



NORTHWESTERN UNIVERSITY

Computer Science Department

Technical Report
Number: NU-CS-2026-02

January, 2025

Designing a Multimodal Collaboration Analytics Tool to Support Students' Collaboration Literacy Development

Khalil Anderson

Abstract

Recent developments in machine learning have enabled significant advances in multimodal collaboration analytics, providing automated, real-time feedback to support people. These advancements have enabled collaboration analytics to support students' development of collaboration literacy, but students have rarely been directly engaged in the design of these systems. Much prior work applies academic models of collaboration or focuses on supporting teachers in developing students' collaboration literacy. Yet the literature shows that directly engaging students is the most effective way to support their development. In this dissertation, I introduce a set of design goals for collaboration analytics tools to engage students in developing collaboration literacy. I analyzed a set of user studies to understand students' conceptions of collaboration and how they wanted to be supported in developing collaboration literacy, thereby aligning the design goals with students' views. These design goals provide a framework for future collaboration analytics systems to align with students and maximize their understanding and engagement in developing collaboration literacy.

Keywords

Artificial Intelligence, Multimodal Learning Analytics, Automated Feedback, HCI, Educational Technology, Collaboration Analytics

NORTHWESTERN UNIVERSITY

Designing a Multimodal Collaboration Analytics Tool to Support Students' Collaboration
Literacy Development

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Computer Science

By

Khalil Anderson

EVANSTON, ILLINOIS

March 2026

© Copyright by Khalil Anderson 2025

All rights reserved

ABSTRACT

Recent developments in machine learning have enabled significant advances in multimodal collaboration analytics, providing automated, real-time feedback to support people. These advancements have enabled collaboration analytics to support students' development of collaboration literacy, but students have rarely been directly engaged in the design of these systems. Much prior work applies academic models of collaboration or focuses on supporting teachers in developing students' collaboration literacy. Yet the literature shows that directly engaging students is the most effective way to support their development. In this dissertation, I introduce a set of design goals for collaboration analytics tools to engage students in developing collaboration literacy. I analyzed a set of user studies to understand students' conceptions of collaboration and how they wanted to be supported in developing collaboration literacy, thereby aligning the design goals with students' views. These design goals provide a framework for future collaboration analytics systems to align with students and maximize their understanding and engagement in developing collaboration literacy.

ACKNOWLEDGMENTS

This dissertation was a journey and the culmination of many years of hard work that could not have been completed without the support of many people throughout my life. Nothing I could say would repay the support they have given me, but in this section, I acknowledge some of the people who provided invaluable support throughout my journey.

My advisor, Marcelo Worsley, is the first person I want to acknowledge and thank. When I first arrived in Evanston, I was unsure what to expect. I had only heard negative things about Chicago, but my first impression here was positive before even meeting Marcelo. Visiting Marcelo's lab, TIILT Lab, made Northwestern the top place I wanted to go. I could see how excited, engaged, and caring people were about their work. Marcelo also radiates this passion and acceptance, and, along with his technical knowledge, it has helped him be a great mentor to me in completing my dissertation.

I want to thank my committee members, Xavier Ochoa, Noshir Contractor, and Nichole Pinkard. Xavier Ochoa and Noshir Contractor specifically supported me through my defense, while Nichole supported me through my prospectus. Each provided a different perspective on my research and was essential in shaping my final dissertation. Thank you for your time in assisting me in completing my dissertation and working with me to schedule my defense around your busy schedules. I'm also grateful to the professors at Northwestern who taught and challenged me throughout my career, and to the many staff members in the Computer Science department and across Northwestern who supported my academic and research endeavors.

TIILT Lab attracted many exceptional collaborators who aided me with my dissertation in various ways, and I want to thank each of them. Natalie Melo and JooYoung Jang contributed to theme development and manual coding of the 7 Cs. They also helped with the second system iteration. Lucas Zurbuchen and Ethan Chan both helped manage the development of the webpage and improved the UI after Sony stopped collaborating with us. Aziza Mirsaidov contributed to

the integration of several NLP machine-learning models. Tochukwu Eze is another collaborator who has been of significant assistance. He helped with the transition from Angular to React, as more undergrads were more familiar with React, and he has been leading the development and integration of the video features. Some of the newer collaborators, such as Bookyung Shin, Chenghong Lin, and Dawei Xi, have also been significant contributors to developing new directions in system design and, along with Tochukwu, have conducted additional system testing in various new aspects. Vishesh Kumar was a Postdoc in the lab, and it was great to have his mentorship and support throughout his time at the TIILT Lab. Overall, the lab has attracted many excellent collaborators, and I thank them all, including those not previously mentioned.

I also want to acknowledge some of my friends who supported me while completing my dissertation. Mike Smith, Sarah Lee, and Stephanie Jones were among my earliest graduate student friends at Northwestern, and we have grown much together throughout our programs here. We have had much fun together, and you've also provided valuable feedback and support in my development as a researcher. I also met Albe Gutierrez, who has been a great partner and support throughout the stress of developing my dissertation. Herminio Bodon has also been a friend I've made here who's supported me throughout my dissertation, and we've shared many fun times working on basketball analytics.

Sony was another significant source of support, and I want to thank them. This project was initiated in collaboration with Sony, which provided part of the initial funding. They managed the webpage, while we designed the machine-learning-based audio processor for the first iteration. They also performed and supported the initial system testing. This was essential in developing the basic system design for the first iteration.

Finally, I want to thank my family, who I could never have done this without the support everyone has provided. My parents both have always been there for me, pushed me academically, and always believed I could do whatever I put my mind to. My father always pushed me academically and exposed me to various technologies and engineering at a young age, while my mother emotionally grounded me and showed boundless care. My brothers, Kadeem and Rasheed, have

always believed in me, and we always have fun and relaxing times together. I'll forever miss my grandmother, Sadie, who passed away too soon, but whose strength was a great example.

ABBREVIATIONS

AI Artificial Intelligence. 87, 100, 106

BLINC Building Literacy In N-person Collaborations. 17, 19, 20, 22, 50, 62, 63, 66, 69–72, 74, 88, 89, 91, 103, 106, 120–122, 130, 131

CPS Collaborative problem solving. 18, 20, 31, 32, 35, 39, 104, 105

DPDP The Digital Personal Data Protection act, 2023 [1]. 92

FERPA Family Educational Rights and Privacy act [2]. 92, 93, 105, 106

GDPR General Data Protection Regulation, 2016 [3]. 92, 93

ICAP Interactive-Constructive-Active-Passive. 132

LA Learning Analytics. 22, 35, 93, 99–101, 103, 105, 131, 132

LGPD Lei Geral de Proteção de Dados Pessoais (General Personal Data Protection Law), 2018 [4]. 92

LLM Large Language Model. 21, 38, 75, 76, 132

MMLA Multimodal Learning Analytics. 24, 38, 58, 60, 70, 73

NISPI Nonverbal Indexes of Students' Physical Interactivity [5]. 25, 31

NLP Natural Language Processing. 76

GLOSSARY

collaboration a coordinated, synchronous activity that is the result of a continued attempt to construct and maintain a shared conception of a problem [6]. 3, 18–26, 28–39, 42, 44–46, 49–52, 54–61, 68, 70–75, 77–89, 91, 94, 96, 98, 99, 101–110, 112–115, 118–122, 124, 125, 128, 130, 131, 150, 151

collaboration literacy the ability to ascertain and respond to changes in the quality of a collaborative experience. From the student perspective, this entails being aware of one's own contribution to a group and having the ability and willingness to intervene to ensure strong collaboration. From the teacher's perspective, this includes awareness of how different groups are progressing, the ability to respond promptly to those groups, and the development of prompts and activities that facilitate effective collaboration. [7]. 3, 19, 21, 23–26, 28–30, 35, 36, 49, 50, 62, 70–74, 87–91, 101, 103–108, 110, 130–132

collaborative problem solving an individual's capacity to use cognitive processes to confront and resolve real, cross-disciplinary situations where the solution path is not immediately obvious and where the content areas or curricular areas that might be applicable are not within a single subject area of mathematics, science, or reading.[8]. 7, 18, 20, 39

multimodal Involving multiple modes of communication such as linguistic, aural, visual, gestural, spatial, and more. 3, 20, 21, 25, 26, 30, 33, 36, 37, 56, 60, 76, 104

scaffolding Cognitive support given by teachers to learners to help them solve tasks that they would not be able to solve working on their own. [9]. 35, 89, 101, 104, 105

DEDICATION

Dedicated to the memory of my grandmother, Sadie,
who was a significant source of support and motivation. Though you are no longer
with us, your bravery, resilience, and warmth are something I think about every day
to drive me further. I know you would be proud, and I miss you.

TABLE OF CONTENTS

Abstract	3
Acknowledgments	4
Abbreviations	7
Glossary	8
Dedication	9
Table of Contents	10
List of Tables	15
List of Figures	16
Chapter 1: Introduction	18
1.1 Dissertation Approach	19
1.2 Dissertation Overview	20
Chapter 2: Literature Review	24
2.1 Multimodal Learning Analytics: Tools, Techniques, and Relevant Constructs	24
2.2 Educational Dashboards and Human-Centered Learning Analytics	26
2.3 Teacher Analytics	28
2.4 Collaboration Analytics	30

2.5	Collaboration Feedback	34
Chapter 3: Measuring Students' Collaborations in Real-Time		37
3.1	Background	38
3.2	System Design	40
3.2.1	Flask Application Service and SQL Database	41
3.2.2	Recording Device	41
3.2.3	Audio Processing Service	42
3.2.4	Angular Front-End and NGINX	45
3.3	User Studies	46
3.3.1	Initial Studies	46
3.3.2	Business Negotiations Class	46
3.4	Results	47
3.5	Discussion	48
3.6	Conclusion	49
Chapter 4: Understanding Students' Conceptualization of Collaboration		50
4.1	Background	51
4.2	Methods	52
4.2.1	Likert Scale Analysis	53
4.2.2	Free Response Analysis	53
4.3	Results	54
4.3.1	Climate	55
4.3.2	Communication	56

4.3.3	Compatibility	57
4.3.4	Conflict	58
4.3.5	Context	59
4.3.6	Contribution	60
4.3.7	Constructive	61
4.4	System Development	62
4.4.1	Overview	62
4.4.2	Refinements	66
4.5	Discussion	70
4.6	Conclusion	72
Chapter 5: Supporting Students' Development of Collaboration with a Learning Analytics Tool		73
5.1	Background	74
5.1.1	Large Language Model Thematic Coding	75
5.2	Study Design	77
5.3	Results	78
5.3.1	Thematic Analysis	79
5.3.2	Co-occurrence	85
5.4	Discussion	87
5.5	Conclusion	89
Chapter 6: Recognizing Students' Ethical and Privacy Concerns		91
6.1	Background	92

6.2	Study Design	94
6.3	Results	95
6.3.1	Contribution	96
6.3.2	Communication	97
6.3.3	Constructive	98
6.3.4	Climate	98
6.3.5	Context	98
6.3.6	Conflict	99
6.4	Discussion	99
6.4.1	Transparency	100
6.4.2	User Agency	100
6.4.3	Learning Over Assessment	101
6.5	Conclusion	101
Chapter 7:	Centering Students in BLINC's System Design	103
7.1	Background	103
7.2	Design Goals	105
7.2.1	Data Transparency	106
7.2.2	Prioritize Learning over Assessment	107
7.2.3	Positive Reinforcement	107
7.2.4	Setting Flexibility	108
7.2.5	Feature Specific Goals	108
7.3	System Design	110

7.3.1	Audio Features	111
7.3.2	Video Features	118
7.3.3	User Interface	121
7.4	Performance Evaluation	127
7.4.1	Experimental Setup	127
7.4.2	Processing Latency	128
7.4.3	Hardware Usage	129
7.5	Discussion and Limitations	129
7.6	Conclusion	130
Chapter 8: Conclusion		131
8.1	Future Work	132
References		134
Appendix A: Gemini Thematic Coding Prompt		150

LIST OF TABLES

1.1	System Iterations and Additional Features	22
4.1	Table of Responses Coded to each Dimension	54
7.1	Participation and Impact Style Measures	114
7.2	Hardware Parameters	127
7.3	Testing Parameters	128
7.4	Testing Results	128

LIST OF FIGURES

3.1	System Design of Chemistry Dashboard	40
3.2	ReSpeakerV2	42
3.3	Server Architecture of Audio Processor	43
3.4	The Chemistry Dashboard features a UI that shows the percentage of directional audio, the emotional features, keywords, and more	45
4.1	Home Page	63
4.2	Group Discussion Data View	64
4.3	Discussion Timeline	65
4.4	Discussion Direction	66
4.5	Customizable Keyword Component	67
4.6	Discussion Features	68
4.7	Discussion Transcripts color-coded based on DoA	69
5.1	Quantitative Overview	78
5.2	Real-Time vs. Post-Hoc responses	79
6.1	Likert question results	96
6.2	Affirmative question results	97
7.1	BlinC Recent Architecture Diagram	110
7.2	Map of audio data flow	111
7.3	BYOD Join and Speaker Fingerprinting	112
7.4	Discussion Transcript with Speaker Labels	113
7.5	Participation and Impact Style	117
7.6	Map of video data flow	119

7.7	Cartoonification Sample	120
7.8	BYOD Live Connection Sample	121
7.9	BLINC Data Display Desktop vs Mobile	122
7.10	Compare and Contrast	124
7.11	Post Discussion QR Code popup	125
7.12	Post Discussion Questions	127

Chapter 1

INTRODUCTION

Collaboration is a highly valued skill within the workplace [10]. Though it is regularly practiced throughout different levels of education, it is rarely taught directly in a subject dedicated to developing collaboration. Previous studies have shown that Collaboration is among the most essential skills for succeeding in the current workforce [11]. Yet, students' Collaborative problem solving (CPS) skills usually remain low even after finishing a professional degree [12]. It seems to be taken for granted, as children regularly learn to play or work collaboratively as they age. Most collaborative work is evaluated based on a deliverable or product. When feedback is given, it is usually on the quality of the deliverable rather than on the active collaboration process at the individual level [13]. Much of Learning Analytics research has focused on the development of collaborative skills.

There have been multiple studies on developing collaboration skills in students. Many studies have shown the importance of individual, real-time feedback on the collaboration process for developing collaborative skills [14, 13, 15]. This is because students need to be able to identify the dynamics of their collaborations in real time and actively respond to them to maintain balanced, high-quality collaborations. With growing interest in Learning Analytics(LA) tools, many universities have attempted to create LA tools of their own for a variety of purposes, such as: increasing retention, improving graduation rates, personalizing students' curriculum, improving the learning experience and performance, determining student success, and analyzing behavior [16]. These systems have succeeded in various ways, but only a few focus on real-time in-person collaboration feedback. Many systems are rarely mapped to theory on understanding and improving collaboration [17, 18], or focus on specific contexts [15] or aspects of collaboration[19]. Several significant issues limit most models.

A common issue is that the student perspective is rarely addressed directly, including their ethical and privacy concerns [20]. Without this view, systems—even those based on sound collabo-

rative theory—won’t engage students naturally and may hinder collaboration. This engagement is vital to students’ retention of collaboration skills [21, 22], so students’ views should be considered when developing collaborative feedback and interventions. Along with this, when teachers drive students’ collaboration literacy development, other factors, such as age, gender, data literacy, and more, can mediate the effectiveness of collaborative feedback. This has driven a focus on directly supporting students and engaging them in their collaboration literacy development [23]. Ensuring students’ maximal participation in their collaboration literacy development is essential also to maximize that development.

This dissertation addresses the stated issues in collaborative analytics by engaging students in designing a collaborative development system. By first developing a formative prototype, the Chemistry Dashboard[24], a sample system was used to examine students’ opinions on engaging with an automated LA system. Surveying students provided new insights into college-aged students’ comprehension of collaboration and their engagement with automated LA collaboration development systems. The insights into the comprehension of collaborations informed the design of a second system iteration, BLINC[25], that addressed students’ concerns while adding new features. Drawing on insights from the desired ways of engagement, I fused the analysis of student survey feedback with prior work on supporting the development of collaborative skills to detail a set of design goals for the final iteration of the system. These design goals and implications from the studies in this dissertation inform the design of other collaboration development tools and are helpful for the design of other student-centered LA tools.

1.1 Dissertation Approach

I took a multistage approach to designing a collaborative development system for students, foregrounding their views through survey research. I was able to describe trends and attitudes toward the system iterations, which helped validate and test our design theories. Surveys were analyzed using a constructivist grounded theory approach [26, 27, 28] to conceptualize students’ understanding of collaboration, their engagement with a collaboration development system, and their privacy

concerns. Following this approach, data are collected and analyzed to refine the system design and to develop new questions based on open areas of inquiry for subsequent collection. The analysis of the feedback will be integrated with theoretical LA design principles for teaching collaboration as seen in Collaborative problem solving ()Coach2.0[29] and A Pragmatic Framework for Multimodal Collaboration Feedback[18]. A final theory of the system design was then developed based on the analysis of the feedback rounds.

1.2 Dissertation Overview

This dissertation comprises five studies that together constitute my research. My overall goal was to determine how to center students' needs in the design of a collaborative development system. This involved multiple user studies to comprehensively document these needs, with a few iterations of the system's design accompanying the studies. The following questions guide this dissertation:

RQ1: What features are necessary to measure college students' collaborations?

RQ2: What is college students' comprehension of collaboration?

RQ3: What support do college students want to engage with their collaborations?

RQ4: What ethical and privacy concerns do college students have when engaging with a collaboration analytics tool?

RQ5: What features should be added to BLINC to align with students' goals, comprehension, and concerns?

The first study in chapter 3 was formative, aimed at assessing the basic viability of the collaborative analytics tool in college classes, and focused on answering **RQ1:** What features are necessary to measure college students' collaborations? Initially, the tool focused on detecting a few simple indicators from transcribed audio data. This transcribed audio was analyzed for affect, participation type, keywords, and direction of arrival. The system helped a professor understand the dynamics of group interactions during mock negotiations and in a few other instances. In successfully implementing a prototype, this system could further test how students want to engage

with a collaboration analytics tool.

The following study in chapter 4 focused on describing how college students conceptualize collaboration by performing and analyzing a survey focused on answering **RQ2**: What is college students' comprehension of collaboration? This involved surveying 131 students about their experiences with collaboration and what constitutes a good collaboration. This work dissected different students' viewpoints and, through consensus coding, aggregated them into major categories: climate, communication, compatibility, conflict, context, contribution, and constructive. Each category was further analyzed to understand how to support students in each area. This analysis allowed me to refine the system to support students' conceptualization of collaboration. Furthermore, this analysis can support the addition of new indicators and the high-level facets of collaboration.

The prototype system informed the subsequent study in chapter 5, which examined the interactions that students want from a collaboration analytics system. This involved allowing 293 computer science students in a college course to pilot the system. They then completed a survey that asked how they want to be supported in their collaborations and in their collaboration literacy development. This was analyzed using the themes of collaboration detailed in chapter 4 to continue supporting students' conceptualization. This survey was investigated through manual human and Large Language Model (LLM) coding, which together aim to account for each bias. This analysis helped answer **RQ3**: What support do college students want to engage with in their collaborations? To this end, I assessed students' preferences for interacting with the system.

Within chapter 6, these students were also surveyed on their privacy and ethical concerns to answer **RQ4**: What ethical and privacy concerns do college students have when engaging with a collaboration analytics tool? The same 293 students from the previous class were then surveyed with a new set of questions about these concerns. This survey was analyzed following the same methods as chapter 5, combining manual and automated LLM coding. Integrating this analysis with prior work on privacy enabled detailed recommendations for the design of multimodal collaboration analytics systems, which are incorporated in the most recent system design.

Chemistry Dashboard (Ch. 3)	BLINC v1 (Ch. 4)	BLINC v2 (Ch. 7)
Audio recording Angular-based UI SQL database storage Transcription Direction of Arrival(DoA) Keyword Detection Textual Sentiment Analysis Question Detection Respeaker integration Bring-Your-Own-Device (BYOD) Mobile UI	Speaker Diarization Colored DoA Increased UI modularity LDA Topic analysis On Device Microphone Control Description Tooltips	Student Interactivity Reactive UI BYOD main use case Video Recording Video Cartoonification Compare and Contrast Participation and Impact Style Analysis Post Discussion Meta-Survey Live Deployment Improved Speaker Tracking

Table 1.1: System Iterations and Additional Features

In chapter 7, recommendations derived from student feedback are combined to refine and detail the design goals, aligning the system with the student perspective. These design goals allow me to finalize the system design and test a live deployment to assess its feasibility. The significant stages of the system’s development included three completed iterations of the collaboration development system, as shown in Table 1.1. The first iteration is a system that uses only audio to detect collaboration indicators, as described in chapter 3. The audio was transcribed to detect a few verbal collaboration indicators. The second iteration followed the study in chapter 4 to align the initial design with students’ views. The third iteration is the most recent and integrates insights from this dissertation and prior research.

In summary, the overall goal of my dissertation is to improve our understanding of how to design automated feedback and scaffolds to best support students’ in-person collaborative skills development. This involves iteratively developing the system and engaging with students between iterations. Early prototypes were used to survey and provide insights for the final system design and implications for other LA tools. Each iteration drove the question and analysis of the next iteration. A final set of design goals was made by integrating these survey insights with LA research. This design will be grounded in students’ collaborative knowledge and their desired approaches

to working with a collaborative analytics tool. Through this analysis and design development, my dissertation will provide information to support the design of future collaboration analytics systems in LA while also developing a new system to improve students' development of collaboration literacy.

Chapter 2

LITERATURE REVIEW

This dissertation focuses on supporting students' development of collaboration literacy. According to Roschelle et al. 1995, a collaboration is "a coordinated, synchronous activity that is the result of a continued attempt to construct and maintain a shared conception of a problem" [6]. This definition of collaboration is frequently cited in other works on collaboration. Meanwhile, Worsley et al. define collaboration literacy as the ability to ascertain and respond to changes in the quality of a collaborative experience. From the student perspective, this entails being aware of one's own contribution to a group and having the ability and willingness to intervene to ensure strong collaboration. From the teacher's perspective, this includes awareness of how different groups are progressing, the ability to respond promptly to them, and the development of prompts and activities that facilitate effective collaboration [7]. These two concepts describe the skill that this paper aims to help students develop. A body of prior research on MMLA, educational dashboards, teacher analytics, and, specifically, collaboration analytics strongly supports the work presented in this dissertation.

2.1 Multimodal Learning Analytics: Tools, Techniques, and Relevant Constructs

Conceptually, this dissertation is heavily motivated by both the affordances and shortcomings of current work in MMLA. At a high level, MMLA aims to examine and triangulate across a variety of modalities (e.g., audio, video, gesture, heart rate, electro-dermal activity, location, and eye tracking) to investigate how users express interest, engagement, or learning. Since its introduction, MMLA has been closely linked to the study of collaboration and collaborative problem-solving. Much of the early work in this subfield emerged out of the Mathematics Data Corpus [30]. The dataset provided researchers with access to audio, video, and digital pen data from groups of high school students completing a collection of mathematics problems. Some papers that used this

dataset demonstrated that predictions of domain expertise and knowledge of the answer could be made with ease using different features. For instance, Zhou, Hang, Oviatt, Yu, and Chen (2014) [31] demonstrate that combining multimodal data (written and spoken) can predict the dominant domain expert within a group with 100% accuracy. This multimodal prediction substantially exceeded what could be achieved with a single modality.

