak, E. (2010). Disentangling Chat. *Computational Linguistics*, 36(3):389–409.

Engelbrecht, J., Llinares, S., and Borba, M. C. (2020). Transformation of the mathematics classroom with the internet. *ZDM - Mathematics Education*, 52(5):825–841. ISBN: 0123456789 publisher: Springer Berlin Heidelberg.

Escueta, M., Quan, V., Nickow, A. J., and Oreopoulos, P. (2017). Education technology: an evidence-based review. Technical Report w23744, National Bureau of Economic Research, Cambridge, MA.

Fehrman, S. and Watson, S. L. (2021). A systematic review of asynchronous online discussions in online higher education. *American Journal of Distance Education*, 35(3):200–213.

Ferrini-Mundy, J. and Graham, K. (1994). Research in calculus learning: Understanding of limits, derivatives, and integrals. In Dubinsky, E. and Kaput, J., editors, *Research issues in undergraduate mathematics learning*, Mathematical Association of America, pages 19–26. MAA.

Fesler, L., Dee, T., Baker, R., and Evans, B. (2019). Text as Data Methods for Education Research. *Journal of Research on Educational Effectiveness*, 12(4):707–727.

Feudel, F. and Biehler, R. (2021). Students' understanding of the derivative concept in the context of mathematics for economics. *Journal für Mathematik-Didaktik*, 42(1):273–305.

Fuchs, D., Fuchs, L. S., Mathes, P. G., and Simmons, D. C. (1997). Peer-assisted learning strategies: making classrooms more responsive to diversity. *American Educational Research Journal*, 34(1):174–206.

Gall, S. N.-L. (1981). Help-seeking: An understudied problem-solving skill in children. *Developmental Review*, 1(3):224–246.

Gao, F., Zhang, T., and Franklin, T. (2013). Designing asynchronous online discussion environments: Recent progress and possible future directions. *British Journal of Educational Technology*, 44(3):469–483. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-8535.2012.01330.x.

Gnanadass, E. and Sanders, A. Y. (2018). Gender still matters in distance education. In Moore, M. G. and Diehl, W. C., editors, *Handbook of Distance Education*, pages 79–91. Routledge, Fourth edition. | New York : Routledge, 2019. | Previous edition: 2013., 4 edition.

Goldie, J. G. S. (2016). Connectivism: A knowledge learning theory for the digital age? *Medical Teacher*, 38(10):1064–1069.

Gonida, E. N., Karabenick, S. A., Stamovlasis, D., Metallidou, P., and Greece, T. C. (2019). Help seeking as a self-regulated learning strategy and achievement goals: The case of academically talented adolescents. *High Ability Studies*, 30(1-2):147–166.

Graesser, A. C., Person, N., and Huber, J. (1992). Mechanisms that generate questions. In *Questions and information systems.*, pages 167–187. Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, US.

Graesser, A. C. and Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal*, 31(1):104–137.

Grimmer, J., Roberts, M. E., and Stewart, B. M. (2022). *Text as data: a new framework for machine learning and the social sciences*. Princeton University Press, Princeton Oxford.

Gueudet, G. (2008). Investigating the secondary–tertiary transition. *Educational Studies in Mathematics*, 67(3):237–254.

Haleem, A., Javaid, M., Qadri, M. A., and Suman, R. (2022). Understanding the role of digital technologies in education: A review. *Sustainable Operations and Computers*, 3:275–285.

Harting, K. and Erthal, M. J. (2005). History of distance learning. *Information Technology, Learning, and Performance Journal*, 23(1):35–44. Place: Morehead Publisher: Organizational Systems Research Association.

Hattie, J. and Jaeger, R. (1998). Assessment and classroom learning: a deductive approach. *Assessment in Education: Principles, Policy & Practice*, 5(1):111–122.

Heinrich, E. and Carvalho, L. (2022). Fostering connections for professional identity formation: two case studies of Discord discipline-focused communities. *ASCILITE Publications*, page e22136.