Ochoa and colleagues did complementary work by surfacing the utility of relatively easy-to-extract features [32]. Some of these features included time spent using calculators, writing speed, and the frequency of use of mathematical terminology, all of which are effective for discriminating between domain experts and domain novices. While this work represented a strong first step for MMLA in supporting the study of collaboration, the learning context was limited, and strategies for real-time student collaboration were absent. This work primarily highlights features relevant to researchers who may wish to develop theories or tools to assess students participating in a collaboration, rather than directly supporting the development of students' collaboration literacy.

Worsley and Blikstein (2015) [33] expand the types of learning settings under consideration by studying pairs of students as they complete hands-on engineering design challenges. They also introduce specific learning interventions and examine how those treatments impact the overall process of the collaborative engineering design task. Like the previous work on MMLA, Worsley and Blikstein triangulate among several modalities: speech, gesture, and skin conductance. This multimodal data is used to construct prototypical multimodal engagement states, which serve as a basis for interrogating each group's design process. Although this work has expanded to a hands-on learning context, it once again focuses on post-hoc analysis. It provides little direct feedback to participants based on their group interactions.

Recent work by Cukurova, Luckin, Millan, and Mavrikis (2018) [5] discusses the Nonverbal Indexes of Students' Physical Interactivity [5] (NISPI) framework, which proposes nonverbal indices of students' physical interactivity as a means to study collaborative problem-solving among high school and undergraduate students completing hands-on tasks. The paper examines explicitly measured variables based on head direction and hand position to represent synchrony, accountabil-

ity, equality, and intra-individual variability. Notably missing from this analysis of collaboration practices, however, are the actual verbal cues, which the authors acknowledge can be central to studying group interactions. Beyond this, the study describes a platform that is utilized post-hoc, leaving the creation of a real-time system for future development.

The three papers mentioned here are a small subset of the many papers from the MMLA community that address collaboration-related constructs [5, 34, 35, 36, 37, 38, 39, 40, 41, 42]. Collectively, prior work in MMLA and collaboration analytics highlights the many ways that MMLA can be used to identify and extract collaboration-relevant constructs. These constructs range from individual- and group-based to unimodal to multimodal, and from literal to semantic [7].

However, a recurring pattern in much of this prior research is that analyses are conducted almost exclusively after the activity is completed [7, 43]. Hence, there is no loop closure with the provided data. While there are instances of interactive tabletops [44] and gesture-based systems that provide real-time information [30] about each student's level of participation, none of these platforms have been developed to improve collaboration literacy. Instead, many existing tools were designed to advance theories of collaborative problem solving, improve the quality of students' work on specific tasks, or provide teachers with insights. In contrast, this dissertation primarily addresses the need to help students develop collaboration literacy, an area that has not been extensively studied within the MMLA community.

2.2 Educational Dashboards and Human-Centered Learning Analytics

While not explicitly about supporting collaboration literacy, Reimers and Noevesky (2015) [45] aim to center student use of learning analytics dashboards as a critical area of inquiry. Their study surveyed K–12 and university students to identify which features should be included in the student view of a learning management system (LMS). Student survey responses highlighted the importance of features such as automatic assignment reminders, the ability to compare progress with peers, and integration with social media platforms. Given the specific context of an LMS and a relatively broad set of platform goals, Reimers and Noevesky (2015) offer few particular

design considerations. However, their work is an early contribution to a larger literature on human-centered learning analytics and educational dashboard design, and underscores the importance of student voice.

More recent work has explicitly considered a broader range of stakeholders within learning analytics [46, 47, 48]. Researchers have highlighted the bridge between theory and practice. Alhadad [46] distinguishes between visualizing data as a means of communication and as a research methodology. Especially when visualizing data for communication, it is vital to leverage prior research on cognitive psychology's potential challenges to human attention and to apply general graphic design best practices to focus user attention [49].

Echeverria et al. (2018) [48] utilize the language of explanatory and exploratory visual learning analytics. Namely, they position researchers as conducting exploratory work, and students and practitioners as using learning analytics for descriptive purposes. In line with the emphasis on supporting explanatory learning analytics, they discuss the importance of task alignment and the visibility of contextual information. When students and practitioners engage with a dashboard, they typically do so with specific tasks in mind and may need only certain information in particular contexts. These ideas generally align with best practices around minimalism and supporting users' ability to tell a story with the data [49, 46, 48]. This idea of storytelling also underscores the importance of supporting multiple interpretations of the data and avoiding generalizations and equating causation with correlation [46]. In addition, there have been other studies on how educational information is presented to students, specifically regarding cognitive load.

Cognitive load theory, initially coined by Sweller in 1988 [50], details how, when teaching a student, a specific limit is placed on their cognitive load. When too many unrelated cognitive tasks are included during learning and problem-solving, a student's education may be impeded by exceeding their limited cognitive resources. In learning, students must perform schema acquisition and automation. Schema acquisition involves learning new concepts, whereas automation automatically recognizes and processes them. Other mental activities involved in learning are referred to as extraneous activities. A student's total working memory capacity limits the number of extra-

neous tasks a student can perform while maintaining high-quality learning. A standard comparison to illustrate this is that working through step-by-step examples typically yields faster learning than solving large numbers of problems. Without being shown the steps, students must first spend time recognizing and remembering them before they can learn the problem's concepts. This emphasizes the need for educational tasks to be designed to mostly only engage related cognitive processes.

By analyzing prior work, Plass et al. [51] identified five specific effects of cognitive load. The worked-example effect is the example above, which shows that studying worked examples reduces cognitive load by allowing students to focus on problem- and solution-steps. The goal-free effect states that goal-free problems reduce load by enabling students to focus on problem states rather than end states. The split-attention effect states that replacing multiple related data sources with a single source reduces cognitive load by minimizing the need to integrate them. The completion effect refers to the reduction in cognitive load that results from working on subsections of partially completed problems, which reduces the size of the problem space. The final effect, the redundancy effect, is similar to the split-attention effect, but in this case, the new single data point doesn't support integration and instead removes redundant information. In designing learning analytics, there has been work applying cognitive load theory directly to the design process [52, 53, 54]. It will be helpful to use these findings in the design. These prior works are all essential considerations when designing and implementing educational dashboards that inform this work.

2.3 Teacher Analytics

Echeverria et al. (2018) [48] push into the domain of teacher analytics, a growing sub-area of learning analytics (LA) that aims to enable teachers to better understand and support their learners across a variety of learning configurations [47]. This work is dually motivated by the challenges posed by teaching large classes and by the increasing availability of data within learning management systems [55, 56, 57]. While the current paper focuses on addressing student needs and preferences, the broader goal of collaboration literacy also involves both students and teachers. To this point, teachers can play an essential role in helping students develop collaboration skills.

Hence, work to support teacher interactions with collaboration data and visualizations is also relevant to the current work. Additionally, many insights from research on teacher analytics can inform thinking about the student experience and the representation of student data. However, as prior researchers have suggested, as in Van Leeuwen et al. 2015 [58], the actual creation and study of these tools for teachers or students has scarcely been discussed within the LA community, even less so in the context of studying or supporting collaboration.

Van Leeuwen [58] provides an early attempt to address this gap by conducting an experimental study with 40 pre-service teachers who were tasked with responding to different collaboration scenarios. More specifically, the paper describes two tools implemented to give pre-service teachers a window into group progress: the Concept Trail and Progress Statistics. Concept Trail tracks the specific topics discussed within a given group and presents them on a timeline. Progress Statistics tracks the number of words students enter into the text editor, enabling teachers to assess students' progress on the assignment. In their study, the authors found that pre-service teachers using these two tools were generally more responsive to students.

However, they did not notice any noticeable differences in pre-service teachers' ability to identify problematic group interactions. Moreover, they found that because the Concept Trail specifically provided information about the topics under discussion, pre-service teachers in the learning analytics condition (i.e., those who used the Concept Trail) tended to focus on conceptual problems while overlooking specific social indicators of collaboration challenges. Pre-service teachers' inexperience may influence some of these factors in the field. Even so, this analysis provides several relevant insights that inform our work.

First, the information provided certainly affects what teachers and students pay attention to. To ensure a smooth interaction across different collaboration levels, we must be aware of what to present and how to present it. Second, this study reiterates the importance of collaboration literacy, as providing data does not necessarily lead to greater awareness of inequitable or otherwise unproductive collaboration practices. Third, the study by Van Leeuwen et al. [58] used fairly simple measures of collaboration quality. This is both promising and an opportunity for improvement. It

is promising that even the simplest measures can alter participants' behavior during a collaborative experience. However, consistent with Worsley and Ochoa's taxonomy [7], Van Leeuwen et al. [58] primarily focus on very literal, individual-level measures. While these can likely be used in beneficial ways, the need is to consider more complex patterns of collaboration through multimodal data. Moreover, as we will see in the following section, many critical dimensions of the collaboration experience are being overlooked in these simple representations. Hence, I position the current work as addressing an essential gap in the information provided to both students and teachers about the collaboration process.

2.4 Collaboration Analytics

There have been multiple models of collaborations [8, 59, 60, 61, 18, 7, 62, 63, 6]. Along with this, in previous studies to support collaboration literacy development, the feedback provided to students is simplified to a few graphs or data points of high-level collaboration constructs [64, 20, 15, 65, 66]. This work, mapping skills to constructs to actions, is vital for determining which modalities and features are necessary to provide effective feedback. Across these models, there are patterns in which features are used to assess collaboration quality.

In many of the models, a popular way of breaking down collaborations is into distinct social and cognitive skills [8, 59, 60, 67, 68]. Still, many follow other breakdowns [61, 6, 7, 63, 69]. Each model is designed for a specific domain, and the models rarely can transfer to other domains.[70]. Some of the models are mainly for assessment of student collaboration [8, 59, 68] while others focus on providing data to researchers [67]. Alongside this, many models were developed during the COVID-19 pandemic, leading many to focus on remote work [15, 20, 71]. This makes using these models for in-person learning difficult, as reinforcement learning models rely on a similar input structure to their training inputs, which can cause them to struggle in new domains. Even so, these models' facets remain valid in other domains, but the indicators must be adjusted accordingly. The aspects must mainly be learner-centered and comprehensible for students. These facets are further explored to see which set best aligns with this goal.

Many models share similar sub-facets but differ in their high-level facets. For example, *ATC2IS* [59] categorizes social skills at the highest level into participation, perspective-taking, and social regulation, while cognitive skills are task regulation and knowledge building. Many social cognitive split models focus on assessment or research rather than on learning. Looking at another recent model, a *Generalized Competency Model of CPS* [61], it shows greater overlap among sub-facets within the collaborative facets' divisions. Maintaining team function encompasses participation, task regulation, and social regulation. Negotiation/Coordination includes participation, perspective-taking, social regulation, and task regulation. Construction Shared Knowledge includes perspective-taking, participation, and knowledge-building. Each of these focuses on human coding of these moments, but some models aim to develop numerical methods for calculating collaborative metrics.

One model that attempts to calculate the dynamics of collaboration is Group Communication Analysis (GCA) by Dowell et al. [72]. GCA detects linguistic patterns of collaboration. This first involves transcribing a collaboration and embedding it in a semantic vector space using latent semantic analysis [73]. This requires a corpus that covers the topics that are going to be discussed to define the LSA space, but there are new methods that can be used to provide semantic vectors I explore later that can get around this need. Using semantic embeddings, multiple collaboration features can be computed, including participation, responsiveness, internal cohesion, social impact, newness, and communication density. By analyzing groups using these metrics, the researchers illuminated the different social roles that students play within collaborations. Using this information could provide students with vital insight into their roles in collaboration and help them understand which skills they may have less experience with. In different contexts, collaboration may not be primarily verbal, and my work focuses on in-person interactions, so nonverbal indicators have shown success in modeling for these interactions [5, 74, 75, 76]. I can look further at how these indicators were used.

Across several models, the Nonverbal Indexes of Students' Physical Interactivity [5] (NISPI) framework [5] uses cameras to analyze students' physical interactions. This study involved high

school and college students. Students were tasked with working on an Arduino-based physical computing kit called TALKOO. This included an Integrated Development Environment (IDE) for coding the hardware, specific sensors, and hardware modules, as well as prototyping materials. Thirty students worked in multiple sessions while being measured with this framework. Within the framework, students were grouped into high- and low-skilled collaboration students, and the distribution of indicators was analyzed to compare the groups. The tested indicators — synchrony, equality, and intra-individual variability — showed predictive validity for the competency in collaboration skills. By measuring these indicators, the study showed that they are associated with high CPS group performance. These are visual indicators of good collaboration; other models have also used them or determined different indicators.

Computer Vision for Position Estimation (CVPE) [76] also uses video data and OpenCV [77] to track movement. CVPE was used in a mock hospital-nurse scenario. This involved attempting to detect scripted movements within the collaboration using a camera placed in a corner. The paper related relative camera-based positioning to the wearable's physical position and observed a correlation between the obstructed angle in the video and the wearable's absolute position. In two slightly different collaborations, it successfully captures estimation in one context, specifically one with a large amount of movement and point-of-interest spacing, which supports the idea that visual position can enable understanding of collaboration quality. Zhou et al. [74] also demonstrated that collaboration quality can be predicted from video data by measuring individual engagement via gaze-behavior types. The gazes are then clustered into types of collaborators, following the approach of Zhou et al. 2022[78]. They then observed that students classified as drivers—the most engaged collaborators—gain significantly more from the post-tests than passengers—the least engaged. There have also been a few other ways of exploring collaboration dynamics.

Five constructs of collaboration have been computationally explored individually in the field of Learning Analytics research. These are social dominance, affect, social cohesion, conflict resolution, and engagement. Social dominance is a common phenomenon that has been explored [79, 80, 81]. This is defined as "a relational, behavioral, and interactional state that reflects the

actual achievement of influence or control over another via communicative actions”[82]. In most collaborations, an imbalance in social dominance is seen as problematic [19]. As dominance is communicative, it is easy to detect, to some extent, across most modalities and has been detected using multimodal fusion of audio and visual data [80, 81].

Another phenomenon explored is Affect [83, 84, 85, 86], the experience of feelings, emotions, or moods [87]. Automated classification of individual emotional states is often combined to approximate group affect and its effects on collaboration outcomes [19]. Emotions are typically distributed across the valence-arousal spectrum. A common method for detecting affect from video is to use facial expressions, such as FACS [88]. Another method is to use text with a lexicon, such as LIWC [89]. Individual emotional states can be aggregated to determine the group’s effect.

Social Cohesion is another aspect of collaboration that is regularly explored [86, 90, 91]. This involves understanding the bonding that shapes each individual’s sense of group membership, and high levels of cohesion are regularly observed in successful groups [91]. A few computational manifestations of cohesion are group-level interest, rapport, attraction, and synchrony. Audio and video features were used to detect group-level interest[92]. Rapport[93] encompasses mutual attention, positivity, and coordination, each of which is best detected visually. Attraction has previously been detected using non-linguistic social signals related to visual engagement, as reported in [94]. Synchrony has been modeled visually by detecting whether the body movements of two people lie in the same plane [95].

A fourth aspect, conflict resolution, involves two or more people seeking a solution to a disagreement. This has traditionally been regarded as necessary, as ongoing conflict will hinder work. Emotion regulation usually plays a significant role in conflicts [83], so emotions are regularly used to model conflict. As a result, similar modalities to affect are applicable. In some cases, robots have been able to detect and moderate conflicts [96, 97].

A final aspect is engagement. This aspect may vary with the task or context. In a general sense, according to Sidner et al., ”it is the process by which individuals in an interaction start, maintain, and end their perceived connection to one another” [98]. In some cases, engagement is

simply attendance; in others, active participation is essential. Key to engagement are interactivity, attention, and the user experience [99]. Visual data has shown strong success [100, 101], but engagement measures should be tailored to the activity.

By combining these indicators, a system can model collaboration dynamics in several ways. To support indicator detection, the final design uses the following modalities: **Visual, Auditory, and Verbal**. Within these modalities, a set of indicators supports the detection of collaboration dynamics. The indicators can help us detect the ground truth of what is happening in collaborations. With a detailed, timely assessment, teachers can receive real-time feedback to help them mediate and improve their students' collaboration [13].

2.5 Collaboration Feedback

Previous work highlights various approaches to developing students' collaboration skills across age groups [102, 103, 104, 13, 105, 106], using a few standard techniques. One shared idea is the need for explicit instruction in collaboration skills, as many students are unsure what constitutes good collaboration [107, 13]. Students also hold misconceptions about what constitutes a good collaboration, and these misconceptions must be addressed to foster the development of effective collaboration skills. The tasks themselves cause some of these misconceptions, as not all tasks that involve working with others are collaborative; some are cooperative [108, 107]. Tasks must be designed to be collaborative to foster the development of collaboration skills [60].

Collaboration requires mutual goals, collective work, and shared understanding rather than only split roles [13, 8, 107]. The different parts of collaboration can be supported by an expert to facilitate effective collaboration and develop collaboration skills. Still, this support must be moderated, as prior studies have shown that excessive teacher intervention can hinder collaboration [64]. This underscores the need for structured, limited interventions or feedback [60] that do not distract from or overshadow active collaboration.

Historically, teachers have been able to provide feedback because they already know students and are physically present. Teachers would need to attend to all groups at all times and would have

difficulty remaining available, even to intervene. Modern Learning Analytics (LA) has sought to automate feedback and the detection of groups in need of support using computational models. One model, CPS Coach 2.0[15], provided an intervention after a 10-minute collaboration to support collaborative skills development, using a Generalized Competency model[61] to detect collaborative states.

Previous studies have shown that students require explicit instruction in collaborative skills within a structured environment to further their development of collaboration literacy. Students can then be supported with structured and limited feedback and interventions to develop their collaborative skills. This information must also adhere to several design principles to support effective skill development. Some design principles have been successful in other cognitive learning studies and should help validate the current design. By analyzing these works, I've developed an understanding of the current design goals for many collaboration analytics systems, which will be integrated with insights from the survey analysis in the prior chapters.

CPS Coach2.0 offered reconfigurable interventions that provided feedback to improve collaborative skills. The feedback display showed overall CPS performance. The facet and indicator overview display shows detailed information for a target facet or indicator. The compare-and-contrast display allows users to compare CPS indicators across a few hypothetical individuals. Lastly, the sample Collaboration coaching tests reinforce knowledge of targeted facets by engaging students as examiners of a collaboration. In this work, the authors highlight several design principles from cognitive research on learning that reinforce the design [109, 110]. This included **Objective Feedback, Formative Feedback, Observational Learning, Interactive-constructive-active-passive(ICAP) framework, and Instructional scaffolding**. These are useful and validated design principles that can be used to ground the system's design.

Objective feedback is seen as a necessary condition for learning [109] while formative feedback supports improvement rather than only evaluation [111]. Objective feedback provides a foundation for developing collaboration literacy, while formative feedback ensures the feedback is framed as a new learning opportunity. This involves ensuring that each student receives individualized, private

feedback rather than collectively. Examples and comparisons with an expert can support observational learning [112]. This means that any feedback should include multiple examples to promote more profound reflection, comparison, and assessment. The ICAP framework [113] posits that interactive and constructive processes are more beneficial than active and passive processes. This means not just being told what to do or take an action, but also thinking about why that action is necessary or how it affects their collaborations. This framework suggests feedback should encourage new cognitive processes through information retrieval, error correction, comparison and contrast, and the generation of open-ended questions. Instructional scaffolding [9] is an approach used in a few other works [114, 14, 71, 115, 116, 117] to provide automated feedback to students following collaborative sessions. This involves incrementally providing more feedback as students slowly master progressively complex sub-collaboration skills. The initial feedback amount decreases until students are asked to explain collaboration dynamics, after which they are increasingly asked to provide explanations.

A similar set of design principles can be seen in Bridging the Gap 2021 [69], which emphasizes feedback that is: **Fundamental, Learner-centered, Actionable, Comprehensible, Computable, and Theory-grounded**. These design goals were used to define high-level collaboration constructs to support the development of collaboration literacy. The constructs were further used to inform the features required to measure them. These principles can also help to guide the system design. Fundamental information is needed for successful collaboration. The following design principle, learner-centered, means presenting information about each individual's behavior rather than the group's complex behavior. Designing actionable feedback details aspects that students could deliberately and readily change. Comprehensible means to ensure the information is relatable and transparent to students. Computability requires that features be computable from low-level multimodal features. Finally, theory-grounded mean features are supported by theoretical and psychometric frameworks. The design goals I've covered will serve as a foundation and support the system design.

Chapter 3

MEASURING STUDENTS' COLLABORATIONS IN REAL-TIME

The most effective teachers tailor their content to learners' levels of understanding. This means guiding learners to progressively build conceptual models for themselves to aid their integration of new knowledge [110]. This means that to support students' development of collaboration skills, an educator must first assess students' current skill level and their conceptions of collaboration. To support this goal, this chapter describes the design of a prototype system called the Chemistry Dashboard. The design focused on helping teachers manage growing classroom sizes by measuring and providing collaborative analytics data to drive teachers' collaborative feedback and interventions.

Work in this chapter was published in ICMI 2019 as *Chemistry Pods: A Multimodal Real Time and Retrospective Tool for the Classroom*[24]. The initial development of the system was conducted in partnership with Sony, as they sought to assess the marketability and viability of a tool to help teachers measure active collaboration. This allowed me to split the system's work initially. Sony focused on UI design, while I focused on implementing and deploying the features required for collaborative analytics. This enabled the rapid development of a high-quality prototype, and it was also helpful to have access to a broader network of prospective user testers. The setup allowed me to focus on identifying and testing the required models for this system, enabling me to study **RQ1**: which features are necessary to measure college students' collaboration.

Prior works on LA and collaboration provided a valuable starting point for defining the different aspects of collaboration. By breaking down and defining collaboration into more measurable components, the indicators and features of collaboration can be determined by measuring these aspects. Works like those by Roschelle and Teasley [6] and, more recently, the PISA 2015 assessment [8] help define these areas, which can be broken down into features to be measured using machine learning and NLP. These features would be added to the prototype to test the functional

viability of combining these metrics with near-real-time transcription and to determine how to create an easily deployable system for teachers to integrate within their classrooms. This included whether a multi-device setup could perform effectively in small to moderate-sized classrooms. I was also able to gauge students' acceptance of the system.

3.1 Background

Collaboration has been measured in various ways over the past few decades as our academic understanding of collaboration has evolved in tandem with advances in knowledge about human cognition. Measurements collaboration can often be facilitated through formal assessments that rely on understanding the joint problem-solving space that collaborators construct through their interactions to coordinate and solve a shared problem together [6]. Assessments of collaboration can be conducted in several common ways, such as experts coding and rating a collaboration, relying on peer assessments, or evaluating based on a deliverable [118]. Over time, various assessments have been developed within the academic community to assess students' collaboration skills.

One that has provided a basis for many other assessments is the PISA 2015 assessment [8], which used a matrix to define and organize a set of skills related to collaborative problem-solving, based on the experience and knowledge of Learning Analytics researchers. Some of the metrics the assessment is interested in include: discovering perspectives or abilities, building a shared representation, negotiating the meaning of the problem, or communicating with team members about the actions being performed. Another Assessment that has influenced many current assessments is the ATC21S method [59], which similarly defines sets of organized collaborative skills and adds indicators for each skill to detail how to detect them in an active environment. With the development of machine learning, many of these features can now be detected quickly using modern MMLA techniques, including LLMs such as GPT-4, Gemini-2.5, and Llama, which excel at high-level textual analysis [119]. This would require converting the audio to text.

The growth of machine learning over the past decade has been driven by the increasing availability of computing power and the growing availability of large datasets necessary to train and run

deep learning models [120]. This has allowed for the deployment of complex systems trained to process audio, video, text, and other modalities to detect emotions [121], generate media [122], or detect or classify other important information [123, 124, 125]. A combination of machine learning models can be used to detect the variety of features associated with collaboration [15]. Some modern systems have attempted to integrate these models.

Looking at a recent system, Collaborative problem solving (CPS) Coach2.0 [15], which provides real-time feedback to students through automated computational modeling, aligns with my design goals. CPS Coach2.0 focuses mainly on remote collaboration and analyzes only language-based indicators. This tool was designed for multiple contexts but was tested with a two-dimensional game called Physics Playground. This involved multiple levels for exploring different introductory physics concepts interactively, with collaborators working together to progress through them. The collaborations were analyzed by transcribing verbal conversations into text and training a BERT-based model [126] to detect manually coded collaboration facets.

Specifically, BERT was prompted to detect indicators of the facets of collaboration. These facets included the construction of shared knowledge, negotiation/coordination, and the maintenance of team function. Each facet is associated with specific indicators. The data collected to quantify these facets were used to provide feedback and interventions to support students' collaborative skill development. The interventions were intended to help refine underdeveloped CPS skills. To this end, the study observed a clear improvement in students' performance during the first intervention with the system. Follow-up interventions require further analysis to determine how to continue CPS skills development. Similar to CPS Coach2.0, other models have also successfully used language-based indicators via modern NLP techniques, such as group communication analysis[72], discourse analysis[127], and related methods. Linguistic models for measuring collaboration provide the basis for the system's analytics.

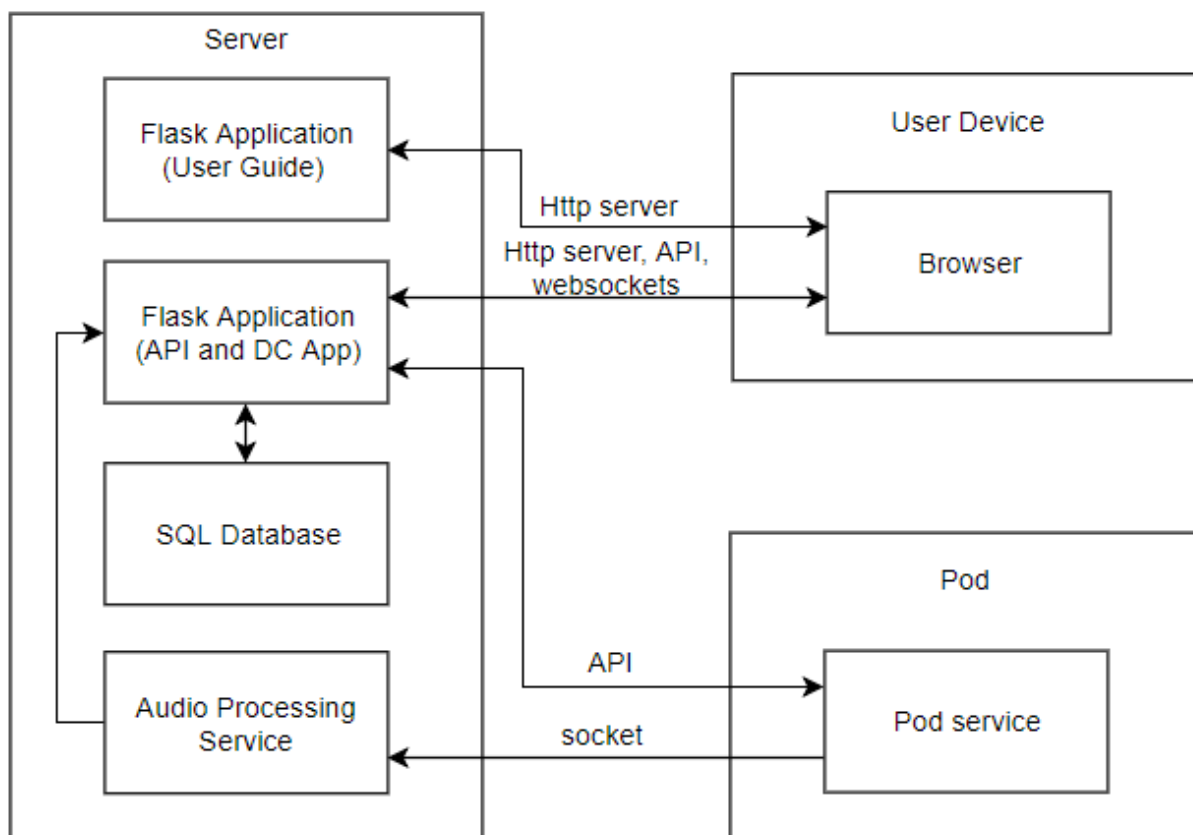


Figure 3.1: System Design of Chemistry Dashboard

3.2 System Design

Here, I go through the major areas of the Chemistry Dashboard's system design. This includes various components: a server for intensive tasks, a user-facing browser interface, and recording devices that connect to the server for processing. The functions performed by the server included processing API data requests, processing audio to compute the desired collaborative analytics, serving browser web pages, and storing the analytics in a database for easy access. SQL Server was used as a database to store the collaborative analytics and other data, including user sign-in information. A Python Flask application was created on the server to process API requests and provide database access. Concurrently, the server runs another Python service, the Audio Processing Service, to process audio data and calculate collaborative analytics. To serve the webpages to the user's device, NGINX is also run on the server, with statically compiled Angular files hosted

there as a source. An overview of how these components interact is shown in 3.1. In the following sections, I will provide further details on each system component.

3.2.1 Flask Application Service and SQL Database

The system used a Flask application server to handle most communication via API requests. This handled communication between the SQL database, the audio processing service, the user's browser, and the Respeaker pod. The server also used SQL Alchemy with Flask to manage the SQL database. The application server contained the table structure, and Flask-Migrate automatically updates the database tables when changes are made. The service consisted mainly of routes that required different levels of authentication to view the results of requests to each route. Importantly, the service had a route for adding, removing, and updating data in the database. The Flask service also handles browser authentication and is where Respeaker connects before starting a session. The SQL database stored all of the user and processed collaborative metrics data.

3.2.2 Recording Device

The design was initially focused on a singular, specific device as the default recording device. This was a multi-microphone array that provided multi-channel audio, improving the accuracy of the analysis. This device would connect directly to the audio processing service. At the same time, teachers could select the desired devices for the discussion from the Angular front-end in a browser.

Initially, two different devices were considered: the ReSpeaker four-microphone array HAT and the Matrix Labs Voice eight-microphone array. Due to the lack of documentation and limited availability of the Matrix Labs Voice, the ReSpeaker was chosen for the initial prototype. The ReSpeaker had a second device, called the ReSpeakerV2, shown in Fig. 3.2, which featured a six-microphone array, providing improved DoA estimation and recording quality. This device was ultimately chosen. A method for connecting a user's device from a web browser was also implemented, but it wasn't the focus of this round of testing. No data was viewable by a device

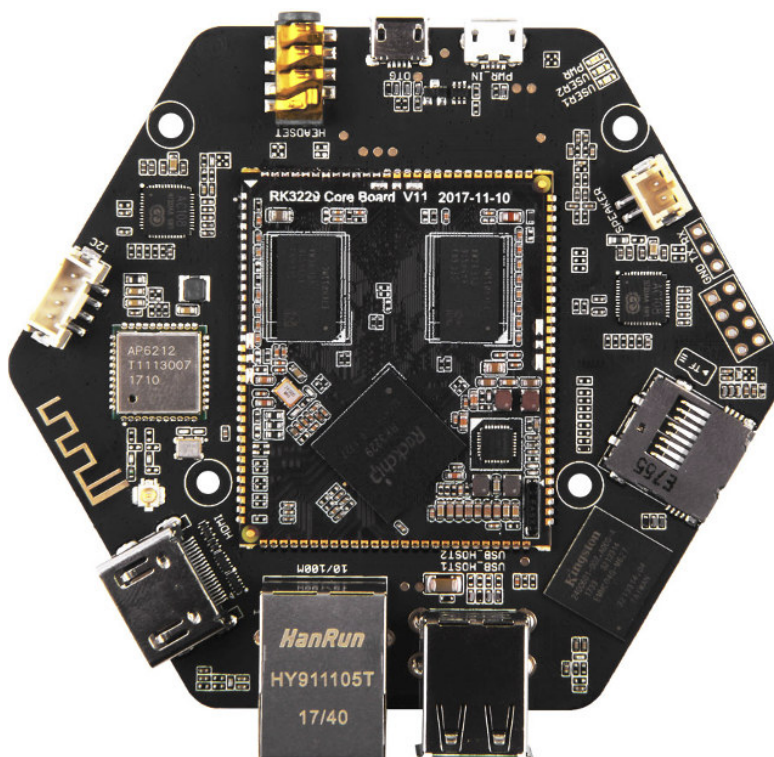


Figure 3.2: ReSpeakerV2

connected through the browser, mimicking the experience of someone using the ReSpeaker device.

3.2.3 Audio Processing Service

The initial features were audio-only. IBM Watson and Speechmatics were initially considered for audio transcription. IBM Watson was initially chosen for its strong performance and its ability to process real-time, streamed audio. GCS also had this capability, but was not initially considered. GCS was eventually selected for its comparative performance and its ability to detect questions, which is vital for tracking the progress of collaborations. Further analysis of these transcripts is possible. Importantly, a robust and accurate transcription is essential for NLP to provide high-quality data on collaborations. Coding of the transcript was performed to provide initial collaboration data.

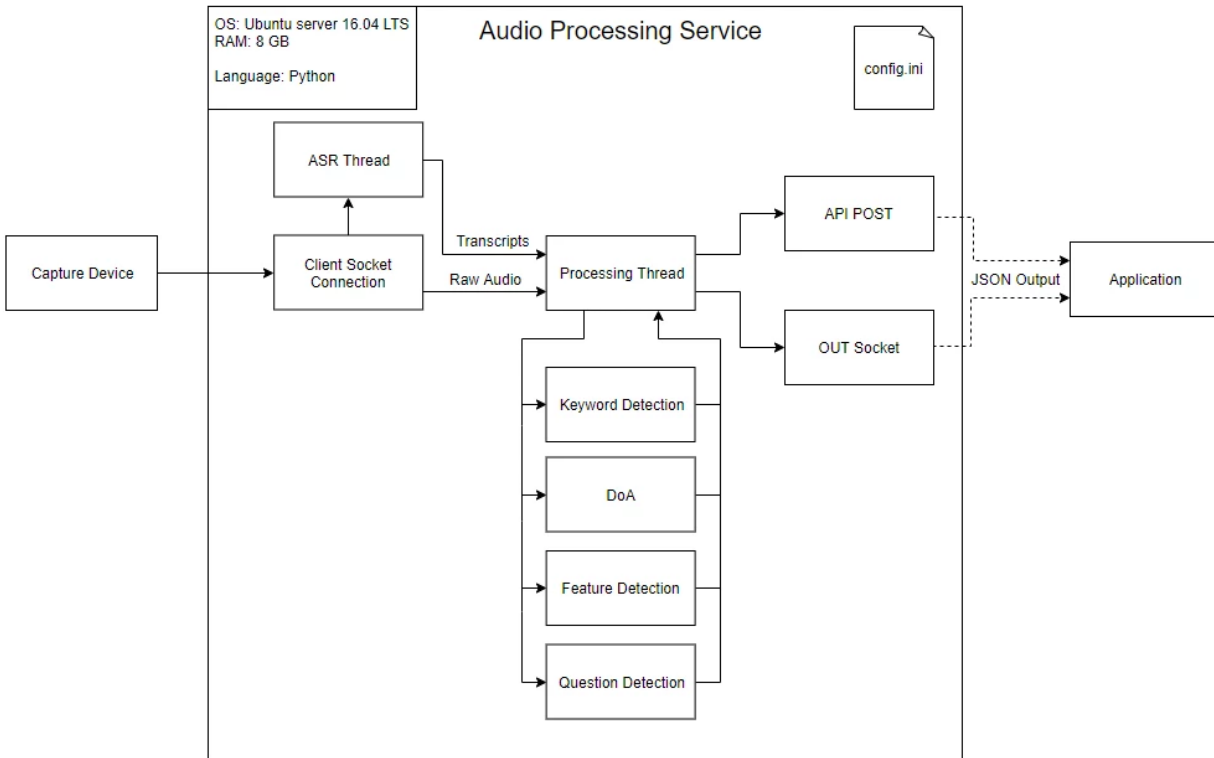


Figure 3.3: Server Architecture of Audio Processor

Direction of Arrival (DoA)

The multi-microphone array of the ReSpeakers enabled us to estimate the Direction of Arrival (DoA) by subtracting the time differences between the arrival times at the microphones. Initially, this analysis was noisy because the data detected non-voice audio segments. Because many speech recognition algorithms perform voice activity detection, the transcription could determine which audio segment to use to calculate the DoA, thereby eliminating most noisy data. This enabled the system to display a sound-source heat map, which could then be used to estimate the number of speakers in the discussion and attribute contributions to them.

Keyword and Question Detection

With a clean transcript, keyword extraction was done to assist the teacher in understanding when students are actively participating versus participating off-topic. Teachers could insert a list of

words representing the current class discussion topics. These words were further analyzed for synonyms, yielding a set of target words to extract from the discussion. These words would be listed and highlighted based on their relationship to the word entered by the teacher. When viewing the transcript, teachers can easily see that a significant portion of the conversation was off-topic or identify what a student may need help with based on the highlighted keyword from their recent discussion. Another approach to segmenting transcripts is question detection. Detecting and displaying patterns in the progression of the collaboration could be helpful. The primary method for detecting questions involved identifying specific words at the beginning of sentences that were associated with questions. GCS's question detection eliminated the need to account for this in the design. This allows for some basic segmentation of the transcripts, which other detailed analyses can further support, providing greater differentiation of the transcriptions and enabling a deeper understanding of the collaboration dynamics.

Psychological Features

Dictionary-based analytics can be utilized to understand the dynamics of collaboration. The dictionary-based approach currently extends two leading psychological research tools: the Linguistic Inquiry and Word Count (LIWC) [89] and the Harvard General Inquirer (Stone, Bales, Namenwirth, & Ogilvie, 1962). Both dictionaries provide between 50 and 200 word categories. These word categories have been carefully selected, curated, and validated using a range of measures in psychological research. To further understand whether a collaboration is productive or progressing, dictionaries can be used to segment transcripts by detailing the likelihood that the speaker is in a specific psychological state. The states I was initially interested in included emotional tone, analytic thinking, clout, authenticity, and confusion. These states could all provide insight into how students' collaboration may be progressing at a surface level. For example, suppose confusion has persisted for an extended period; a teacher may want to review the transcripts to understand what is confusing the group, then engage with the group to alleviate the confusion.

3.2.4 Angular Front-End and NGINX

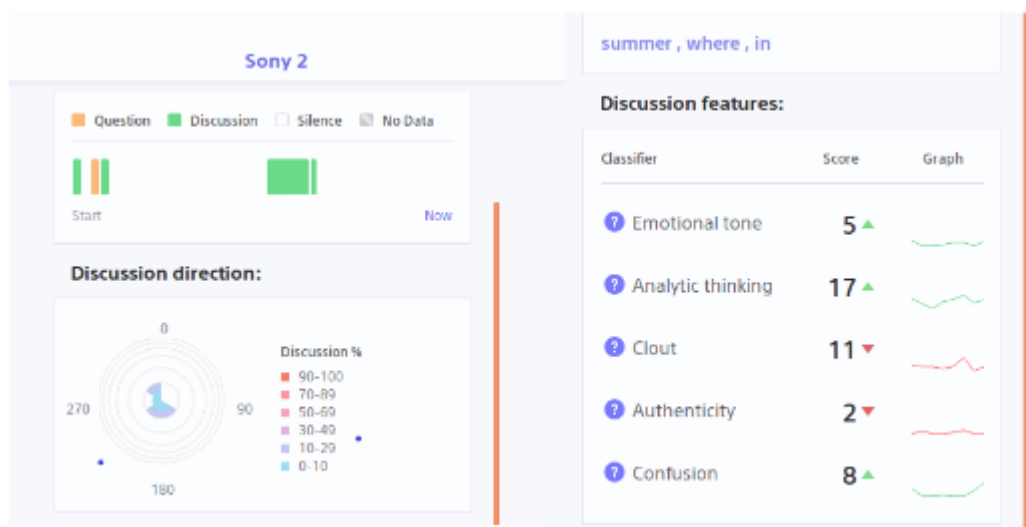


Figure 3.4: The Chemistry Dashboard features a UI that shows the percentage of directional audio, the emotional features, keywords, and more

With these features, a web interface was designed to allow teachers to view the data, manage the device setup, and support students in maintaining productive collaborations. A sample of some of these analyses is shown in Fig. 3.4. First, I wanted to validate that this would be a tool that teachers would be comfortable using. This design enabled teachers to coordinate all aspects of setting up the collaboration feedback system. This initially limited the feedback I received from students, but I was still able to present the data after discussions. This would enable us to assess potential student interest in such a system and determine whether there would be widespread opposition to its introduction.

Within the interface, teachers can set up a single session to coordinate multiple devices simultaneously. A comparison of the devices could be displayed, with each device selectable for a detailed view. Teachers could view a list of keyword occurrences, a timeline of the discussion, questions, silence points, a graph showing the discussion direction by percentage, trends, and the current LIWC text analysis values. It was essential to determine whether teachers could respond to students based on feedback and to assess whether the presence of a recording device affected

students' engagement in collaboration.

3.3 User Studies

In this section, I review studies conducted with the system to inform the iterative design of the prototype. These studies allowed me to validate various components of the system's design and identify ineffective parts.

3.3.1 Initial Studies

The initial user studies aimed to determine the functional viability of the planned design. To this end, initial testing was conducted with a single device in a classroom, followed by testing with 2-3 devices simultaneously in a room. This allowed testing several features of the system. For a single device, testing would help determine how well the system can detect signals in the presence of minimal background noise. While this isn't representative of the main use case of managing multiple groups, it is a crucial step for validating the accuracy and limitations in a simple setting. This was done first in a 180-minute class.

This was followed by multi-device testing with three devices connected, with prompts provided. The first step was to determine the device's sensitivity while other groups were also in the same room. Afterward, I calculated detection accuracy by prompting users to ask only questions. To assess functional viability, User testing was then conducted with a business negotiations class and a small computer science class, with six devices concurrently connected.

3.3.2 Business Negotiations Class

Within the negotiations class, Students were divided into two opposing teams, each working in small groups to negotiate on opposite sides. There were eight groups, and each was recording concurrently. The students were then shown the data at the end of the discussion, after the teacher had reviewed it. This allowed students to engage with data, even though we had not yet developed an interface for direct student interaction during discussions. The testing involved a variety of

system setups to accommodate different settings, as the system had not yet been deployed to a live server.

The procedure to set up and test the system required that I bring a Laptop and a set of ReSpeakers with me to a session. Specifically, a set including extra ReSpeakers was always carried in case a single one experienced technical problems, although this was rare. The laptop hosted the running application, comprising all its different components. Along with the ReSpeakers, I carried a Wi-Fi router that the ReSpeakers were connected to during setup in the lab before attending a session. This allowed us to create an impromptu network in any testing location. The only requirement was a wired connection to an active network. This was done to avoid programmatically reconnecting each ReSpeaker when a new network is connected.

3.4 Results

The initial 180-minute test showed that using voice-activity detection to detect the start and end of utterances was ineffective due to excessive background noise. This meant that audio was continuously chunked and sent to the server for processing. This required a persistent socket connection, which led me to use IBM Watson initially before I pivoted to GCS, which provided question detection and processed streamed audio over a web socket. This testing also allowed me to refine the system setup, as some limitations to the initial design were apparent. In particular, the ReSpeakers experienced connection issues during testing. I refined the setup and added a custom Wi-Fi network to support the devices, which highlighted the system's limitations in its application. Further thinking of how to minimize connectivity issues is essential.

Within the negotiations course, Students enjoyed comparing their perceptions with the data. The system successfully collected data across multiple concurrent sessions, with only minor transcription errors. By showing the student their data after the negotiation, I observed some of their reactions to it and to the system. This provided insight into aligning the system with students. Many students first analyzed their discussion features and were curious about how those values were determined. After explaining how the system analyzed transcripts of their conversations,

many questions arose about how they could directly explore their emotional tone from the audio, such as whether they were yelling or whispering. Students were also curious about how textual tone analysis differed from auditory tone analysis.

3.5 Discussion

In this section, I discuss some of the takeaways from the results. Students enjoyed seeing the data on the screen, although they were dismayed when errors occurred. Overall, most students remained interested in their data after multiple sessions, despite making mistakes. Further testing of accuracy could be conducted to assess how often it incorrectly transcribes. These mistakes can be corrected over time by adjusting the recording device or filtering the audio.

Initially, a basic router with 600 Mbps capacity was used, but in some tests, students were widely dispersed across different rooms. This led to connections failing in specific setups. To address this issue, I began using a Google Mesh Wi-Fi router. This allowed us to wire one mesh device to the local network, while the other mesh devices extended the system's connection range. This worked smoothly for the remainder of this testing round. While this did work, it adds another part to the system. It may be possible to use microphones in phones or laptops already connected to the internet, rather than a custom device; however, this will limit the ability to determine the speaker's direction.

Throughout testing, the system handled a single device as smoothly as it did eight devices concurrently. This was while the system was not running on a dedicated server, but on a 2015 MacBook Pro. This suggests the system could scale well to more devices. More features are planned, so continual testing will be necessary to ensure the system's maintenance, scalability, and ability to handle multiple recording devices simultaneously. Furthermore, it would be beneficial to test the system's capacity on a dedicated server to obtain the most accurate and reliable measurements.

While displaying the feedback to the students, I observed initial reactions to the data and heard ideas about how students might actively use it. Understanding this further could help design the

concepts built into the system to build on college students' conceptual understanding of collaboration, rather than relying solely on scientific and teacher perspectives. Ideally, this would provide students with more actionable information that is easier to digest at their level of understanding. Along with this, I was able to clearly see that this data is of interest to students, and further understanding and analysis of students' perspectives would be valuable for determining how best to engage students with this data to support the development of collaboration literacy.