Henri, F. and Pudelko, B. (2003). Understanding and analysing activity and learning in virtual communities. *Journal of Computer Assisted Learning*, 19(4):474–487. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1046/j.0266-4909.2003.00051.x.

Herbst, P. and Chazan, D. (2012). On the instructional triangle and sources of justification for actions in mathematics teaching. *ZDM*, 44(5):601–612.

Herbst, P. G. (1995). *The construction of the real number system in textbooks: A contribution to the analysis of discursive practices in mathematics*. PhD thesis, University of Georgia.

Hershkowitz, R., Schwarz, B. B., and Dreyfus, T. (2001). Abstraction in context: epistemic actions. *Journal for Research in Mathematics Education*, 32(2):195.

Hiebert, J. and Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In Lester, F., editor, *Second Handbook of Research on Mathematics Teaching and Learning*, pages 371–404. Information Age, Charlotte, NC. 00000.

Hirschberg, J. and Manning, C. D. (2015). Advances in natural language processing. *Science*, 349(6245):261–266. Publisher: American Association for the Advancement of Science.

Holub, O. (2023). Discord chat exporter.

Hwang, S. and Cho, E. (2021). Exploring latent topics and research trends in mathematics teachers' knowledge using topic modeling: a systematic review. *Mathematics*, 9(22):2956.

Hähkiöniemi, M. (2006). Associative and reflective connections between the limit of the difference quotient and limiting process. *The Journal of Mathematical Behavior*, 25(2):170–184.

Inglis, M. and Foster, C. (2018). Five decades of mathematics education research. *Journal for Research in Mathematics Education*, 49(4):462–500.

Jay, V., Genevieve M. Henricks, Henricks, G. M., Anderson, C. J., Lawrence Angrave, Angrave, L., Bosch, N., Williams-Dobosz, D., Shaik, N., Najmuddin Shaik, Bhat, S., and Perry, M. (2020). Online discussion forum help-seeking behaviors of students underrepresented in STEM. *International Conference of the Learning Sciences*, pages 809–810.

Jurafsky, D. (2009). *Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition*. Prentice Hall, 2 edition.

Ketterlin-Geller, L. R. and Yovanoff, P. (2009). Diagnostic assessments in mathematics to support instructional decision making. *Practical Assessment, Research, and Evaluation*, 14(16). Publisher: University of Massachusetts Amherst.

Kim, Y., Wijaya Ong, C. I., and Fung, F. M. (2023). Supplementary discourse - forming an online learning community with asynchronous discussions. *Journal of Chemical Education*, 100(2):496–506.

Kummerfeld, J. K., Gouravajhala, S. R., Peper, J. J., Athreya, V., Gunasekara, C., Ganhotra, J., Patel, S. S., Polymenakos, L. C., and Lasecki, W. (2019). A large-scale corpus for conversation disentanglement. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3846–3856, Florence, Italy. Association for Computational Linguistics.

Kurianski, K. M., Marzocchi, A. S., and Soto, R. C. (2022). Tools for humanizing mathematics classes in a virtual world (and beyond). *International Journal of Mathematical Education in Science and Technology*, 53(3):698–707.

Kwon, S., Kim, W., Bae, C., Cho, M., Lee, S., and Dreamson, N. (2021). The identity changes in online learning and teaching: instructors, learners, and learning management systems. *International Journal of Educational Technology in Higher Education*, 18(1):67.

Lamon, S. J. (1993). Ratio and proportion: connecting content and children's thinking. *Journal for Research in Mathematics Education*, 24(1):41–61.

Lave, J. and Wenger, E. (1991). *Situated learning: Legitimate peripheral participation.* Cambridge University Press, New York, NY, US. Pages: 138.

Likwambe, B. and Christiansen, I. M. (2008). A case study of the development of in-service teachers' concept images of the derivative. *Pythagoras*, 0(68):22–31. Number: 68.

Lin, S. Y. and Overbaugh, R. C. (2007). The effect of student choice of online discussion format on tiered achievement and student satisfaction. *Journal of Research on Technology in Education*, 39(4):399–415.

Liu, J. and Cohen, J. (2021). Measuring teaching practices at scale: a novel application of text-as-data methods. *Educational Evaluation and Policy Analysis*, 43(4):587–614.

Major, L. and Watson, S. (2018). Using video to support in-service teacher professional development: the state of the field, limitations and possibilities. *Technology, Pedagogy and Education*, 27(1):49–68.

Marchand, G. and Skinner, E. A. (2007). Motivational dynamics of children's academic help-seeking and concealment. *Journal of Educational Psychology*, 99(1):65–82.

McCallum, A. K. (2002). MALLET: A machine learning for language toolkit.

Mesa, V. (2004). Characterizing practices associated with functions in middle school textbooks an empirical approach. *Educational Studies in Mathematics*, 56:255–286. tex.ids= mesa2004a tex.mag_id: 2059757071 tex.pmcid: null.

Miller, K. E. (2021). A light in students' lives: K-12 teachers' experiences (re)building caring relationships during remote learning. *Online Learning*, 25(1).

Moreno, S. E. and Muller, C. (1999). Success and Diversity: The Transition through First-Year Calculus in the University. *American Journal of Education*, 108(1):30–57.

Noble, T., DiMattia, C., Nemirovsky, R., and Barros, A. (2006). Making a circle: tool use and the spaces where we live. *Cognition and Instruction*, 24(4):387–437.

Nurwahyu, B., Maria, G., and Mustangin, M. (2020). Students' concept image and its impact on reasoning towards the concept of the derivative. *European Journal of Educational Research*, 9(4):1723–1734.

Orton, A. (1983). Students' understanding of differentiation. *Educational Studies in Mathematics*, 14(3):235–250.

Patterson, N. (2009). Distance education: A perspective from women's studies. *International Women Online Journal of Distance Education*, 1(2):1–14. Publisher: Emine DEMİRAY.

Pepin, B. (2014). Student transition to university mathematics education: transformations of people, tools and practices. In Rezat, S., Hattermann, M., and Peter-Koop, A., editors, *Transformation - A Fundamental Idea of Mathematics Education*, pages 65–83. Springer New York, New York, NY.

Pratt, N. and Back, J. (2009). Spaces to discuss mathematics: communities of practice on an online discussion board. *Research in Mathematics Education*, 11(2):115–130.

Rasmussen, C., Marrongelle, K. A., and de Carvalho Borba, M. (2014). Research on calculus what do we know and where do we need to go. *Zdm*. tex.ids: rasmussen2014a tex.mag_id: 2078929928 tex.pmcid: null.

Reich, J., Slama, R., Kizilcec, R., Joyner, D., and Koedinger, K. (2023). Research infrastructure: million tutor moves observatories project (mt mop): incubating digital infrastructure for an equitable, intelligent, and resilient stem education system. NSF Project Proposal NSF Award # 2321499, MIT.

Roscoe, R. D. and Chi, M. T. H. (2007). Understanding tutor learning: knowledge-building and knowledge-telling in peer tutors' explanations and questions. *Review of Educational Research*, 77(4):534–574.

Roth, V., Ivanchenko, V., and Record, N. (2008). Evaluating student response to WeBWorK, a web-based homework delivery and grading system. *Computers & Education*, 50(4):1462–1482.

Roundy, D., Dray, T., Manogue, C. A., Wagner, J. F., and Weber, E. (2015). An extended theoretical framework for the concept of the derivative. In Fukawa-Connelly, T., Infante, N., Keene, K., and Zandieh, M., editors, *Proceedings of the Eighteenth Annual Conference on Research in Undergraduate Mathematics Education*, pages 919–924.

Roy, P. K., Singh, J. P., and Banerjee, S. (2022). Is this question going to be closed? Answering question closibility on Stack Exchange. *Journal of Information Science*, page 01655515221118665. Publisher: SAGE Publications Ltd.