3.6 Conclusion

In this chapter, I detailed the design of a prototype system to track basic features of collaborations. The system records audio from users' collaborations for analysis. The analyzed data is then shown to teachers to allow them to manage or understand students' conversations. This tool can primarily help teachers manage large class sizes when conducting group-based work. Supporting teachers drove the initial ideas for the set of metrics to be provided to them. The Education and Learning Science literature was practical in offering information and metrics to measure or assess a collaboration. Using a set of desired metrics, machine learning algorithms were implemented to detect them within a multi-part server architecture. The server handled serving the webpage to users, processing the audio, and storing the data.

With this prototype working on basic collaboration features, I can further examine students' opinions about the system-provided data, beginning with how they conceptualize collaboration. The findings showed that we can measure and display to teachers a breakdown of certain collaboration features. By further engaging with the discussion areas in future user studies and design iterations, we can enhance the system's ability to support students' development of collaborative skills. This includes conducting additional user studies to better understand college students' comprehension of collaboration, integrating additional features to fully break down collaboration, and regularly testing the system's ability to support multiple groups simultaneously.

Chapter 4

UNDERSTANDING STUDENTS' CONCEPTUALIZATION OF COLLABORATION

In chapter 3, I designed and built a prototype tool to measure students' collaborations from audio input. In this chapter, I conduct a comprehensive evaluation of college students' conceptualization of collaboration, and the prototype was iterated to better align with this understanding. Work in this chapter was published in the Journal of Learning Analytics as *Designing Analytics for Collaboration Literacy and Student Empowerment* [25]. Within this chapter, a survey was conducted to address **RQ2**: What is college students' comprehension of collaboration?

The survey included a few short free-response questions and a few multiple-choice Likert-scale questions. By engaging students with these questions, I better understood how to present the collaboration information to empower them in developing collaboration literacy. As the previous design focused on a theoretical understanding and teacher-driven collaborative measurement and support, this study allowed me to integrate students' perspectives into the design. This enables collaboration literacy development to be student-driven, an essential component for maximizing the development of collaborative skills.

The surveys were analyzed by coding the short-answer responses using a consensus coding approach, in which a set of topics was selected. Following a grounded theory approach, four coders analyzed 10 responses and, after reaching consensus, narrowed the topics to a specific subset. Coders then coded 300 total responses each, refined categories, and reached a consensus on the code for each answer. The final set of topics was seven major categories: **Climate, Communication, Compatibility, Conflict, Context, Contribution, and Constructive**. The coded responses were further analyzed to identify common trends in students' opinions towards each topic and to determine patterns in how students discuss and understand collaborations. The design of the prototype is then centered on aligning with the trends identified in the analysis to address **RQ5**: What features should be added to BLINC to align with students' goals, comprehension, and concerns?

4.1 Background

Learning Sciences has evolved over the past few decades to explore various approaches to measuring and supporting students' educational progress. This exploration has provided a strong understanding of the processes teachers can use to help students develop their academic skills. A book published by the National Academies of Sciences, Engineering, and Medicine, *How People Learn II* [110], explored the development of learning processes over these decades and highlighted that effective feedback must be immediately applicable to the user. This means that any feedback must engage the student at their level to help bridge the gaps in conceptual understanding. Furthermore, the book emphasized the importance of supporting students' conceptual modeling to enhance their overall knowledge and long-term integration of concepts. Engaging students' conceptual development requires direct interaction with them and an understanding of their current conceptualization of collaboration.

Student engagement in learning collaboration is as essential as the content taught about collaboration, as engaging in collaboration itself has not been shown to improve collaboration skills over time [60]. Studies have demonstrated that explicit, active collaborative instruction shows marked improvement over control groups without instruction. Chen et al. [128] studied college students performing collaborative tasks and provided active intervention to one group while providing none to a control group. The experimental group had higher teamwork knowledge and better expert ratings of their skills. A similar effect is observed in McKinney and Denton [129], where an experimental group receiving collaborative instruction achieved significantly higher peer-review scores than a control group. This means prioritizing students' engagement with the presented information is essential for improving students' collaborative skills. To do this, understanding how to tailor information to engage students most effectively is a priority. The first step in engaging students is to ensure that the analytics align with their conceptualization of collaboration. The rest of the chapter explores this.

4.2 Methods

Overall, there were 131 students in a university class. Most participants (124) were undergraduate students, two were master's students, and five were PhD students. First- and second-year students represented the majority of the pool, with 40 participants in each year. 22 students were in their third year, and 26 were in their fourth year. Two of the graduate students were in their fifth year of study. Individuals received \$5 compensation for finishing the survey and were also entered into a raffle for five \$50 gift cards. In this survey, students were asked a variety of free-response questions about collaboration, which enabled me to analyze their responses and identify trends in their views. This was achieved through group consensus coding, in which four trained researchers coded the student responses using an agreed-upon set of classification topics. Once the responses were organized thematically, the subgroups were further analyzed to uncover deeper insights. The surveys included short answers(Questions 3-8) and Likert scale questions(Questions 1 and 2) and a final other open comments question:

1. How often do you collaborate on classwork in school settings (inside of classrooms, for class)?
2. How often do you collaborate on classwork outside of school settings (work done outside of the classroom, etc)?
3. How would you describe a good collaboration?
4. What are some characteristics of a bad collaboration?
5. How do you deal with poor collaboration experiences?
6. What are some things you've been taught about collaboration and collaborating?
7. What advice would you give a high school student about how to collaborate most effectively in college?
8. In your ideal world, how would you want to learn how to collaborate?

4.2.1 Likert Scale Analysis

To the two Likert scale questions included in the survey — “How often do you collaborate on classwork in school settings (inside of classrooms, for class)?” and “How often do you collaborate on classwork outside of school settings (work done outside of the classroom, etc.)?” — students could respond Never, Sometimes, Regularly, or Everyday. To compare the results from these two questions, responses were coded into ordinal values ranging from 1 to 4 (Never = 1, Sometimes = 2, Regularly = 3, and Everyday = 4). This order aligns with the response options presented to participants. A two-tailed Wilcoxon signed-rank test was used to compare the mean response value, specifically to determine if students reported collaborating at different rates in class versus outside of class.

4.2.2 Free Response Analysis

I began the text analysis by leveraging various computational techniques to mine student responses. For example, I generated term-frequency word clouds. I also employed topic modeling to examine underlying patterns that could inform our analysis. However, I found that, as implemented, these two approaches were insufficient to achieve the level of specificity I wanted in the study. Instead, I drew on techniques from grounded theory [26] to analyze students’ responses to the free-response questions. I supplemented this human coding with a brief text analysis of the most common words associated with each category.

Four team members coded the data. The first step was for individual team members to highlight student responses that surfaced one or more key ideas. From these initial exemplars, I identified a set of overarching themes. These initial themes included broad topics such as Emotion, Communication, and Equal Work. After some discussion, a preliminary set of categories was proposed. This preliminary set included Content, Context, Climate, Communication, and Contribution. Group members met to collectively code a selection of 10 responses taken from different questions. The team reached consensus on those responses, refining the categories to: Communication, Compatibility, Conflict Resolution, Context, Contribution, and Constructive. The Communication category

was combined with Climate, Constructive replaced Content, and Compatibility was added. The four research team members each coded 300 of the 567 total responses.

After individually coding 300 responses, the research team met to reassess the list of categories and proposed a few modifications. Namely, the Climate category was re-added, and Conflict Resolution was changed to simply Conflict. Next, the research team conducted an additional round of coding using updated categories and their definitions. Finally, each team member reviewed all assigned codes for each response, adding and removing codes until the research team reached consensus on all student responses. Hence, the data reported in this chapter comprises the codes that the entire research team agreed applied to each student's response.

4.3 Results

We received completed surveys from 87 of the 131 participants; however, partial responses are also included within this analysis. The breakdown of responses is shown in Table 4.3. The two Likert-scale questions about the contexts in which students practice collaboration provide the starting point for this analysis. Students were independently asked about how frequently they collaborate in classrooms and outside of school. Students could respond Never, Sometimes, Regularly, or Every Day. On average, students reported collaborating more frequently outside of class than in class ($Z = -3.67$, $p < 0.001$). This, as I discuss later, has significant implications for the design and deployment of collaboration analytics technologies.

Dimension	Percentage
Communication	30%
Contribution	22%
Constructive	22%
Compatibility	12%
Climate	9%
Conflict	6%
Context	6%

Table 4.1: Table of Responses Coded to each Dimension

Moving now to the free-response questions, our iterative coding process identified seven over-

arching topics, which are called **The 7 C's of Collaboration**. I suggest that these topics reflect how students perceive and experience collaboration. These topics include **Climate, Communication, Compatibility, Conflict, Context, Contribution, and Constructive**. The research team reached consensus that these seven topics encompass the broad categories students mentioned in response to our survey questions. As I describe each category or dimension, it is essential to note that they are not meant to be mutually exclusive or orthogonal. On the contrary, as the reader will see, many of the categories reflect, or are influenced by, other categories. Even so, each category does appear to have several characteristic sub-dimensions. I do not consider these sub-dimensions predictors of collaboration quality; instead, they reflect how students may perceive and interpret a given collaborative experience.

Our presentation of each dimension will follow the same format. First, I provide a brief description of the category, followed by a list of terms that most frequently appear within responses coded under this dimension. Next, I include an exposition of the different sub-dimensions of that category as evidenced through one or more quotations. Finally, I discuss whether the given dimension is addressed in prior research on collaboration and learning analytics.

4.3.1 Climate

The Climate dimension refers to emotional or affective considerations of the collaboration. When I examine the most frequently occurring words in responses that received this code, a few exemplify the idea of Climate: ideas, open, help, try, and respectful. Notably, students are considering how openly they can share ideas in a respectful and welcoming environment. For example, one student wrote that bad collaboration can be characterized as an “overall toxic atmosphere.” Another student described it as “Feeling like I want to withdraw. A place where I don’t feel welcome.” In contrast, they described good collaboration as a space where “everyone is given a fair chance to voice their opinion and does not feel scared to give that opinion on the basis that others will make fun of them.”

Consistent throughout this category is the realization that feelings, emotions, and affective

states are integral to a collaborative experience. Our Climate category is somewhat similar to psychological safety. Edmondson and Lei (2014) [130] define psychological safety as the level of comfort people feel taking interpersonal risks within a given context. Psychological safety is associated with several measures of team performance and organizational change.

Outside the psychology community, several LA researchers have discussed the inclusion of emotion- or affect-related measures in studies of collaborative groups. For example, Vrzakova et al. 2020 [41], Worsley and Blikstein 2018 [131], and Furuichi and Worsley 2018 [132] all use electrodermal activation wristbands that can serve as proxies for feelings of stress or arousal. These measures, however, can be challenging to disentangle from correlated measures, such as cognitive load. Others have utilized text mining, albeit post-hoc, to surface positive and negative sentiment in spoken utterances [133]. A primary challenge, then, is how the field might incorporate a more complex set of affective states without being subject to the biases present in many tools for perceived-affective-state recognition [134]. Additionally, there is the challenge of distinguishing between the various multimodal signals and the corresponding felt emotions.

4.3.2 Communication

The Communication dimension refers to the quantity and quality of information shared among group members participating in a given collaboration. Terms frequently used in responses labeled as Communication included: communication, everyone, talk, share, open, and visualization. The words “communication” and “talk” highlight how this category generally encompasses the simple act of talking to one another when problems arise or when people are uncertain about next steps. In a more concrete sense, responses coded as Communication often included people establishing and being aware of expectations. For instance, one student wrote, “Constant communication is important to make sure everyone is on the same page about goals and expectations.” Another student writes, “Everyone must be open to compromise, and people should try to have open communication.” Open communication also touches on another key idea: listening. Connecting these two ideas, one student wrote that “Collaboration requires active listening and understanding from both

sides.” Finally, several responses also mentioned using visual representations of data to communicate the information to students.

Communication is the area where prior research has been most prolific. Many projects have endeavored to quantify talk using speech signals [135, 136] or chat-based messages [41]. These systems can approximate speaking turns and use body language to proxy for participant attention. For instance, Vrzakova et al. 2020 [41], Worsley and Blikstein 2018 [131], and Cukurova, Zhou, Spikol, and Landolfi 2020 [137] use head pose to identify moments when participants’ attention appears to be directed away from their collaborator. These existing platforms have done this as part of post-hoc analysis. Other research explicitly measures visual attention using mobile eye trackers [36]. Albeit prohibitively expensive, mobile eye trackers can monitor participant attention.

4.3.3 Compatibility

Compatibility refers to the extent to which group members use complementary strategies in work and interactions. Common compatibility words included: working, together, team, help, and friends. As these terms suggest, students assess the effectiveness of group members’ collaboration and support. Other responses noted ways that students may have “compatible work styles.” Importantly, however, students did not view compatibility as static. Instead, some indicated an awareness that “team building and bonding activities give people a sense of closeness and community.” The term “friend” arose from student responses that highlighted the advantages of working with and without friends. In advising younger students, one respondent wrote, “Find people who work well together; these people may not necessarily be your closest friends.” Another student advised, “Be friends with your teammates.” The view suggests that students consider friendship integral to effective collaboration.

To our knowledge, prior research has not focused on compatibility as a key topic in education collaboration or collaboration analytics. Research on pair programming [138] and group formation strategies [139] is likely most relevant to this topic, but these works rarely explicitly discuss team building. There is a complementary body of research on team-building strategies, but I scarcely see

this work discussed within the collaboration or MMLA communities. Part of this may be related to the contexts I typically use to study student collaboration, which I discuss in more detail shortly.

4.3.4 Conflict

The Conflict dimension encompasses students' approaches and needs for responding to challenging or contentious situations that may arise during a collaborative experience. Some of the most common words associated with this category include: conflict, different, bad, team, and professor. One student's response, in particular, seems to encapsulate frequently mentioned approaches for dealing with conflict: "I usually switch up which people I am collaborating with if possible. If not, I will confront the responsible people in the group, or else just deal with it." Hence, the term "different" was often used to describe a common approach that students take when faced with conflict: "finding a different team." In addition to exiting the collaboration, students discussed ways to confront the responsible individual(s), including "just dealing with it" and bringing their concerns to the course instructor or an external mediator.

On this topic, one student voiced that, in the ideal world, they would like "having experienced educators facilitating collaboration would be nice, someone being there to make sure everyone is heard and gets a chance to speak, also sort of a conflict mediator." Another student similarly noted, "If conflict arises, it would be helpful to have someone more experienced mitigate this to show how issues can be resolved." This is a significant observation because the student is not satisfied with simply having the conflict resolved. Instead, they are interested in learning how to address such challenging situations.

Existing research on classroom orchestration (Dillenbourg et al., 2011; Roschelle, 1999) begins to scratch the surface of this dimension by alerting instructors to students' needs. Other research has documented the emergence of conflict in collaborative settings (Jung, Toprani, Yan, & Borge, 2017; Roschelle, 1999; Barron, 2003), yet little of the leading research on collaboration analytics considers conflict as an essential dimension of the collaboration landscape. Put differently, researchers have approaches for identifying when conflict has arisen, but these are seldom included

within collaboration analytics tools, nor are tools developed to support conflict resolution.

4.3.5 Context

Context refers to who participates in a given group, why they participate, and where the collaboration takes place. Some of the words associated with Context responses include learn, experience, problem, different, and taught. “Problem” refers to learning collaboration on real-world problems. “Learn” and “taught” are emblematic of students wanting to learn collaboration more explicitly. One student notes, “In an ideal world, there would be a mandatory class teaching students how to collaborate using research-backed ideas.” Part of what this student suggests is that collaboration is never explicitly taught as an end goal but is instead treated as a secondary skill. For instance, one student responded, “I would want to learn general principles in collaboration through some form of class/lecture, but focus more on actually collaborating with other people and learn from experience.” Context also addresses questions of why and where students learn to collaborate.

Students consistently expressed a desire to learn collaboration across various settings. For example, a student responded, “I think [collaboration] is best learned by doing it both in class and out of class.” Another student echoed that sentiment but proposed another dimension of Context that is neither explicitly in-class nor explicitly out-of-class. The student wrote, “I would want to learn how to collaborate in a setting where I feel comfortable, since sometimes I am shy and being forced into collaboration can be counterproductive.” This is another important consideration, as the Context can often influence people’s willingness to contribute. To this point, one person noted that they would ideally “[a]lways do it with friends, where you can communicate your needs and ideas comfortably. Strangers are frustrating.” The final sub-dimension of context concerns how actions and interactions in a collaborative setting may be interpreted. One student’s comment makes this point explicitly. The student wrote, “Oftentimes, outside perspectives on group collaborations do not reflect the opinions of the collaboration members.” This student is concerned with how an external person or tool may interpret what is happening within that group. Namely, that an external evaluator may make judgments about a group that reflect a lack of understanding of the context for

those actions. Accordingly, the context for the collaboration has important implications for how these tools and research studies are conducted. It also has implications for how I choose to surface information to collaborators or instructors.

Regarding Context, much of the existing conversation within the MMLA community has focused on the diverse settings in which collaboration occurs. Compared to some of the early MMLA research that looked at a very decontextualized mathematics learning environment (Oviatt & Cohen, 2014), more recent work has begun to move outside of laboratory environments and into Fablabs and Makerspaces (Schneider & Blikstein, 2015; Spikol, Avramides, & Cukurova, 2016; Worsley & Blikstein, 2018). In many of these studies, students are solving open-ended learning problems with tangible materials and microcontrollers, which necessarily shifts the context away from a mandated set of tasks and towards self-directed tasks. Other work has shifted from traditional educational settings to hands-on healthcare learning environments (Martinez-Maldonado et al., 2017). Less common, however, are studies and platforms designed for user-selected contexts. This need for user-selected contexts is corroborated by student survey responses, which found that they reported spending more time collaborating outside of class than in class. Ochoa's multimodal selfie (Domínguez, Chiliza, Echeverria, & Ochoa, 2015), a mobile device equipped with a camera and a microphone, could meet these cross-contextual needs. However, it was designed with a much more individualistic perspective in mind.

4.3.6 Contribution

The Contribution dimension highlights what individual participants are, and are not, bringing to the collaboration. Among the most common words for this dimension were work, ideas, contribute, make, time, effort, and project. Students discussed their contributions across two main sub-dimensions. First, examples of responses to verbal contributions to group discussions included new ideas and problem-solving solutions. At times, these were with people adding ideas. In other instances, the issues involved discredited ideas. For example, a student discussed poor collaboration, noting that “[o]ne person dominating, ideas being shut down.” In other instances,

contribution is more about “work” or “effort” toward a group project. To this point, another student wrote, “A good collaboration involves all parties contributing equally to complete a task to a degree that satisfies all parties involved.” These two examples highlight the presence of people over- and under-contributing, themes that pervade the survey responses.

Much of the prior research on collaboration and collaboration analytics examines this dimension (Borge & Rosé, 2016; Tse et al., 2007; Worsley & Blikstein, 2018). Students’ verbal and gestural contributions to a given collaboration are quantified and represented, allowing them to reflect upon their contributions (Praharaj et al., 2018; Dillenbourg et al., 2011). In collaborative problem-solving, researchers may wish to identify who contributed to the eventual solution or idea that the group ultimately adopted. Aside from dependent measures, assessing student contributions has been a driver of collaboration analytics.

4.3.7 Constructive

The Constructive dimension relates to how students conceptualize or discuss the overall goals of collaboration. Some of the words commonly used in responses about this dimension include ideas, team, goal, help, make, and better. This dimension is typically aligned with the concept of team progress towards individual or group learning goals or project outcomes. To this point, one student wrote, “Good collaboration is mutually beneficial. Every individual involved in the collaboration has something unique to contribute, and the net result is that every person gains insights that they would not have had on their own through the collaboration.”

The end goals of a collaboration can vary widely based on many of the other dimensions previously mentioned, particularly Context. Much of the prior research on collaboration in engineering design contexts has focused on documenting the progress of project teams towards a specific final project. This work will often involve documenting the different phases of collaboration a group is in, based on the types of discussions it is having. For instance, Gero and Kannengiesser’s (2004) work on Function-Behavior-Structure mapping has been applied to the study of groups during collaborative problem-solving. The orientation towards products or outcomes is also evident across

other research on collaborative problem solving (Hmelo-Silver & Barrows, 2008; Praharaj et al., 2018). Many prior research studies have the objective of improving student learning or helping them produce better artifacts (Schneider et al., 2018; Worsley & Blikstein, 2018; Spikol, Superiore, Anna, & Landolfi, 2020; Martinez-Maldonado et al., 2017), but are not necessarily concerned with team dynamics or how much each person is contributing.

4.4 System Development

At a high level, the BLINC platform comprises a distributed system of wireless audio/video data-collection devices, a centralized server, and a secure web-based interface for real-time or post-hoc data access. The tool is designed to be useful for researchers, practitioners, and students. Embedded in this tool are capabilities that can provide valuable, actionable information to different stakeholders. I will present a general overview of the platform, followed by a description of its core capabilities and their alignment with each category. As I discuss each feature, I also describe ways that it might be used to support collaboration literacy. I recognize, however, that effective use of these features for collaboration literacy requires more than simply including them in the interface. Instead, there needs to be a complementary set of intentionally designed student experiences that incorporate this data.

4.4.1 Overview

The participant-facing component of BLINC consists of four main views. All of these views are user-specific and require users to log in to their accounts to access the data. The first view after logging in is a user-specific home screen that lets the user adjust configurations (Figure 4.1). Some configurations include defining relevant keywords, uploading documents to construct topic models, configuring WiFi settings, and organizing previous data-collection sessions. From this home screen, users can view previous data-capture sessions or start a new one by selecting the Discussions button.

After clicking the Discussions button and starting a new session, users can specify keywords

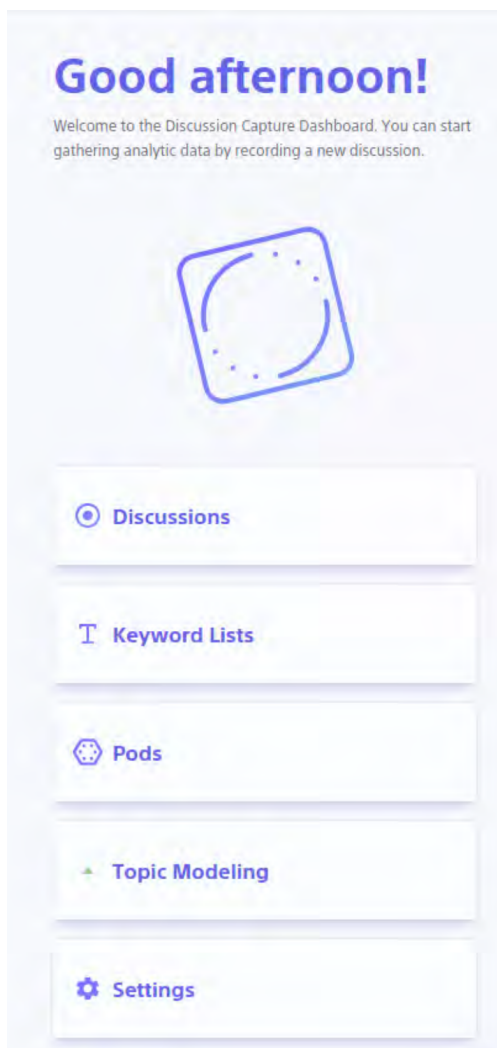


Figure 4.1: Home Page

and a topic model to associate with it. Users can then begin synchronizing data collection across all available devices. This includes microphone arrays running custom software to process multi-dimensional audio and video data, or smartphones simultaneously logged into the BLINC interface via the user's account. After selecting the appropriate devices and initiating data collection, users are directed to a screen that displays real-time data from each device. This group-specific data is presented using four primary representations. The data depicted within these representations can be adjusted using the Timeline control at the top of the interface. Adjusting this slider allows the user to specify the time interval they want to review.