Roy, T. (2023). Discourse in discord: An analysis of authority on the UTM math server. *Writing for University and Beyond: A Journal of First-Year Student Writing at UTM*, 3(1):73–78.

Ryan, A. M., Shim, S. S., Lampkins-uThando, S. A., Kiefer, S. M., and Thompson, G. N. (2009). Do gender differences in help avoidance vary by ethnicity? An examination of African American and European American students during early adolescence. *Developmental Psychology*, 45(4):1152–1163.

Schenke, K., Lam, A. C., Conley, A. M., and Karabenick, S. A. (2015). Adolescents' help seeking in mathematics classrooms: Relations between achievement and perceived classroom environmental influences over one school year. *Contemporary Educational Psychology*, 41:133–146.

Schoenfeld, A. H. (1983). Episodes and executive decisions in mathematical problem solving. In Lesh, R. and Landau, M., editors, *Acquisition of mathematics concepts and processes*, pages 345–395. Academic Press.

Sfard, A. (1992). Operational origins of mathematical objects and the quandary of reification: the case of function. *The concept of function: Aspects of epistemology and pedagogy*, 25:59–84.

Shaffer, D. W. (2017). *Quantitative ethnography*. Lulu. com. 00000.

Shen, D., Yang, Q., Sun, J.-T., and Chen, Z. (2006). Thread detection in dynamic text message streams. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '06*, page 35, Seattle, Washington, USA. ACM Press.

Siemens, G. (2005). Connectivism: a learning theory for the digital age. *International Journal of Instructional Technology and Distance Learning*, 2(1):3–10.

Smith, S. U., Hayes, S., and Shea, P. (2017). A critical review of the use of wenger's community of practice (cop) theoretical framework in online and blended learning research, 2000-2014. *Online Learning*, 21(1).

Stewart, J. (2012). *Calculus : early transcendentals*. Brooks/Cole, Cengage Learning, Belmont, Cal.

Strogatz, S. (2019). *Infinite powers: how calculus reveals the secrets of the universe*. Houghton Mifflin Harcourt.

Suresh, A., Jacobs, J., Harty, C., Perkoff, M., Martin, J. H., and Sumner, T. (2022). The talkmoves dataset: k-12 mathematics lesson transcripts annotated for teacher and student discursive moves. arXiv:2204.09652 [cs].

Suthers, D. D., Vatrapu, R., Medina, R., Joseph, S., and Dwyer, N. (2008). Beyond threaded discussion: Representational guidance in asynchronous collaborative learning environments. *Computers & Education*, 50(4):1103–1127.

Swales, J. M. (2016). Reflections on the concept of discourse community. *ASp. la revue du GERAS*, 69:7–19. Number: 69 Publisher: Groupe d'étude et de recherche en anglais de spécialité.

Tall, D. and Vinner, S. (1981). Concept image and concept definition in mathematics with particular reference to limits and continuity. *Educational Studies in Mathematics*, 12(2):151–169. 02999.

Tareen, H. and Tahir Haand, M. (2020). A case study of UiTM post-graduate students' perceptions on online learning: Benefits & challenges. *International Journal of Advanced Research and Publications*, 4(6):86–94.

Thomas, M. (2002). Learning within incoherent structures: the space of online discussion forums. *Journal of Computer Assisted Learning*, 18(3):351–366.

Van De Sande, C. (2011). A description and characterization of student activity in an open, online, mathematics help forum. *Educational Studies in Mathematics*, 77(1):53–78.

Van De Sande, C. (2013). Experiences of a newbie helper in a free, open, online, mathematics help forum community. *Journal of Education and Training Studies*, 1(1):p194–203.

van de Sande, C. C. (2008). *Open, online, calculus help forums: Learning about and from a public conversation*. phd, University of Pittsburgh.

Vayansky, I. and Kumar, S. A. (2020). A review of topic modeling methods. *Information Systems*, 94:101582.

Vergnaud, G. (1981). Quelques orientations théoriques et méthodologiques des recherches françaises en didactique des mathématiques. *Recherches en Didactique des Mathématiques Grenoble*, 2(2):215–232.