The first representation is a discussion timeline that approximates the times when the group was

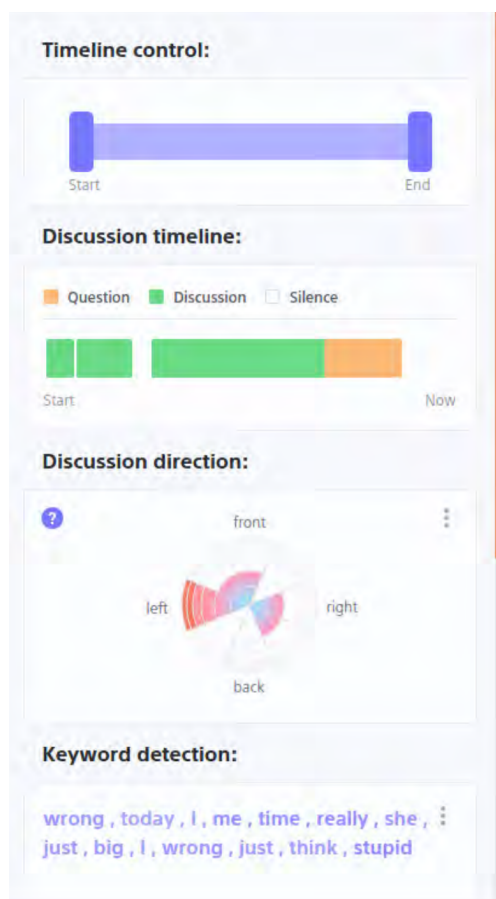


Figure 4.2: Group Discussion Data View

discussing, asking questions, and being silent (Figure 4.3). This data is derived from automatically transcribed audio data.

The second representation, called “Discussion direction,” provides the direction of audio arrival captured by the microphone array (Figure 4.4). This representation can offer an approximate distribution of participants’ talk within the group. This representation can be adjusted to different levels of granularity. For example, it can divide the 360-degree space surrounding the microphone array into 2, 4, 6, 8, or 12 equal-sized parts.

The next data representation shows the keywords stated (Figure 4.5). Recall that the desired set of keywords was for users to select before starting the discussion. Each occurrence of a keyword is highlighted within the “Keyword detection” component.

The fourth representation is a customizable set of sentiment classes, labeled “Discussion fea-

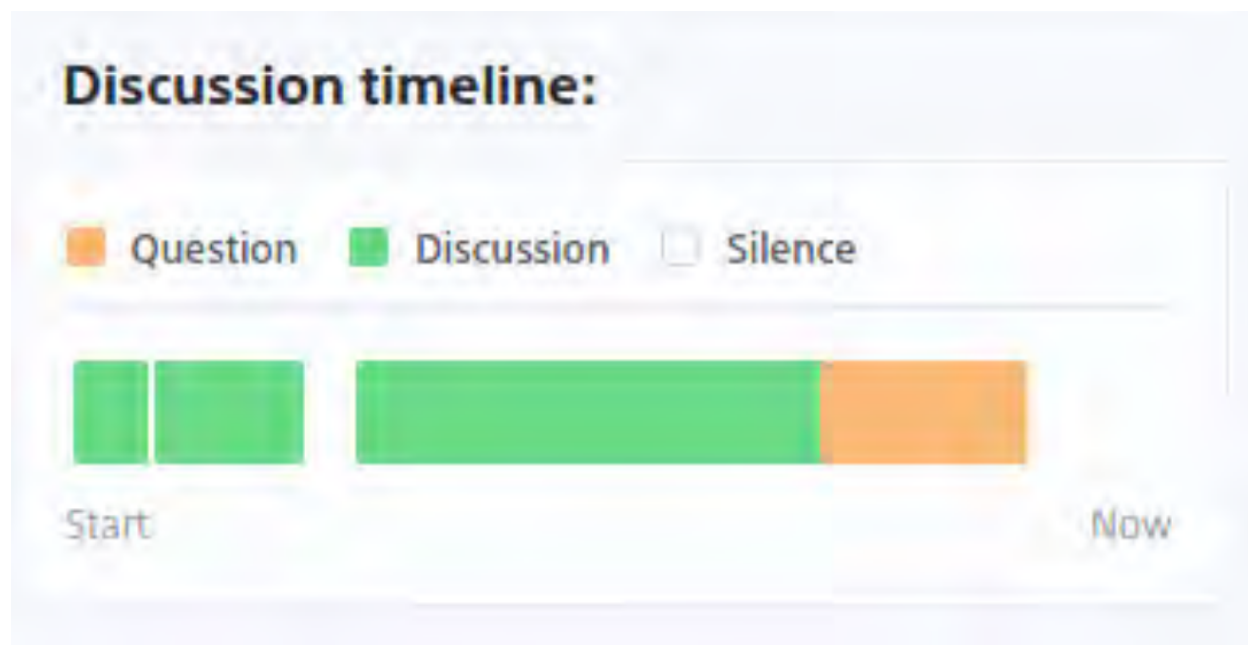


Figure 4.3: Discussion Timeline

tures” (Figure 4.6). This component appears just below the keywords at the bottom of the main view. Automatically transcribed audio is used to conduct fundamental sentiment analysis by leveraging extended forms of the Harvard General Inquirer (Stone, Bales, Namenwirth, & Ogilvie, 1962) and the Linguistic Inquiry Word Count (LIWC; Tausczik & Pennebaker, 2010). Hovering over any of the sentiment categories provides a tooltip that specifies which words are associated with that specific sentiment class.

Across all four representations, users can click on desired items. They will be taken to the fourth primary view, which includes a complete transcript of audio session data from that specific audio capture device (Figure 4.7). When entering the transcript view from any of these four representations, the transcript immediately loads into the corresponding context of the selected keyword, sentiment, direction of arrival, or question segment.

The transcript view is color-coded by audio direction of arrival and also highlights keywords in the transcripts. The transcripts can also be labeled with the detected speaker using speaker diarization. The inclusion of anonymized speaker labels or color-coded direction-of-arrival information provides pertinent, individual-specific information, even when participants use a smartphone rather



Figure 4.4: Discussion Direction

than a microphone array with direction-of-arrival capabilities.

Given this general overview of the BLINC platform, I now discuss how it connects to the different dimensions I covered in previous sections.

4.4.2 Refinements

As previously noted, specific design components can align with multiple dimensions. In this section, I analyze these alignments in the different components.

Communication and Contribution

When I examine the various representations, particularly the discussion timeline and the direction-of-arrival chart, I see how they can support questions about Communication and Contribution. Both allow the user to quantify the amount of communication within the group, both collectively in the timeline and on a more fine-grained basis in the direction-of-arrival chart. These components also



Figure 4.5: Customizable Keyword Component

include information on the types of verbal contributions and, more broadly, on their distribution across the group.

The transcript view similarly lets users see the type of communication and the origin of each contribution. By following along with the transcripts, users can also see how much each verbal contribution is taken up in subsequent conversations.

Constructive

The different representations also align with the Constructive category, as they provide information that can be used to review a group's stated goals and expectations against the timeline and to assess how well the group is progressing towards those goals. Keyword representations can be used to determine how well students are using domain-specific language, for example, or, at the very least, discussing content relevant to the activity at hand.

Context

The salience of context is evident across various platform features. First, the home screen enables users to specify keywords and topic models used in the backend, thereby enabling analytics to be

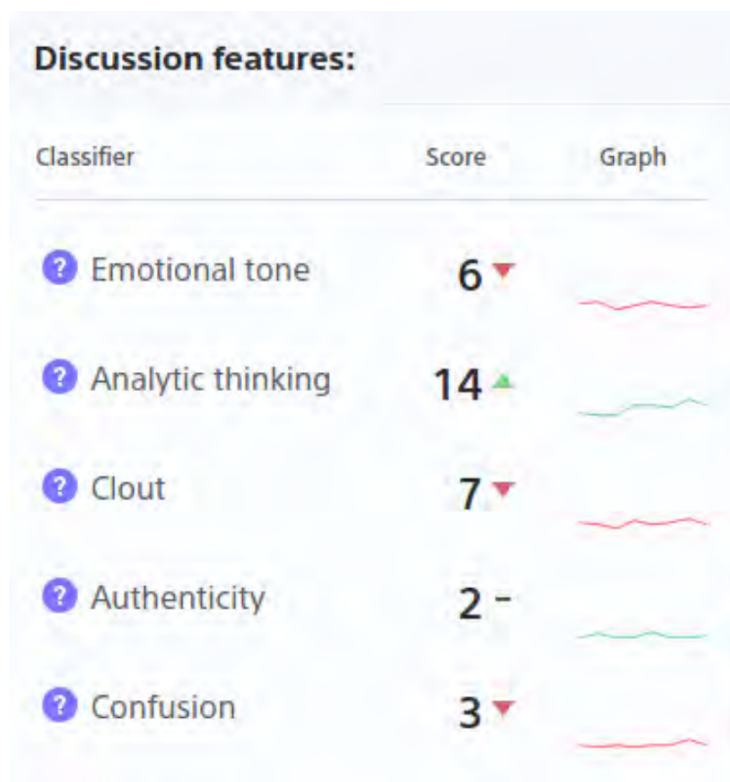


Figure 4.6: Discussion Features

easily adapted across multiple disciplinary contexts in which students may practice collaboration. The platform also uses contextualized data representations, allowing users to specify the data interval to display and to drill down to specific utterances in which a given keyword or question was used. This closely aligns with Echevarria's suggestion to contextualize information. Additionally, the platform is intentional about not presuming what constitutes a good or bad collaboration. Instead, it provides numerical and graphical data that users can interpret in the context. Perhaps a more important feature is participants' ability to capture collaboration data on their own terms. The ability to collect and receive real-time analysis in any internet-enabled space using a smartphone enables students to examine their collaboration data outside the classroom. Based on Likert-scale responses and student quotations, students considered this helpful.

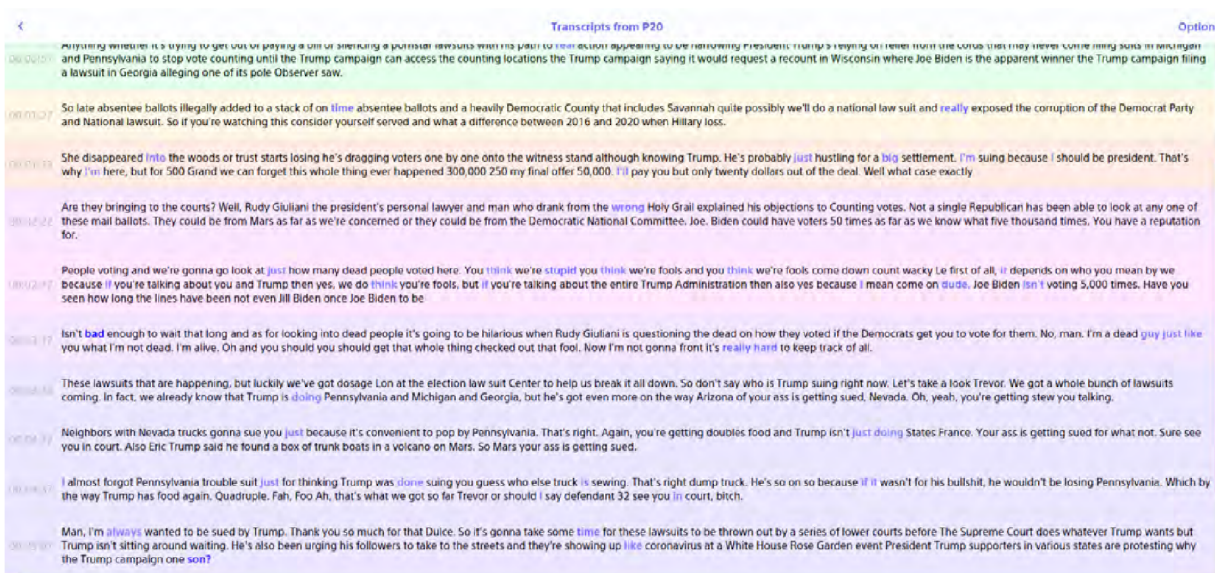


Figure 4.7: Discussion Transcripts color-coded based on DoA

Climate

Regarding climate, the text-based sentiment analysis features enable both students and teachers to gain insight into the emotional tone of group interactions. This information could be helpful in real-time to redirect the group conversation to address concerning behavior. They may also use this information post-hoc to better understand how climate changes propagate across different group participants and for students to reflect on how these changes affected their participation in the group. As the survey respondents described, this is often an important dimension that is not captured by examining only the content of a given group's discussion or the quality of the work product.

Compatibility and Conflict Resolution

The current BLINC implementation supports dimensions of compatibility and conflict resolution by combining transcript and sentiment data, enabling teachers and students to reason about group dynamics and assess how well students participate in conflict resolution. For example, teachers might use transcript information to see how well different students' ideas are taken up by their

peers. The direction-of-arrival chart also provides a high-level view of how well a group is working together and/or the different roles students are taking on within their groups. In terms of conflict resolution, the discussion transcripts, while still anonymous, can help external facilitators intervene appropriately and effectively when conflicts arise. It also provides an external individual with important contextual information from the prior conversation, making it a valuable tool for engaging with the group.

While I do not enumerate each feature of the BLINC platform and how it aligns to the goals of collaboration literacy, the information above should have highlighted three key ideas: 1) there are many ways to instantiate the different dimensions within a collaboration analytics platform; 2) a single interface design component can feasibly address multiple dimensions; and 3) the design components, can be used to support collaboration literacy, albeit this may require intentional connections to instructional strategies and/or teacher interventions.

4.5 Discussion

The seven categories identified from student survey responses help me understand how students think about collaboration. These seven categories include apparent ties to prior research and several novel considerations. The dimensions of Communication, Contribution, and Constructive are readily apparent across a broad spectrum of research projects in collaboration analytics, MMLA, and educational dashboards. Specifically, these three dimensions address primary questions about what people are discussing within a given collaborative group, who contributes, and how effectively the group achieves its goals. On the other hand, some of the most intriguing considerations come among the dimensions of Context, Climate, and Compatibility, which move outside of the content of group interactions and, instead, focus more on questions of where collaboration takes place, the affective dimensions of group activities, and questions of overall group dynamics.

While less prevalent in responses to our survey questions, these dimensions also appear relevant for supporting student collaboration, both in terms of direct information for learners and information that teachers might use to support student collaboration and collaboration literacy. Moreover,

in our analysis of category co-occurrence, I see that while Communication and Contribution frequently co-occurred with most other dimensions, the Context code was a notable outlier. Namely, in nearly 30% of the instances that the Context code was utilized, it was the only code applied to that response. This suggests that considering the Context dimension is particularly important.

In the second portion of the analysis, I examined how these dimensions might be implemented in a collaboration analytics platform. I situated these examples within the design of the BLINC tool. Upon reviewing those examples, I demonstrated how a given design or interface component can satisfy multiple categories. I also note that there are numerous ways to address the different dimensions, and that the objective is not necessarily to explicitly name each dimension within the interface. The goal of considering the other dimensions is to ensure that the data provided aligns with how students may want to think about or reflect on their collaboration experiences.

Additionally, as I described in our use of users' smartphones as data-collection devices, intentionality in the design process can enable users to practice and reflect on collaboration analytics across a broader range of contexts. The discussion of implementing the different components as supports for collaboration literacy also reiterated that these tools may reasonably need to be used alongside intentional instruction or interventions by an educator.

The current analysis has provided an opportunity to closely examine how students think about collaboration and collaboration literacy. I position collaboration literacy as an essential re-orientation for the collaboration analytics community. Considering collaboration as literacy can prompt us to consider how to help people learn to read and respond in different collaborative settings. Developing such a skill set requires that I support more thorough reflection of collaborative experiences. Importantly, this reflection should carefully align with the seven categories (or dimensions) I identified in this chapter: Climate, Compatibility, Communication, Conflict, Context, Contribution, and Constructive.

4.6 Conclusion

Throughout this chapter, I present and analyze the results of a survey conducted among college students on collaboration. The questions sought to understand how students conceptualize collaboration to inform the further design of the system. The answers to these targeted questions were then thematically analyzed to classify their conceptualizations into the 7 C's of Collaboration: Climate, Communication, Compatibility, Conflict, Context, Contribution, and Constructive. This classification was then used to align the system with this conceptualization, thereby improving students' engagement with it. By categorizing students' thinking about collaboration, the system can be designed to support and build on students' proximal zone of development.

Importantly, these classified categories are not intended to suggest specific language to use within a collaboration analytics interface. Instead, researchers and designers should reflect on the categories and ensure that the tool's different dimensions are incorporated into its design or use. Additionally, I observed that the categories are interrelated and encompass a diverse set of ideas students hold about collaboration. From this perspective, researchers and designers may find that a single cleverly designed interface component can address multiple dimensions.

I also note that the categories are not prescriptive in identifying what constitutes good or bad collaboration. Instead, they address dimensions along which someone may reflect on how they — or their group — are collaborating as desired. While I recognize that these dimensions are based on a relatively targeted sample of students, the emergent categories build on existing research in the collaboration analytics and collaborative problem-solving communities in promising ways. I suggest that, through the seven dimensions outlined in this chapter and tools such as BLINC that aim to support collaboration literacy, I can develop the system to employ pedagogical methods to support collaborations across a variety of contexts. Throughout this chapter, I provide examples of how to use the categorization to develop refinements to a collaboration analytics Tool.

Chapter 5

SUPPORTING STUDENTS' DEVELOPMENT OF COLLABORATION WITH A LEARNING ANALYTICS TOOL

This chapter analyzes a survey on the types of collaborative support that students desire and presents design recommendations for student-facing collaboration analytics systems. In Chapter 3, I developed a prototype MMLA system to measure and support students' development of collaborative skills. Then, in Chapter 4, I focused on students' comprehension of collaboration to align the system design with this conceptualization. This was an essential step in tailoring the system to provide data and analytics that start engagement with students in their current zone of development.

While the data may be understandable, it is also imperative to engage students in using the system for it to be effective. This is especially crucial when learning collaboration, as people rarely improve their collaboration skills without direct engagement with collaborative concepts [60]. Engaging with students in their conceptualization is essential throughout education, as is evident in the development of collaboration literacy [110]. Along with this, naturally engaging students is vital to empower students to own their data [140]. Understanding how students currently regulate their collaborations was a significant goal of the survey and subsequent analysis. The questions of the study were tailored to assist in answering **RQ3**: What support do college students want to engage with their collaborations?

Surveying was conducted after students had a chance to use the prototype system, providing them with a concrete idea of an MMLA tool to support their thinking about the survey questions. I performed a thematic analysis of the survey responses, backed by Google's Gemini AI, using the collaborative areas identified in Chapter 4. This enabled me to derive insights from the survey and align them with students' conceptualizations of collaboration. I then provide a set of design recommendations to allow the system to support the collaborative interactions that students desire.

This will help me answer **RQ5**: What features should be added to BLINC to align with students' goals, comprehension, and concerns?

5.1 Background

Learning collaboration requires active coordination of multiple cognitive processes across numerous brain regions [110]. Furthermore, a few significant insights into providing effective educational feedback were identified through analysis of output from multiple student collaborative assessments. This includes task- and learning-focused feedback, supportive and aligned feedback, timely feedback, and a receptive, capable learner. This report also importantly notes how to effectively support students in building their conceptual models. This involves providing information that students can immediately apply, while also allowing students to develop their own conceptual models. This can only be achieved by engaging students in their collaborative development, as excessive direct support may limit their growth in collaboration literacy by reducing their thoughtful engagement with collaborative concepts and preventing them from developing their own internal conceptual models. In this work, the researchers found that students need to create their own explanations or attempt to help or assess other students to fully integrate complex mental representations.

The complexity of collaboration engages multiple brain regions, making it challenging for a student to learn and manage these processes without support or organization for the associated cognitive processes. This support for students is just as important as the activity they are engaged in for building their collaboration skills. This tailored advice is a central tenet of K-12 education [141]. Within collaboration, there are multiple skill levels. Without explicit instruction, we rarely see people move between different levels of collaborative skill [60]. Importantly, within this support, there needs to be a balance between providing information and allowing students to make sense of it in their own way, thereby developing their internal conceptual model of collaboration. This further highlights the importance of specific feedback on collaborations, as students need direct support in developing their collaboration literacy and the essential collaborative cogni-

tive processes. By identifying aspects of learners' observable collaborative behavior, the feedback could specifically emphasize the development of those processes. Identifying students' current conceptualization and understanding of collaboration can help tailor instruction to their needs.

By connecting students directly to the design, interventions and feedback can be designed to support students' mentalities [21]. Importantly, I can make sure the input is easy for students to integrate. While supporting students' collaboration development, the effectiveness of any tool depends on multiple contextual factors. Aligning the system's support with students' contexts will improve the effectiveness of their collaborative development. Many students lack a clear understanding of what constitutes good collaboration [64], so they may not recognize the critical areas of collaboration. This underscores the need to clearly explain why metrics are included and how they relate to the management of collaborations.

Additionally, any feedback must account for students' potential misprioritization of specific measures. This means that explainability is essential to supporting students' engagement in their collaborative development, as the system must help them understand why the data is presented and how it relates to their collaboration. Having the right training materials is essential to ensuring this development and moderating prior misconceptions about collaborations. By surveying students, I can understand how they conceptualize collaboration and use this to ensure the feedback engages them in ways they know, so they can readily integrate the knowledge the system provides.

5.1.1 Large Language Model Thematic Coding

Before the emergence of Large Language Model (LLM)s, thematic coding was performed manually. This required multiple people to code a subset of the data and reach consensus before coding the remainder. This is a significant approach to interpreting user survey results, leveraging researchers' expert knowledge to extract and synthesize insights. LLMs have grown to be able to support and perform this task independently, but there are a few factors to consider when using LLMs to accomplish this task.

OpenAI's GPT model [119] was a pioneer in LLMs. It was one of the first general-purpose

models to achieve significant success when transferred to different domains within NLP. Eventually, this model also achieved success with multimodal data, demonstrating its ability to understand natural language. This has led to multiple studies analyzing the potential for an LLM to thematic code surveys. While there are many benefits to this, some minor considerations must also be made when using LLMs in this task.

When using an LLM, the specific model used can have a significant effect. Different models vary in precision, applicability, and user-friendliness, and these factors should all be considered before selecting an LLM [142]. Even after choosing a model, steps should be taken to ensure precision. This could involve a comparative analysis of manual and automatic encoding, or guaranteeing the validity of the selected codes by manually defining them. This is important as any LLM has some level of bias, and human discretion is needed to mediate this bias [143]. Accordingly, human coding also involves some bias. By combining manual and automatic coding, these sources of bias can moderate one another, thereby improving the final quality of the codes [144].

To further support the strength of LLMs in performing thematic coding, studies have been done to examine how to effectively use LLMs in this task. These studies have allowed for structured methodologies on how to best prompt LLMs to perform different steps of thematic coding, such as theme generation, labeling, and even analysis of labels [142, 144, 145]. In one study, it was compared with other models for unsupervised theme extraction from text and was found to be more detailed than prior thematic modeling approaches [145].

During coding, ChatGPT exhibited an error rate of approximately 1 in 20. A minor caveat about these models is that, while they are good at understanding descriptive information, they are less effective at interpreting figurative or more nuanced phrasing. This led to a suggestion within each of these studies of a combined approach. Using LLM to moderate human bias, while human oversight can moderate the bias within the model. This could be done by combining manual and automatic coding, or by using an LLM to perform an initial pass of the data, then manually interpreting and confirming those codes afterward. These works, which use LLMs for thematic coding, show significant promise when applied correctly.

5.2 Study Design

I asked 293 college students to test our system and then complete a multi-part survey. All participants were university students from a private university in the United States. These students comprised 201 undergraduates, 88 master's students, and four Ph.D. students. There were 47 first-year, 62 second-year, 96 third-year, 80 fourth-year, and eight fifth-year students who participated. I received 205 fully completed survey responses, yet we still included partial responses in our analysis. This set of questions sought to gauge students' attitudes toward engaging with the systems and the collaboration analytics system. The questions included seven free-response questions:

1. Please provide 2-3 sentences about the purpose and functionality of the data capture platform described in the video.
2. What kinds of information would be useful to evaluate the collaboration quality of individual participants?
3. What kinds of information would be useful to evaluate collaboration quality at a group level?
4. What kind of collaboration quality information would you like to see in real-time?
5. How would you use this real-time information?
6. What kind of collaboration quality information would you like to see after collaboration?
7. How would you use this post-hoc information?

A quantitative analysis was conducted by thematically coding each response according to the 7 C's of Collaboration: **Communication, Compatibility, Conflict Resolution, Context, Contribution, and Constructive**. A single response could be assigned multiple codes if it touched on different dimensions of collaboration. To this end, Gemini was used to assign thematic labels to the responses. The prompt used with Gemini is detailed in appendix A. To ensure the acceptability of the Gemini labeling, I analyzed the inter-rater agreement between the prompt's performance in the previous thematic coding in chapter 4 and the consensus codes. I determined the absolute percent agreement between the previous consensus codes and Gemini's coding, and Gemini achieved **79%** agreement, exceeding the standard minimum threshold of **75%**. With this acceptable performance,

Gemini was then used to code the new responses in this chapter further. I also performed my own coding and compared the responses where they aligned. These final codes were used to analyze the reactions further and provide design recommendations to align the system with students' preferred ways of regulating their collaborations.

5.3 Results

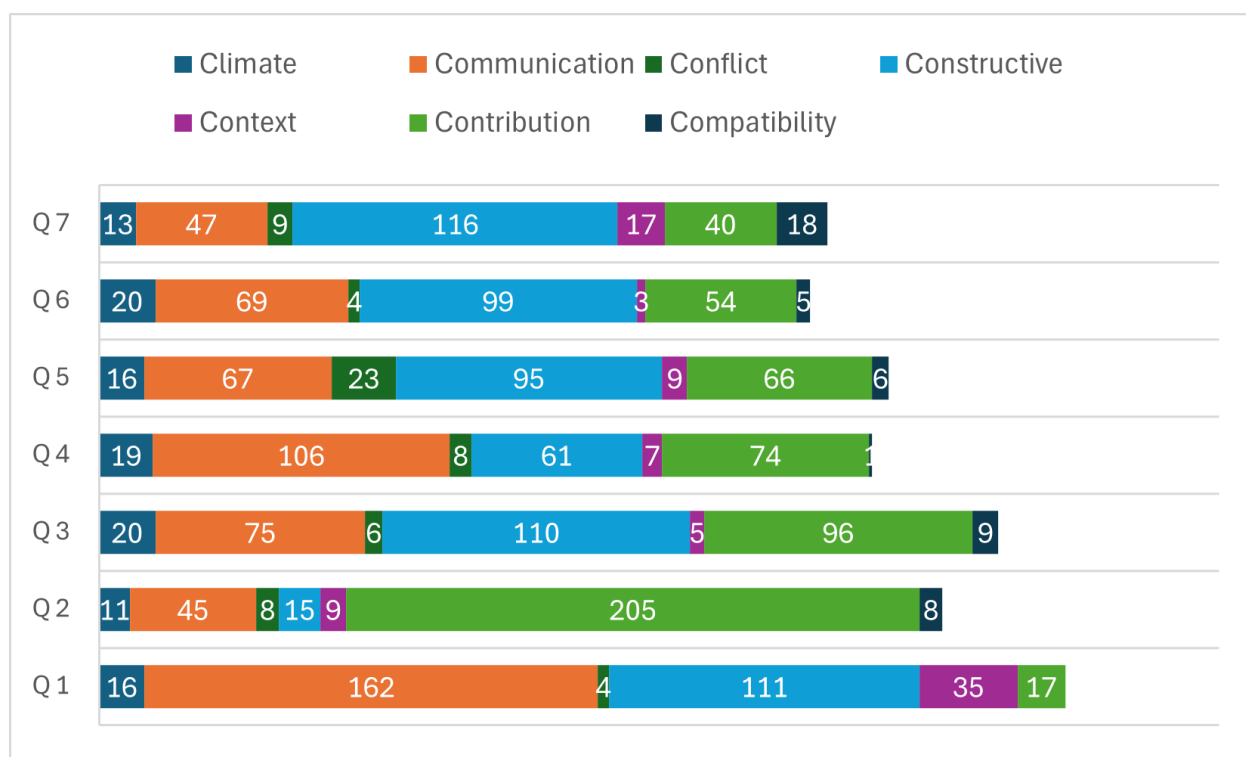


Figure 5.1: Quantitative Overview

The thematic labeling revealed clear patterns in the students' responses. Figure 5.1 shows the frequency of each theme across the survey questions, providing an overview of the aspects of collaboration that students are most focused on. The data clearly indicates that Contribution and Constructive are the most dominant themes. Students are primarily concerned with knowing who did what and how that work contributes to the outcome. Climate and Communication follow, indicating a strong desire for healthy interpersonal dynamics and effective information exchange. Notably, Context is most prominent in Q5, where students discuss using post-collaboration infor-

mation to learn and improve for the future. Conflict is the least frequent theme, suggesting students may be less focused on active disagreement and more on the foundational aspects of making collaboration work.

5.3.1 Thematic Analysis

In this section, I review responses using the seven collaboration themes and detail the most frequent student feedback for each question. Using the seven C's of Collaboration helped ensure the insights provided could be tailored to continue supporting students' conceptualization of collaboration, as detailed in the previous chapter. The analysis reveals a multi-layered set of needs students have for collaborative support. A desire for fair and productive outcomes primarily drives them. Still, they are acutely aware that these outcomes depend on a foundation of healthy interpersonal dynamics and a commitment to learning from their experiences. Additionally, the information students sought varied between real-time and post-hoc, and an overview of those responses is shown in Figure 5.2. I will analyze these differences below.

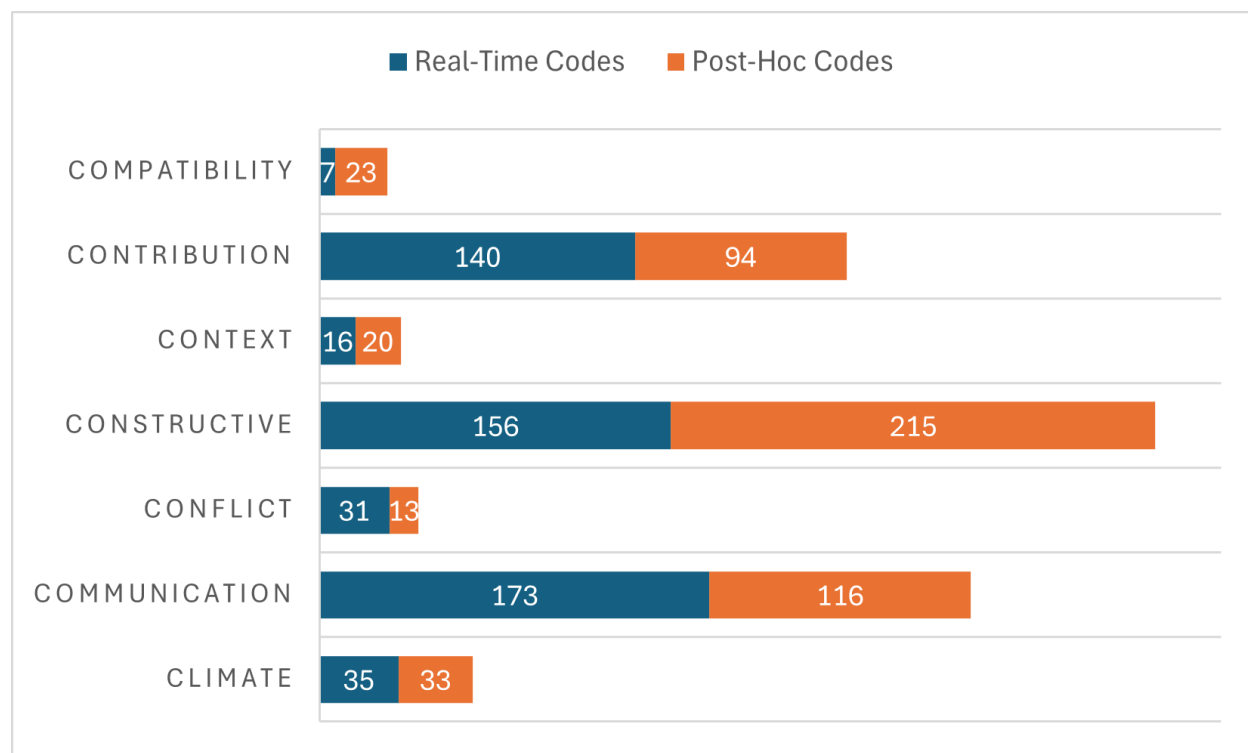


Figure 5.2: Real-Time vs. Post-Hoc responses

Contribution

The theme of Contribution was the most frequently identified, indicating a deep-seated student concern with fairness, accountability, and the validation of effort. This goes beyond simply wanting credit; it's about ensuring a functional and equitable team process, often stemming from past negative experiences with unbalanced workloads.

In real-time and post-hoc, students express a strong desire for objective, data-driven evidence of work. This is a direct response to the challenge of managing "free-riders." Students want tangible metrics that cannot be easily disputed, such as "edit history (if some sort of paper)", "The number of lines of code they committed on a coding project", and tracking "which parts of each assignment each person completed". This desire for evidence is about creating an accountability system that makes contributions visible to everyone in the group.

In real time, some students requested the ability to actively track their work. Specifically, one student asked for a "tasks board where we can clearly assign tasks and see live progress". Something that could also notify them when the group was not following their assigned tasks. One student noted, "If I noticed things going south I would step in and explicitly encourage those who aren't participating". Another stated they would use the info to "warn them to get back on task". This shows students' desire for real-time tracking of their contributions to keep themselves accountable to themselves and each other.

Students' desire for accountability was also important post-hoc. Students demonstrated a nuanced understanding that not all contributions are equal, and the ability to understand the practical impact of contributions on the collaboration was also desired. Students want to know "who worked on which parts of an assignment" and the "percentage of the total sections they contributed to." These were important to students as evidence for evaluations or as justifications for a fairer assessment. To this end, students recognize the potential flaw in simple metrics, noting that "just because you talk a lot doesn't necessarily mean you are collaborating". They are interested in the quality and impact of contributions, not just their volume. One student wanted to measure "how creative or unique the group members have come up with". At the same time, another focused on

whether an individual provided "thought-provoking ideas that initiate further discussion and allow for progress". This suggests that a tool must go beyond simple counts to help students identify impactful moments and track context-specific actions to understand how much they contribute to a collaboration.

To capture the nuances that automated systems might miss, students place high value on peer evaluations. They see this as a way to account for intangible contributions, such as providing moral support, organizing meetings, or clarifying complex ideas. Many students advocate for an "anonymous survey for each individual to explain their opinions on the other group members" or a system of "anonymous peer ratings on collaboration". This human-in-the-loop approach is considered a crucial check on purely quantitative data, providing a more holistic view of an individual's role within the team.

Constructive

The Constructive theme is closely intertwined with Contribution, as students view the entire collaborative process as oriented toward a specific goal. They want support not only in completing their individual tasks but also in ensuring that the group as a whole moves efficiently toward a high-quality, synthesized final product. In real-time, students want visibility into the group's progress to manage their workflow and maintain momentum. They desire tools that show "Everyone's task progress and plan, as well as current problems and approximate completion time". This information serves a dual purpose: it helps individuals "adjust my work pace". It holds the entire group accountable to deadlines and milestones, ensuring that the "goals that have been outlined are being accomplished at the pace that the group hoped for". Students want real-time alerts with their collaborations to ensure they do not drift away from their original goals. They are keenly aware of wasted time, such as when collaborators "get off topic". Therefore, a support tool is valuable if it can help them "maximize work efficiency" and ensure that meeting time is spent productively.

After the discussion, students' constructive needs focused more on how the deliverables of a collaboration came together as a whole and their quality. According to students, a key indicator of

a successful collaboration is the cohesiveness of the final product. A good outcome is not merely a collection of individual parts but a seamless whole in which "all the components [are] seamlessly integrated together". However, some students understood that a delivered "final project doesn't always mean good collaboration." An essential aspect of this was the desire of group projects not to look like a "collection of separate projects stitched together arbitrarily". This shows students' concern for cohesion when working with others. While there is a primary concern with how the product comes together at the end of a collaboration, ensuring that the collaboration produces a cohesive outcome is also of interest to students.

Climate

Students explicitly recognize that the emotional and interpersonal Climate of a group is not a secondary concern but a critical factor that can enable or hinder success. This aligns with the emphasis on the "psychological safety climate," in which members feel comfortable taking risks. The need for a safe and respectful environment is a recurring theme. Students describe good collaborations as spaces where "everyone is given a fair chance to voice their opinion and does not feel scared" and where they feel "comfortable with expressing their opinions within the group". The fear of a "toxic atmosphere" is a significant concern, suggesting that tools to monitor negative sentiment or promote inclusive practices would be highly valued.

Importantly, when positive collaboration climates are mentioned, students don't desire real-time tone moderation but mainly post-hoc tone analysis. This suggests that, although students want to maintain a positive environment, minute, instantaneous monitoring of tone is not perceived as a positive influence. This shows the need to balance the tool's monitoring with students' need to feel comfortable openly participating and expressing themselves. Another significant component of a favorable climate is ensuring that all voices are heard. Students worry about dynamics in which one person "dominating the conversation" or some members are "overshadowed". They express a desire for real-time information that could help them "find who isn't talking as much or isn't as focused and try to engage them more" or encourage "more silent people to speak up". This reflects

a desire not only for politeness but also for the active inclusion of all perspectives.

Communication

Students identify Communication as the fundamental mechanism that drives all other aspects of collaboration. They seek support for the logistical, conceptual, and relational aspects of information exchange within the group. In real-time questions, this was the most coded theme. The primary goal of student communication is to ensure everyone is aligned. They repeatedly use phrases like making sure "everyone is on the same page" and that "all members of the group understand the assignment in the same way". A breakdown in this shared understanding is seen as a primary cause of project failure. To support this, students expressed a desire for ways to prevent one student from dominating the collaboration, such as "if there is one person speaking their ideas" or "overlapping talking". Conversely, students also wanted to know "who isn't talking as much... and try to engage them more". This shows a strong desire to balance communication in students' collaborations.

Beyond simple clarity, students are interested in the quality of their discussions. They value collaborations in which the group can "bounce ideas off of each other" and in which members actively "try to truly understand their peers' unique contributions and try to provide a meaningful response back". This indicates a need for tools to visualize or analyze the flow of conversation and the process of co-constructing ideas. Students also requested additional quality information about the communication, such as "volume level" or a "timeline... to determine the tone". Students show a strong interest in actively tracking different parts of their communication in real time during collaborations.

Context

The theme of Context reveals that students view collaboration as a meta-skill to be learned and honed over time. This aligns perfectly with the central goal of "collaboration literacy" outlined in chapter 4. They want to use the data from their experiences not just to get a better grade on the

current project, but to become better teammates on the next one. Students overwhelmingly view post-collaboration data as a tool for reflection. They want to use it to "sum up this experience", "recognize mistakes I made", and "reflect on my performance and do better for the next projects". This desire for a structured debrief is sophisticated, with one student even comparing it to the "post-sprint analysis of agile".

The goal of this reflection is to produce actionable insights. Students want to use the information to "identify potential issues or flaws that need to be corrected or fixed" and to "figure out what people are better at, and also where they struggle". This focus on specific, improvable behaviors demonstrates a clear desire to use these tools for personal and professional development, moving from simply completing a task to mastering a skill.

Compatibility

The Compatibility dimension refers to "how well group members use complementary working and interaction strategies". Students recognize that a group's success often depends on a specific "chemistry" or alignment in how members approach work and interact with one another. Students actively think about whether their own methods of working will align with their teammates'. A primary use for post-collaboration information was to diagnose this alignment for future projects, with one student wanting to "see if I am compatible to work with that person and who I can better work with". Another student asked for group members' "schedules, [and] background knowledge" up front to assess compatibility.

Crucially, students do not see compatibility as a fixed, static trait. This suggests an understanding that compatibility can be cultivated. Some survey responses reflect this, with students hoping that after a collaboration, "We can build some connections", indicating that the process itself can foster compatibility. The search for compatibility leads to a common dilemma: is it better to work with friends or strangers? The paper highlights this tension, with one student advocating for working with friends "where you can communicate your needs and ideas comfortably". However, another student gives the opposite advice to younger students: "Find people who work

well together; these people may not necessarily be your closest friends”. This shows that there is no single answer, and that students want help navigating these team-formation choices to find the most effective partners.

Conflict

The Conflict dimension captures students’ “approaches and needs for responding to challenging or contentious situations”. While it was the least-coded theme, students’ discussions of conflict reveal a desire for tools and strategies to manage it constructively, rather than letting it derail a project. In the survey, students identified specific behaviors that lead to conflict and degrade collaboration quality. One student wanted to track the “Number of interruptions”, noting that if a person “is [interrupting] others often, then they are decreasing collaboration and should be warned or removed”. Another described a poor collaboration as “waste time being blocked by other people or waiting on other people”. These examples show that students can pinpoint concrete actions that signify conflict.

Students don’t just want to avoid conflict; they want to learn how to resolve it. The paper highlights a significant student need for support in this area. One student wished for “an experienced educator... [to act as a] conflict mediator”. Importantly, the goal isn’t just to solve a problem for them, but to teach them the process. Another student wanted someone more experienced to “show how issues can be resolved”. This demonstrates a clear need for tools that either flag moments of conflict for instructors or provide students with scaffolds and strategies for constructive confrontation and resolution.

5.3.2 Co-occurrence

To better understand the relationships between these themes, a co-occurrence analysis was performed, counting how often pairs of themes appeared in the same response. This reveals how students conceptually link different aspects of collaboration, showing that these dimensions are not isolated but deeply interconnected. The most significant pairings highlight a pragmatic, goal-

oriented view of collaboration, moderated by essential interpersonal factors. This analysis shows that students have a sophisticated mental model of collaboration. They are centered on achieving Constructive outcomes through fair Contribution. Still, they recognize that this is only possible with a foundation of a favorable Climate and clear Communication, all while viewing the entire experience within the Context of their own learning and development.

Contribution and Constructive

The strongest relationship, by a significant margin, is between Contribution and Constructive (83 co-occurrences). This demonstrates that students fundamentally see individual effort as a means to an end. They don't evaluate contributions in a vacuum; they link them directly to the group's ability to produce a high-quality result efficiently. This connection is often explicit in their responses. For example, one student stated that the key information needed is "the individual's contribution to the team's results", perfectly merging the two themes. Another linked them by defining group quality as the ability of "all members in the group to explain all parts of a project", implying that genuine Contribution leads to a Constructive and well-understood outcome.

Constructive and Context

The second most frequent pairing, Constructive and Context (42 co-occurrences), reveals that students often frame the success or failure of a project as a learning opportunity. When they consider the outcome (Constructive), they frequently do so with an eye toward future improvement (Context). They want to analyze what went right or wrong to "improve the group for the future" or "recognize mistakes I made and use them to prepare for my next collaboration". This pairing underscores the importance of the "collaboration literacy" concept, which emphasizes not only succeeding on a single task but also becoming a better collaborator over time.

Climate, Communication, and Constructive

The analysis shows that students see a favorable Climate and clear Communication as essential foundations for achieving their pragmatic goals. Both themes co-occur frequently with Contribution and Constructive. For example, students believe that a poor Climate, where members don't feel comfortable, directly harms outcomes. As one student put it, underlying conflicts create tension, and it's "important to sort those out for efficiency", linking Climate to Constructive goals. Similarly, good Communication is seen as a prerequisite for success, as the "cohesive[ness of the] final results... tend to indicate how much communication the team had". This shows that students understand that interpersonal dynamics are not just "soft skills" but critical components of an effective workflow.

Communication and Conflict

While Conflict was the least-coded theme, its co-occurrence patterns are revealing. The survey analysis in chapter 4 found that Communication appeared alongside approximately three-fourths of the responses that included Conflict, suggesting that students primarily view conflict as a breakdown in communication. The current survey supports this; when students discuss addressing conflict, it's often through communicative acts, like when one student described dealing with uncooperative members by trying "to communicate with those people... trying to ask them to finish their own work".

5.4 Discussion

Based on the thematic analysis, recommendations for designing AI-powered collaboration tools should center on fostering collaboration literacy—the ability of students to understand and improve their collaborative processes. The tool shouldn't just be a passive measurement device but an active support system for reflection and growth. In this section, I discuss different ways to support the insights from the thematic analysis.

The thematic analysis shows that students are most concerned with the fairness of contribution tracking and the progress toward a quality outcome (constructive). Features supporting these areas should be emphasized in the design. To ensure fairness, dashboards should transparently visualize individual participation and provide context on how students are measured. This could include linking words to the specific emotions detected or other ways to show how the metrics were calculated. Students also want features that help them monitor their constructive progress. This means linking collaborative activities to constructive outcomes. To support this, tools could include shared milestone trackers, keyword monitors to ensure discussion stays on topic, and visualizations of team productivity over time.

There was also a strong student desire for a healthy group climate and effective communication. A tool, therefore, should provide insights into groups' interpersonal dynamics to support these concerns. Move beyond how much people talk to how they talk. Features such as sentiment analysis, as in the BLINC platform, should be integrated to provide a window into the group's emotional tone. This directly addresses the climate dimension by helping students see if the atmosphere is respectful and open. The tool could identify potentially harmful patterns, such as frequent interruptions or one person dominating the conversation, which students in the survey mentioned. It could also highlight positive patterns, such as equitable speaking turns and members building on each other's ideas, to reinforce good communication habits. The climate could also be affected by feelings of surveillance [146], so addressing students' privacy concerns is a top priority for future studies.