Vincent, B., LaRue, R., Sealey, V., and Engelke, N. (2015). Calculus students early concept images of tangent lines. *International Journal of Mathematical Education in Science and Technology*. tex.mag_id: 2145520322 tex.pmcid: null.

Vinner, S. (1983). Concept definition, concept image and the notion of function. *International Journal of Mathematical Education in Science and Technology*, 14(3):293–305.

Voigt, R., Camp, N. P., Prabhakaran, V., Hamilton, W. L., Hetey, R. C., Griffiths, C. M., Jurgens, D., Jurafsky, D., and Eberhardt, J. L. (2017). Language from police body camera footage shows racial disparities in officer respect. *Proceedings of the National Academy of Sciences*, 114(25):6521–6526.

Vollstedt, M., Heinze, A., Gojdka, K., and Rach, S. (2014). Framework for examining the transformation of mathematics and mathematics learning in the transition from school to university: an analysis of german textbooks from upper secondary school and the first semester. In Rezat, S., Hattermann, M., and Peter-Koop, A., editors, *Transformation - A Fundamental Idea of Mathematics Education*, pages 29–50. Springer New York, New York, NY.

Wang, R. E., Zhang, Q., Robinson, C., Loeb, S., and Demszky, D. (2023). Step-by-step remediation of students' mathematical mistakes. arXiv:2310.10648 [cs].

Wardle, E. (2004). Identity, authority, and learning to write in new workplaces. *Enculturation*, 5(2).

Weatherly, G. (2011). Seen but not heard: women in online learning are a source of knowledge, if asked. *Distance Learning*, 8(3):5–12. Place: Charlotte Publisher: Information Age Publishing.

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity.* Communities of practice: Learning, meaning, and identity. Cambridge University Press, New York, NY, US. Pages: xv, 318.

Wenger-Trayner, E. and Wenger-Trayner, B. (2015). Introduction to communities of practice: a brief overview of the concept and its uses.

White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., and Schmidt, D. C. (2023). A prompt pattern catalog to enhance prompt engineering with chatgpt. arXiv:2302.11382 [cs].

White, P. and Mitchelmore, M. (1996). Conceptual knowledge in introductory calculus. *Journal for Research in Mathematics Education*, 27(1):79. 00252.

Yonekura, F. (2006). *A study of millenial students and their reactive behavior patterns in the online environment*. PhD thesis, University of Central Florida, Orlando, Florida. MAG ID: 1585941193 S2ID: 52a8869c2d2c39ee8343f42cb3051c1167a60816.

Yu, B., Kaufmann, S., and Diermeier, D. (2008). Exploring the characteristics of opinion expressions for political opinion classification. In *The Proceedings of the 9th Annual International Digital Government Research Conference*, pages 82–91, Montreal, Canada.

Yu, T. and Joty, S. (2020). Online Conversation Disentanglement with Pointer Networks. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6321–6330, Online. Association for Computational Linguistics.

Zandieh, M. (2000). A theoretical framework for analyzing student understanding of the concept of derivative. *CBMS Issues in Mathematics Education*, 8:103–127.

Zandieh, M. J. and Knapp, J. (2006). Exploring the role of metonymy in mathematical understanding and reasoning: The concept of derivative as an example. *The Journal of Mathematical Behavior*, 25(1):1–17.

Zhang, W. and Watts, S. (2008). Online communities as communities of practice: a case study. *Journal of Knowledge Management*, 12(4):55–71. Publisher: Emerald Group Publishing Limited.

Zhao, W., Chen, J. J., Perkins, R., Liu, Z., Ge, W., Ding, Y., and Zou, W. (2015). A heuristic approach to determine an appropriate number of topics in topic modeling. *BMC Bioinformatics*, 16(S13):S8.

Zhu, R., Lau, J. H., and Qi, J. (2021). Findings on Conversation Disentanglement. arXiv:2112.05346 [cs].