Another key finding was that students want to use collaboration data to learn and improve for the future. The tool's design must support this reflective context. After a session, the tool could generate a summary report. This report could highlight key statistics, notable moments, and areas for improvement, directly enabling the reflection that students requested in the survey. This aligns with moving beyond in-the-moment interventions to building long-term collaboration literacy. The tool could prompt students with reflective questions based on their data. For example: "We noticed one person spoke for 60% of this session. How did this impact the group's dynamic?"

Another feature could be to enable peer evaluation to encourage students to reflect deeply on their collaborations. These encourage students to interpret the data within their own context, rather than having the tool make prescriptive judgments.

The paper found that students collaborate more frequently outside of class than inside it. This finding, combined with students' concerns about misinterpretation by external tools, means the tool must be adaptable and user-driven. The tool should be mobile and easy to deploy in various contexts, from a library study room to a dorm lounge. Relying on users' smartphones for data capture, as BLINC does, is an effective strategy for meeting this need. Let users adapt the analytics to their specific needs. Allowing users to define and customize their metrics makes the feedback more relevant to their particular collaboration. Importantly, this gives students ownership of the analytics process and helps ensure the data align with their group's unique context.

5.5 Conclusion

In this chapter, a survey was conducted to examine how college students want to be engaged to support their collaborations. Students were allowed to use the collaboration literacy development system and then completed a survey. This survey was analyzed using the 7 C's of collaboration, as detailed in chapter 4, to support the development of the system within students' conceptualization of collaboration. This is a crucial step in ensuring the system provides scaffolding that naturally engages students in their collaborative development. Further analysis of coded surveys enabled me to provide insights into aligning the system with students' preferred methods of learning collaboration literacy.

The significant insights from the survey included: transparent contributions, goal tracking, interaction-quality information, highlighting communication patterns, providing post-hoc summaries, encouraging self-assessment, setting flexibility, and student customization. Some of these features are already discussed in previous chapters and in the collaboration analytics literature, but this study highlights their importance to students. By incorporating these insights into the design, the system can better engage and support students. These insights can also inform the future design

of other collaboration literacy development systems.

Chapter 6

RECOGNIZING STUDENTS' ETHICAL AND PRIVACY CONCERNS

Chapters 4 and 5 examine how students think about and want to manage their collaborations. In contrast, this chapter focuses on the privacy and ethical issues that arise when performing automated collaboration analytics with students. Mainly, the survey will attempt to address **RQ4**: What ethical and privacy concerns do college students have when engaging with a collaboration analytics tool? This is a critical area to address when designing any learning analytics system, ensuring students' privacy while providing responsive feedback.

While several collaborative analytics systems using Machine Learning have been recently developed, such as CPSCoach [147] and CoBi [65], students' ethical or privacy concerns have rarely been a primary focus. This is worrying as ethics and privacy should be the significant focus of the design of any learning analytics system. In particular, any tool that aims to support students should consider the moral obligations (or ethics) arising from power dynamics in education and respect students' legal and implied rights to data privacy. The importance of data rights can also vary by location, as different cultures have different perspectives on data privacy. In previous chapters, we saw how vital psychological safety is to students, and failing to address their ethical and privacy concerns will undermine it. This, in turn, will prevent them from fully engaging in the system's collaboration literacy development and participating in collaborative learning.

To better understand students' ethical and privacy concerns, this chapter uses the 7 C's of Collaboration from Chapter 4 to analyze the results of a student survey. This allows me to classify the feedback into the previously learned areas of students' conceptualization of collaboration. This helped to ease the integration of the survey feedback into the system design. Through this classification of students' concerns, I identify a set of design recommendations to prioritize that will alleviate these concerns and help answer **RQ5**: What features should be added to BLINC to align with students' goals, comprehension, and concerns?

6.1 Background

When engaging in data collection and analysis, a learning analytics practitioner has multiple ethical responsibilities. In some jurisdictions, there are legal requirements for all data handling, such as the General Data Protection Regulation, 2016 [3] (GDPR), the The Digital Personal Data Protection act, 2023 [1] (DPDP), or the Lei Geral de Proteção de Dados Pessoais (General Personal Data Protection Law), 2018 [4] (LGPD). Adapting to different standards across countries can be challenging. Additionally, there are legal requirements for the recording of student data in schools, such as the Family Educational Rights and Privacy act [2] (FERPA) in the US. These requirements are the best starting point for addressing privacy and ethics, as they provide a basis for students' legal rights to privacy. When recording students, acknowledging only these rights may not fully assuage student concerns, and this is where further ethical practices can build trust and rapport between students and learning analytics tools

The GDPR [3] is a good starting point as it applies to all EU countries, so by engaging with these requirements, we can ensure the system's compliance in many countries. When collecting personal data, this regulation involves seven major principles: lawfulness, fairness and transparency, purpose limitation, data minimization, accuracy, storage limitation, security, and accountability. These principles are accompanied by data rights for individuals in the EU, including: the Right to be Informed, the Right of Access, the Right to Rectification, the Right to Erasure, the Right to Restrict Processing, the Right to Data Portability, the Right to Object, and the Rights Related to Automated Decision-Making, including Profiling.

Under the first of the seven principles of the GDPR [3], lawfulness, fairness, and transparency require that one of the six lawful grounds for processing data be met, such as consent, contract, legal obligation, vital interests, public task, or legitimate interests. This lawful ground should be clear to users. For purpose limitation, there should be justification for the use of any personal data. Data minimization involves limiting the data necessary by applying three crucial principles—adequacy, relevance, and limitation—to each piece of personal data. Accuracy is the next area,

which entails taking measures to ensure that critical data is regularly maintained and validated. Storage limitations require that data no longer needed be deleted promptly. Security is about granting users access only to the data they are authorized to access, while ensuring that essential data remains recoverable. Finally, accountability is the principle that each organization is individually responsible for ensuring compliance with the requirements.

The GDPR [3] focuses on general data protection requirements, FERPA [2] focuses on data collection within school settings in the US. Under FERPA, audio and other data signals are considered education records and are therefore subject to the standards detailed in this regulation. The standard details a set of rights for students, including the following: the right to inspect and review records, the right to request corrections, the right to request a hearing, and the right to control disclosure. Students should be directly informed of these rights in academic settings, and, for students under 18, their parents must acknowledge their rights to their child's information. These legal obligations are the minimum standard, and a good design should be held to a higher standard to support students' psychological safety (or climate) best while being recorded.

Directly with LA, specific issues arise regarding the data collection and processing. A significant source of these issues is the unequal power dynamics in most educational settings. These power dynamics have at times led to learning analytics tools being seen as surveillance, and this has been shown to harm students' classroom participation [146]. Another source of the issue is algorithmic bias [148], which is especially important given the power dynamics. This can lead to misrepresentations or misunderstandings of students' identities and to the imposition of prior biases on new generations of students without proper engagement with these biases [149]. It is vital to provide students with accurate data while allowing sufficient openness for them to grow and develop when assessed algorithmically [150].

One way to moderate students' concerns is to involve them in interpreting and managing their data. This is evident in modern social media such as Facebook and Google Maps, where users voluntarily share large amounts of detailed data with companies [151], a practice known as crowdsourcing. An additional effect is improved data accuracy and representation. Another common

way to moderate these concerns is also to provide transparency in the data [152]. This allows the data to be non-deterministic, arising from a specific action or effect. These are some good guidelines, but further exploration is needed to ensure that the variety of students' concerns is fully accounted for. In the rest of this chapter, I explore students' ethical and privacy concerns and detail a set of design recommendations for collaboration analytics systems to address them.

6.2 Study Design

I asked 293 college students to test our system and then complete a multi-part survey. All participants were university students from a private university in the United States. These students comprised 201 undergraduates, 88 master's students, and four Ph.D. students. There were 47 first-year, 62 second-year, 96 third-year, 80 fourth-year, and eight fifth-year students who participated. These mainly focused on privacy concerns. The questions included six Likert-scale questions(1-6), one affirmative question(7), one mandatory free response(8) and one optional free response question(9):

1. I am concerned about how I could be represented by the data.
2. I find that the device invades my privacy if used during classes.
3. I find that the device invades my privacy if used during out-of-class group discussions.
4. Compared to a Google Home or Amazon Echo, this device seems more invasive or intrusive.
5. I'd rather not have a piece of technology actively recording me in class
6. I would prefer to have a device that offers more interactivity (i.e., answering questions, and providing a visualization for the students).
7. I would be okay with having a camera also capturing video while I participate in group discussions.
8. Please provide 2-3 sentences about the purpose and functionality of the data capture platform described in the video.
9. If you have any comments about how this data might be used, please provide them here

To understand the survey, a quantitative analysis of the multiple-choice questions will be performed. Then, a thematic analysis of the free-response questions was conducted using the 7 C's of Collaboration: **Communication, Compatibility, Conflict Resolution, Context, Contribution, and Constructive**. To this end, Gemini was used to assign thematic labels to the responses. The prompt used with Gemini is detailed in appendix A. These thematic labels, combined with the quantitative data, provided insights into how to moderate students' ethical and privacy concerns when interacting with the system.

6.3 Results

The quantitative analysis of the responses reveals a significant tension between users' strong privacy concerns and their desire for more interactive tools. Within Fig. 6.3, a majority of respondents feel that being actively recorded is invasive. At the same time, they also show a clear preference for a more interactive device, with 57.1% of users wanting one with more interactive features. Notably, the proposed tool is perceived as less intrusive than common smart speakers such as Google Home and Amazon Echo. Overall, students viewed collaborative analytics in their classrooms negatively, though some recognized their value. Further thematic analysis of the free-response questions was conducted to identify the most appropriate next steps for the system's development.

The quantitative data revealed a general sense of user apprehension, but a detailed co-occurrence analysis indicates that this is not a single fear. The data showed a significant number of users are concerned about data representation (48.9% agree/strongly agree), feel the device is invasive (53.0% agree/strongly agree in-class), and prefer not to be actively recorded (46.4% agree/strongly agree). It is a composite anxiety rooted in three distinct, yet related, themes: the emotional Climate of the group, the interpretive Context of the data, and the potential for Conflict over individual Contributions.

Based on the seven themes established in Chapter 4, the thematic analysis indicates that user descriptions of the tool's purpose (Q8) primarily emphasize Contribution, Communication, and Constructive goals. However, their optional comments and concerns (Q9) are dominated by the

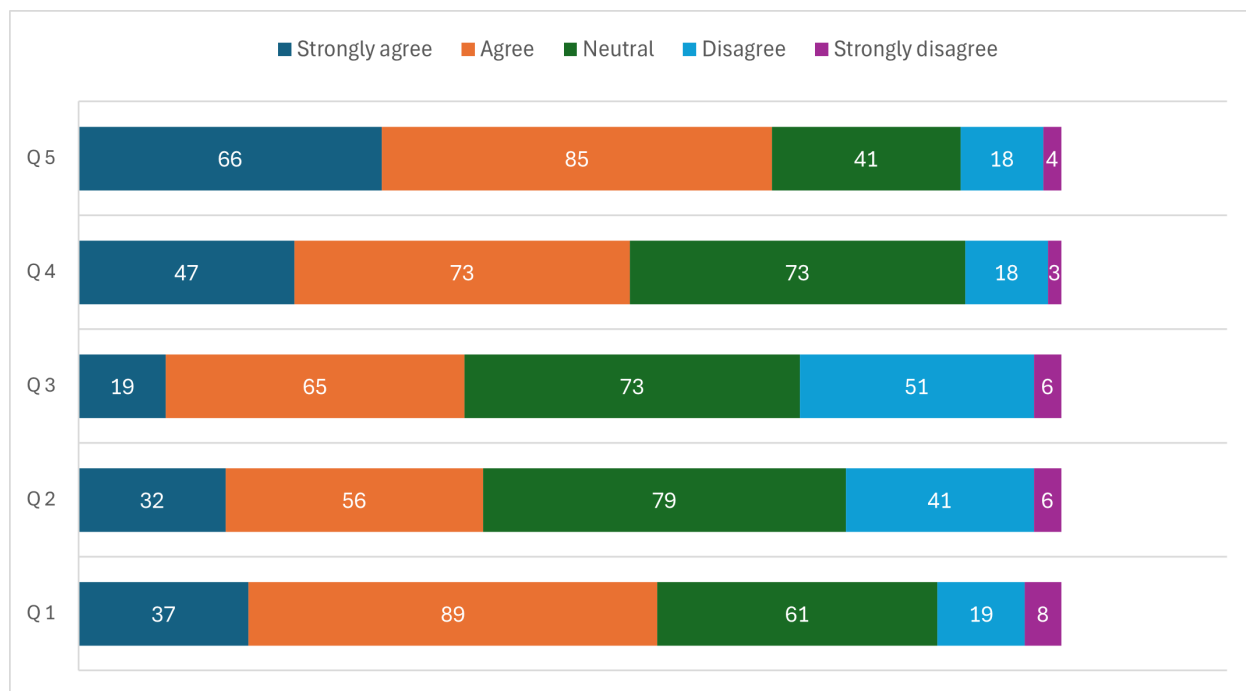


Figure 6.1: Likert question results

more personal and social themes of Climate and Context, with some references to Conflict, revealing deep-seated anxieties about fairness, interpretation, and psychological safety. Additionally, the themes of Conflict and Compatibility were not dominant in the survey responses. It seems users are focused on the tangible aspects of the task. Users may not yet see these complex social dynamics as problems that technology can—or should—solve.

6.3.1 Contribution

Contribution highlights what individuals bring to the collaboration, including equal work, effort, and verbal input. This was one of the most dominant themes in both Q8 and Q9, as users saw the tool's core function as tracking participation but also worried about the fairness of that tracking. In Q8, users repeatedly described the tool as a means of monitoring who is contributing. One response mentioned it allows facilitators to "be able to understand who is participating and who is not". Another pointed to the "Discussion direction" feature as a way to "provide an approximate distribution of participant talk within the group".

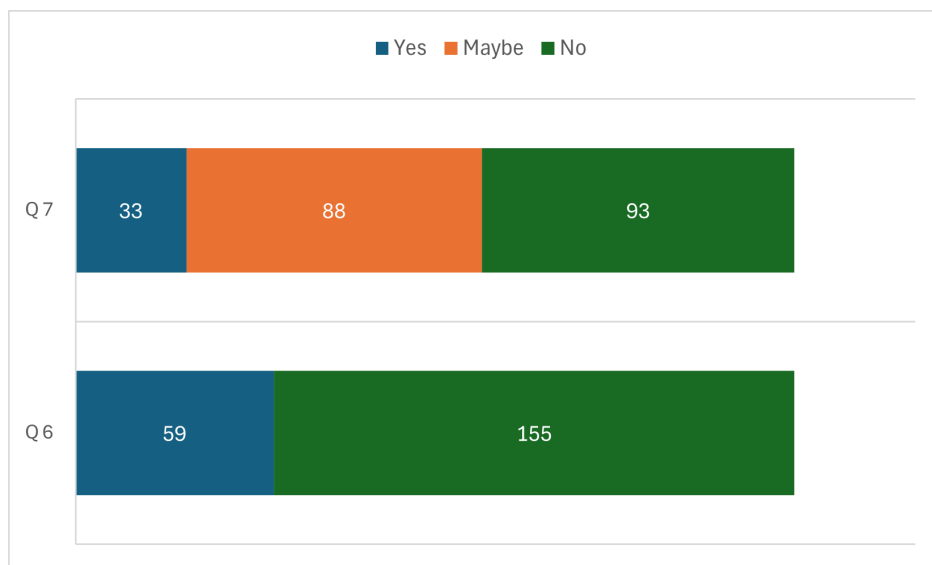


Figure 6.2: Affirmative question results

In Q9, concerns were raised about whether the tool could accurately and fairly assign credit for contributions. A key concern was, "How good is it at distinguishing voices so that I can be sure I get credit for all of my participation?". This fear can easily lead to Conflict, a dimension that encapsulates "responding to challenging or contentious situations". If a student's contribution is misattributed, this can create a conflict with the instructor over grades or with peers over perceived effort.

6.3.2 Communication

This dimension relates to the quantity and quality of information shared among group members. Descriptions of the tool in Q8 were rich in this theme, with users viewing its primary function as capturing and clarifying the flow of information. Users frequently described the tool's ability to create a transcript as a key function for improving communication and ensuring everyone is on the same page. For example, one response noted the tool's purpose is to "provide transcripts of what exactly each student says". Another user highlighted that it "creates a transcript of audio and separates it into different sections to make it easily accessible".

6.3.3 Constructive

Users' descriptions of the tool in Q8 clearly aligned with this theme, framing the technology as a means to an end: improving collaboration. Many responses focused on how the tool's analytics could help groups work more effectively. One user stated that the purpose is to "analyze discussions and use machine learning tools to identify words in discussions that could imply whether collaboration is going well." Another saw the goal as helping "to ensure that group work time is spent productively and time isn't wasted as much".

6.3.4 Climate

User concerns in Q9 strongly reflect this theme, focusing on how surveillance might create a hostile or fearful atmosphere. Users worried that constant monitoring would stifle open expression and creativity, creating a climate of fear rather than safety. One user stated, "I think people might be more scared to participate and speak their opinions if they know that what they're saying is constantly being recorded and analyzed. It might prevent people from truly expression their creativity and thoughts". Another shared this sentiment: "I imagine that this would just make group mates afraid to interact genuinely and hamper communications instead of aiding them". Another user articulated this by worrying that constant recording would make them "afraid to interact genuinely", damaging the psychological safety necessary for effective collaboration.

6.3.5 Context

This theme was central to the anxieties expressed in Q9, in which users questioned the tool's ability to capture nuance. Users are concerned that a purely data-driven analysis would lack the contextual understanding needed to accurately assess a conversation. One user noted, "This data could be harmful given that exact meanings for people can differ, especially in the context of things like tone analysis, which rely on machine learning that can be biased". Another expressed concern that "technology may be biased or not understand the full implications/nuances of collaboration". One student expressed they worried that "machine learning... can be biased" or might not "understand

the full implications/nuances of collaboration”. They fear the tool lacks the context to interpret their actions correctly.

Another aspect of context is that students see the tool as less invasive than commercial smart speakers. 55.4% of users found the device less invasive than a Google Home or Amazon Echo, while only 19.0% found it more invasive. An academic tool has a defined, limited purpose. A commercial device has a broad corporate purpose, which creates less trust. However, this “contextual trust” is challenged by another key data point: students report they collaborate more frequently outside of class than in class. As the tool moves from the controlled context of a classroom to the less formal setting of a dorm room—where students do most of their collaboration—developers must intentionally work to maintain that trust.

6.3.6 Conflict

While less prominent than other themes, it appeared in Q9, with concerns about how the tool might create or exacerbate conflict. Users raised concerns that the tool’s data could be used to unfairly penalize participants. One user hoped the tool “would not be used to penalize students with softer voices; I have a lot of trouble with being talked over”. This points to the tool’s potential role in mediating—or failing to mediate—a common source of group conflict.

6.4 Discussion

Based on the comprehensive analysis, the primary recommendation is to develop the tool around an empowerment philosophy rather than a judgmental one. This means prioritizing features that provide transparent, user-controlled data for self-reflection while actively avoiding those that feel prescriptive and top-down. This approach directly addresses the central tension in the data: users’ desire for the tool’s utilitarian benefits is in direct conflict with their fear of its social and emotional risks. In this section, I will discuss the major areas to focus on to address this tension and explain how they relate to current ethical practices regarding LA.

6.4.1 Transparency

The analysis shows users want more interactive features but are deeply skeptical of how the tool reaches its conclusions. Trust with students can be built by making the analytics transparent and interpretable. While most educational institutions are legally obligated to provide this transparency [2], this study further underscores the need for transparency as an ethical practice to mitigate students' privacy concerns. The data showed it is crucial to focus on features that objectively visualize Communication and Contribution patterns, as these are the most desired functions.

One way to provide transparency is to replace a single, opaque score for a concept such as "Emotional tone" with an interface that allows users to click the score to see which words or phrases in the transcript contributed to that analysis. Further steps that enable students to link the output to the input would support this transparency. This approach to explainability is a common way to help demystify the black-box nature of AI and to address users' concerns about understanding the context surrounding the data [153].

6.4.2 User Agency

Another major insight was significant anxiety about being recorded and judged, which corresponds directly to the Climate theme and the need for psychological safety. This anxiety is related to the power dynamics in most academic settings and the fears about surveillance [146]. This can create a "toxic atmosphere" or make users "afraid to interact genuinely" in the presence of LA systems. The most effective way to counter this is to give users meaningful control over their data. Importantly, students should feel that this data is in their service, not in the service of administrators, teachers, or researchers [149]. By enlisting students as co-interpreters, we will make them more aware of their collaborative actions.

This means implementing robust user controls for data capture and sharing, and empowering students' autonomy in their collaborative development. One feature could be providing an easily accessible "pause recording" function for sensitive parts of a discussion. Another option would be to allow students to review their transcripts and delete entire sessions. Another essential feature

could be the ability to create granular permissions that allow the group to decide which data, if any, to share with an instructor. These steps could help maintain a favorable collaborative climate while students are being recorded by empowering them to control and use their data.

6.4.3 Learning Over Assessment

The analysis revealed that the tool's "contextual advantage" is its greatest asset; it's perceived as less invasive than commercial technology due to its academic purpose. The fear of unfair grading is a significant source of potential conflict. Using the tool for assessment would confirm users' worst fears about contextual misinterpretation and unfair evaluation of their contributions, destroying the favorable climate needed for collaboration. Intertwined with this educational context is students' sense of agency, which can limit their feelings of commercialization [154]. This educational view must be actively maintained through careful implementation of features.

The tool should be explicitly framed as a formative resource for developing collaboration literacy, not as a summative tool for grading. One feature to support this emphasis is to provide clear guidelines for students on how to use the data. Another feature could be for the tool to generate reflections rather than focus on participation scores. Using the data, the reflection could provide instructional scaffolding based on notable events or data from the collaboration. Students must know that the system can help them understand collaborations, but that the patterns are not causal or deterministic, maintaining this educational and inquisitive environment rather than a grading and assessment environment [155].

6.5 Conclusion

This chapter used the seven C's of Collaboration to analyze ethical and privacy concerns arising from students' interactions with collaboration analytics systems. This is a crucial missing piece in the design of many current collaboration analytics systems. This analysis was used to provide a set of design recommendations for the next iteration of the system and for other student-centered Learning Analytics systems. Importantly, LA tools should prioritize transparency, user agency,

and learning over assessment. Applying these recommendations will help ensure students' psychological safety when engaging with collaborative analytics systems. This is not only important as an ethical researcher, but psychological safety is also vital for maintaining good collaborations.

Chapter 7

CENTERING STUDENTS IN BLINC’S SYSTEM DESIGN

This chapter attempts to synthesize insights from previous chapters to design an LA tool that supports students’ collaboration literacy development called Building Literacy In N-person Collaborations (BLINC). First, in chapter 3, a prototype system was designed that used basic collaboration metrics. Next, in chapter 4, students’ conceptualization of collaboration was analyzed to understand a starting point for students’ collaboration literacy development. In chapter 5, I explored how students want to be supported in their collaborations. Finally, in chapter 6, I analyzed students’ priorities regarding their privacy and the ethical treatment of their data. These chapters provided vital information on how best to design a collaboration literacy development system.

Building on insights from previous chapters, I detail how to align BLINC’s design with students’ perspectives in the following sections. First, I’ll review the overall design goals informed by the survey analysis in the previous chapters. Then, I’ll detail the system’s features, each accompanied by a justification for inclusion based on alignment with the design goals. Finally, I’ll discuss the implications, limitations, and possible future work to follow up on this design. Through this, I will fully answer **RQ5**: What features should be added to BLINC to align with students’ goals, comprehension, and concerns?

7.1 Background

I’ll review some key concepts for developing students’ collaboration skills. The following outlines a set of design principles to support the development of collaboration through tailored feedback and data. These design principles, along with students’ feedback, will guide the system’s design.

For many years, educators have had to judge the quality of a collaboration in real time or from recordings and decide how best to support students in developing collaboration literacy [8]. With advances in Machine Learning, collaboration analytics has developed to automate the detection of

collaboration features in recent work [15, 69, 64]. There have been multiple previous studies on providing collaboration feedback. These studies typically use collaboration analytics to identify collaboration dynamics and provide feedback based on them.

Two different tools, CPSCoach2.0 [15] and Bridging the Gap 2021 [69], use slightly differing design goals to support collaboration literacy development. CPSCoach2.0 was a tool that provided interventions for individual students at the end of a collaboration, after recording and analyzing the collaboration, to improve their collaboration literacy. This tool also had a real-time feedback display that showed overall CPS performance. The system design was guided by the following principles: **Objective Feedback, Formative Feedback, Observational Learning, Interactive-constructive-active-passive (ICAP) framework, and Instructional scaffolding.**

Bridging the Gap 2021 focuses on detailing low-level features that support the calculation of constructs that define collaboration. These constructs were detailed from another set of design principles. These principles were: **Fundamental, Learner-centered, Actionable, Comprehensible, Computable, and Theory-grounded.** Alongside these design principles, cognitive load theory has been extensively used to inform the design of learning analytics systems. Oviatt et al. [54] explicitly analyzed the effects of multimodal data on interface design. They observed that working memory was expanded when engaging with students in a multimodal manner, and students had better recall and performed better when learning and engaging in a multimodal way.

Previous work has shown students require explicit instruction in collaborative skills within a structured collaborative environment to develop collaboration literacy effectively. Students should be supported with structured and limited feedback and interventions aligned with cognitive load theory. This information must also adhere to several design principles to support effective skill development. Some design principles have been successful in other cognitive learning studies and should help validate the current design. By analyzing these works, I've developed an understanding of the current design goals for many collaboration analytics systems, which will be integrated with insights from the survey analysis in the prior chapters.

7.2 Design Goals

Here, I detail the design goals, motivated by the survey analysis in the prior chapter and by previous work in collaboration analytics. These design goals inform the overarching system design. The system's design focused on aligning with students' conceptualization of collaboration and supporting how students want to participate in their collaboration literacy development. This design builds on prior research on automated CPS skills development and emphasizes the essential features to best support students' perspectives. The main goal of this design is to be learner-centered by prioritizing these perspectives. Chapter 4 was the first significant step toward this goal by examining students' conceptualization of collaboration. The breakdown of students' understanding of collaboration was essential for the system to ensure its interactions properly perform instructional scaffolding by engaging students at their level of collaboration literacy. By combining findings from the previous chapters and prior work in LA, I can detail a set of design goals centered on students. In this section, I describe these goals, followed by a detailed description of the commonly requested features.

When working with students in formal education settings, it is vital to center ethics given the power dynamics between students and instructors. This dynamic has been shown to elicit a negative sense of surveillance when using LA [151]. At the same time, there are also serious legal obligations in these settings, such as FERPA [2]. Significantly, the feeling of surveillance can harm the collaborative climate and limit students' participation. Climate was a primary concern for students, as the analysis in chapter 5 showed, and psychological safety is already known to be important in collaborations [130]. Throughout chapter 6, I explored students' privacy and ethical concerns regarding the recording and provision of collaboration feedback, primarily in formal educational settings. This allowed me to detail a few design goals to address students' views in these areas.

Along with the climate, other students also wanted feedback on different aspects of collaboration; this was further explored in chapter 5. The exploration of how students want to interact

with a collaboration literacy development system in that chapter illuminated multiple areas for the system's design, including peer evaluation, goal tracking, and highlighting additional communication patterns. This was important for centering learners' perspectives and ensuring that system interactions are supportive of students. Here, I review the main goals of BLINC's design, which include: **Data Transparency, Student Autonomy, Prioritizing Learning over Assessment, Positive Reinforcement, Student Customization, and Setting Flexibility**, as well as feature-specific goals.

7.2.1 Data Transparency

Within a formal educational setting, FERPA requires students to be able to request access to their data at any time. This is a good baseline, but further transparency is necessary when working with AI. People are not necessarily quick to trust things they don't understand, especially when it comes to AI, and one of the best ways to build trust is to explain the methods required for replication [153]. This means users should be able to know which words or actions are associated with specific feedback or analyses. Not only is this important for building trust, but it also allows students to trace the metrics to their actions. This further enables the feedback to be easily actionable, as it is clear which behavior is associated with specific collaborative metrics. This will further develop students' collaboration literacy by enabling them to identify specific collaborative behaviors.

Student Autonomy

Student autonomy in the presence of BLINC is significant for a variety of reasons. Collaboration is rarely passively learned [60]. Engaging users in their collaboration literacy development is key to actually seeing effective development. Additionally, students expressed concerns about being judged. To support both goals, providing users with control over their data and its interpretation can help build students' sense of ownership of the data. By strengthening these feelings of student ownership, students are more likely to be invested in their development of collaboration literacy. This will help improve the climate of collaborations using BLINC by limiting feelings that data is

for others to measure or monitor them, while also improving the accuracy of the system if students can correct their data, as is seen in many crowd-source apps in modern social media where user voluntarily share their own data [151]. Allowing user autonomy and customization can also help align feedback with the specific context of a collaboration.

7.2.2 Prioritize Learning over Assessment

Students must see the tool as supporting their collaborations, not as an assessment. Already, in Chapter 6, the tool is considered less intrusive than other mainstream commercial recording products. Still, more should be done to alleviate students' concerns and empower them in their learning. Much of the student apprehension about the system stemmed from the possibility of being graded and from worries about whether grading with the system was fair, given the inherent bias of algorithms. Supporting a sense of learning over assessment involves giving students autonomy over their data, as this autonomy is a strong way to mitigate feelings of commercialization or monitoring [154]. This will allow students to see the tool as a resource to integrate into their collaborations to support them. Another way to support learning is to encourage students to reflect on their collaborations to emphasize the tool's learning goals. As self-regulation is essential for the development of collaboration literacy, students must engage directly with collaborative concepts.

7.2.3 Positive Reinforcement

Communication was a significant area for most students. Many students wanted a way to understand their communication patterns. Some students specifically hoped that understanding these communication patterns would provide insights into conflicts. This would provide monitoring of the measured collaborative indicators to notify or label points of conflict. This requires a balance, as being too aggressive in labeling negative experiences or actions could harm the psychological climate and reinforce students' previous fear of assessment. Monitoring is still beneficial, but negative feedback doesn't necessarily promote positive behavior on its own. Positive reinforcement of good collaborative actions is a stronger approach to teaching collaborative skills than negative

reinforcement [110], and this should be emphasized. By focusing primarily on positive feedback, the system aims to foster a healthy collaborative climate and to monitor and reinforce collaborative behaviors.

7.2.4 Setting Flexibility

In chapter 4, students indicated that they collaborate more frequently outside of the classroom than inside it. This is very important to factor into the system's design, as if I hope to support student collaborative development, the tool should be usable where students collaborate most. During testing, devices struggled with connectivity, underscoring the need for flexibility. Different settings may impose various constraints on connectivity or on the availability of recording devices. A design that promotes versatile deployments is critical to support this goal of flexibility. Within different settings, the common discussion vocabulary can also change. Any system must be able to moderate feedback based on the specific context and to account for these changes across different settings.

7.2.5 Feature Specific Goals

Students also had specific features commonly requested by the system; integrating these features is essential to further engage students in developing collaboration literacy. The following covers these feature-specific goals, which are: **Goal Tracking, Interaction Quality Analysis, Summative Information, and Peer Evaluation**

Goal Tracking

Students specifically asked for a way to track their progress within a collaboration. This included tracking their constructive progress and whether the collaboration is derailing. Ideally, students want a way to link their constructive progress with their collaborative actions. Contribution and Constructive were the most commonly co-occurring concepts in the survey analysis in chapter 5, indicating that students associate the constructive outcome with the extent of people's contribu-

tions to the collaboration. Goal tracking could provide data to support these connections and help students understand how their actions affect their progress toward their collaborative goals. This can also support students' self-regulation of their collaborations as they can manage their actions based on their distance to their goal.

Interaction Quality Analysis

Students within chapter 5 expressed worry about the value of the volume of contributions. This concern motivated an interest in assessing the quality and impact of contributions. Collaboration quality analysis is essential, as more contributions within a collaboration do not necessarily indicate a high-quality collaboration [8]. It is then vital to balance quantitative and qualitative metrics to provide a holistic picture of a collaborative experience. Qualitative data include assessing the relevance of a contribution within a collaboration. An example is topic modeling, which can evaluate the relevance of contributions to predefined topics extracted from documents related to the collaboration. Other quality analyses could be numerical, such as Group Communication Analysis [72], which details metrics that measure different aspects of collaboration using semantic embeddings.

Summative Information

Many students expressed a desire for more summative information. This information is valuable for helping students reflect on how to change future collaborations. This reflection could yield actionable insights from the whole collaboration and detail specific aspects to focus on. This summative information can also support students' development of collaboration literacy by prompting them to consider their collaboration as a whole.

Peer Evaluation

Although one design goal was to limit students' self-assessments, students nonetheless wanted peer evaluations, including anonymous surveys of other group members. This can allow students to

raise points during the collaboration that the system may have missed. This also supports students' feelings of autonomy by engaging them in assessing their peers rather than having a teacher evaluate them. Another significant benefit of peer evaluation is that it forces a student to think about the field being evaluated, along with the interactions involved in that field [110]. Significantly, then, these peer evaluations support students' collaboration literacy development by prompting them to reflect on each person's role in the collaboration.

7.3 System Design

In the following section, I go through the most recent system design. I then detail the components of the system design that achieve those goals and relate them to the 7 C's of Collaboration: **Communication, Climate, Compatibility, Conflict, Context, Contribution, and Constructive**. The architecture of the technical components is shown in Figure 7.1.

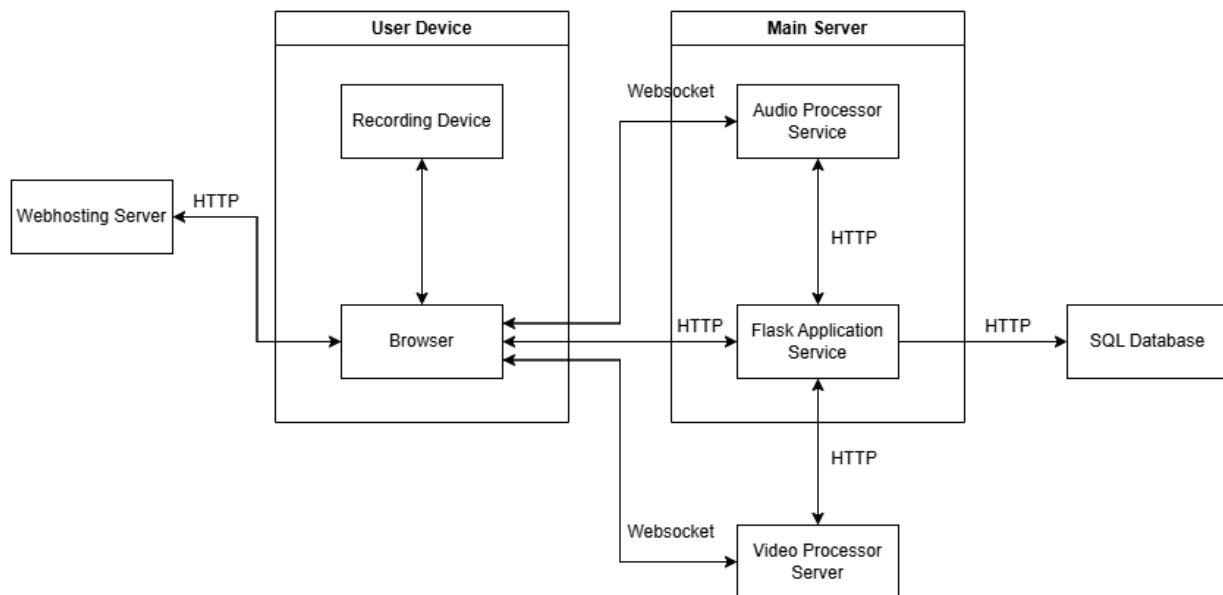


Figure 7.1: Blinc Recent Architecture Diagram

7.3.1 Audio Features

Based on the previous system refinement, I first expanded the audio analysis. This was important to provide individualized per-speaker metrics, as the students had requested from prior surveys. Figure 7.2 shows an overview of how the audio data is processed to calculate different features. I expanded the processing by improving speaker tracking and adding a participation-and-impact style analysis. The participation and impact style analysis is enabled by speaker tracking, which accurately segments the audio to detect the desired per-speaker metrics. I'll go through how I implemented speaker tracking in the system, then detail how I added the participation and impact style analysis.



Figure 7.2: Map of audio data flow

Improved Speaker Tracking

A new speaker-tracking algorithm was implemented to improve diarization accuracy. Speaker-tracking supports understanding collaboration's **Contribution**, **Communication**, and **Context**. Speaker tracking accurately records contributions and adds context to transcripts by indicating who is participating in the collaboration. This tracking also supports further communication analysis by enabling discourse analysis, as the flow of the collaboration can be dissected with speaker labels. The speaker-tracking system used a neural network, ECAPA-TDNN [156], trained to embed short audio segments for speaker identification. The generated embeddings can then be compared using cosine similarity, since two segments from the same speaker will differ by a threshold. This newer neural network model was found to be much better at handling background noise and much easier to implement across different audio formats than the pyAudioAnalysis [157] model, which was initially implemented for speaker tracking.

Join Discussion

Please type your name and passcode to join a discussion.
If rejoining a discussion, type the same name you used previously.

Device Name:
Name

Numbers of Collaborators:
0(Automatic)

Passcode:
Passcode (4 characters)

Join With:
Audio

Connect to Server

Khalil

Please add a Speaker Fingerprint for each speaker
Each speaker must record and temporarily store a short 5-7 second sample of their voice. This is used to track each speaker's position throughout a discussion and is deleted upon ending the discussion.

Speaker 186

Speaker 187

Join Discussion

Figure 7.3: BYOD Join and Speaker Fingerprinting

To use this model, the current speaker-tracking feature allows students to specify the number of speakers in a recording, as shown in Fig. 7.3. Then, users record a 5-second "fingerprint" of the voices they want to track on the device, and they can also rename the speaker labels. The fingerprint set is temporarily stored on the server and embedded using ECAPA-TDNN. On each

incoming audio segment with a transcript, the segment is also embedded using the same model. Each fingerprint is used to calculate the cosine similarity between itself and a new audio segment. The speaker label of the fingerprint with the highest value above a matching threshold matches that of the new segment.

Transcripts from COLLEGEBOUND		Options
00:08:05	Oh, for now just leave it.	
00:08:07	Yeah, just leave it just leave it. Okay so we have	
00:08:12	Speaker 178: 3 minutes. You can't really do anything with the compass that you have no idea. Yeah.	
00:08:19	Speaker 179: I don't know if we're trying to like Escape or just survive, but probably just survive first	
00:08:23	Speaker 179: What do you think? I said I was saying I'm like, I don't know if we're trying to like Escape or survive. I feel like we're not going to be able to escape. Well it it could be it could be either.	
00:08:55	Like um, it's not just going to run out of battery too.	
00:09:00	Speaker 179: And then what are we going to do with the Rope though? I feel like I don't know.	
00:09:07	Going to do anything with that cuz like you can set up a trap a trap that's true you could also use it to have like a cover like some kind of protection.	
00:09:21	Speaker 179: So, we have to attend the hammer and the roing Nicolas fishing. Yeah.	
00:09:40	Speaker 179: You can also use the hammer to like secure the tent so it doesn't like blow. That's true.	
00:09:46	Speaker 180: Nothing but really cut, like, cut stuff there.	
00:09:50	Speaker 180: I would like a knife. Yeah, yeah.	
00:09:57	Speaker 179: Walking somewhere and like get something on our foot. We could like	
00:10:01	Speaker 179: die from that. Yeah.	
00:10:03	Speaker 178: Or something. Yeah, but I feel like you can just use like this thick and put like the fish in there.	
00:10:12	Just rotate it or something	

Figure 7.4: Discussion Transcript with Speaker Labels

While students can enter speaker fingerprints, which increases the difficulty of speaker diarization, an automatic speaker detection method is also implemented when users select zero speakers. Because the number of speakers is unbounded, this can sometimes result in an unexpectedly large number of speakers in a discussion. By enabling the addition of speaker fingerprints, the task shifted from continually segmenting and clustering the audio to finding matches. Notably, a more robust and accurate speaker-tracking algorithm offers numerous benefits beyond improved audio segmentation. With speaker labels available, the data can reliably estimate individual collaboration metrics for each speaker. This will further allow student to identify their transcripts as shown in Fig. 7.4. This aligns with students' survey responses, aiming to identify their contributions better and track both their own and the group's collaborative performance.

Participation and Impact Style Analysis

To help students better understand their collaboration, the system added new, individualized speaker metrics. Participation and Impact Style analysis quantifies several measures of collaboration [72] by providing methods for measuring participation, internal relevance, external relevance, influence, freshness, and compactness individually. These features enable a deeper understanding of the **Communication** flow and the effects of people’s **Contributions**.

Feature	Description
Participation	General level of an individual’s participation
Internal Relevance	The tendency of an individual to relevance or novelty in their contributions
External Relevance	The tendency of an individual to respond, or not, to the contributions of their collaborative peers
Influence	The tendency of a participant to evoke responses, or not, from their collaborative peers
Discussion Novelty	The tendency to provide new information or to echo previously stated information, irrespective of the originator of the information
Compactness	The extent to which participants convey information in a concise manner
Discourse Analysis	Dissecting the topical and rhetorical structure of contributions to detect specific collaboration indicators in speech

Table 7.1: Participation and Impact Style Measures

Given the speaker labels, the participation value is the fraction of the speaker’s spoken audio segments among the total transcribed audio segments. To calculate the other metrics, semantic modeling is required to assess similarity between audio segments. The transcribed audio segments must first be embedded into the vector space provided by a sentence transformer. All-mpnet-base-v2 is a strong model that maps to a 768-dimensional vector space. This model was trained to support clustering and semantic search. Once the segments are embedded, cosine similarity can be computed between pairs of segments to calculate the other Participation and Impact Style metrics.

Using the speaker labels, the participation can be calculated. The speaker-label sequence, which contains the speaker index for each segment, also serves as the contribution sequence C used in the following calculations. Within C , c_t is the t^{th} contribution, while $n = |C|$, the length

of the sequence. P is the set of participants, and $k = |P|$, the number of participants. a and b are sample participants. The participation function is then defined below as:

$$p_a(t) = \begin{cases} 1, & \text{if contribution } c_t \text{ was made by participant } a \in P \\ 0, & \text{otherwise.} \end{cases} \quad (7.1)$$

This then means each participant, a , has their participation function represented as,

$$P_a = \{p_a(t)\}_{t=1}^n = \{p_a(1), p_a(2), p_a(3), \dots, p_a(n)\} \quad (7.2)$$

By summing the participation sequences for each student, the total number of contributions for each student, as shown in (7.3), can be used to determine their contribution proportion relative to the total, as shown in (7.4).

$$\| P_a \| = \sum_{t=1}^n p_a(t) \quad (7.3)$$

$$\bar{p}_a = \frac{1}{n} \| P_a \| \quad (7.4)$$

The proportion of students' participation can then be used to score their participation based on the number of students involved in the collaboration equation, as shown in (7.5).

$$\bar{p}_a = k\bar{p}_a - 1 \quad (7.5)$$

After the participation score, a few other precursor metrics must be calculated before determining the other Participation and Impact Style metrics. This was the cross-cohesion. Cross-cohesion is essential for relating two temporal series of data. In this case, I want to compare two contributions. To determine cross-cohesion, a means of comparing two contributions is needed; here, semantic similarity is used. Semantic similarity refers to the difference in meaning between two different contributions.

Calculating semantic similarity involved using a sentence transformer embedded vector calculated from the incoming audio. With these embeddings, cosine similarity as shown in (7.6) can be used to determine the semantic similarity as defined in (7.7). Mainly, within a specified time segment called a lag τ , all contributions will be compared in semantic similarity. Cross cohesion is finally defined as the average semantic similarity between two people's contributions with a lag τ as shown in (7.8).

$$\cos(\vec{d}_t, \vec{d}_u) = \frac{1}{n} \| P_a \| \quad (7.6)$$

$$s_{ab}(t, u) = p_a(t) \cdot p_b(u) \cdot \cos(\vec{d}_t, \vec{d}_u) \quad (7.7)$$

$$\xi_{ab}(\tau) = \begin{cases} 0, & \| P_{ab} \| = 0 \\ \frac{1}{\| P_{ab} \|} \sum_{t=\tau+1}^n s_{ab}(t - \tau, t) & \| P_{ab} \| \neq 0 \end{cases} \quad (7.8)$$

The sum of similarity within the lag τ is averaged by the number of contributions between the two speakers with the same lag window.

$$\| p_{ab}(\tau) \| = \sum_{t=\tau+1}^n p_a(t - \tau) \cdot p_b(t) \quad (7.9)$$

ξ is the cross cohesion function while $\xi_a b$ is the τ -lagged cross-cohesion of a to b. This measures the semantic similarity of all τ -lagged contributions between speakers a and b. With this, if in a given conversation there are $k=|P|$ participants, the cross-cohesion can be represented by a $k \times k$ matrix $X(\tau)$ where row i and column j in the matrix is equal to $\xi_{ij}(\tau)$. Then, relevance across a window w is defined as in (7.10). This creates a matrix of moving averages of the relevance of each speaker to each other, where $r_{ab}(w)$ is the w-spanning relevance of a to b.

$$R(w) = \frac{1}{w} \sum_{\tau=1}^w X(\tau) \quad (7.10)$$

This matrix can be used to determine Internal and External Relevance and Influence. First, the internal relevance can be seen as the diagonal of this relevance matrix or as $r_{aa}(w)$. For the external relevance calculation, we ignore these values in the matrix and compute the mean of the

row sums. In this case, each row will be representative of a speaker as shown in (7.11)

$$\bar{r}_a(w) = \frac{1}{k-1} \sum_{i=1; i \neq a}^k r_{ai}(w) \quad (7.11)$$

Influence is very similar to external relevance, but instead of the column sum, the column average is calculated.

$$\bar{i}_a(w) = \frac{1}{k-1} \sum_{j=1; j \neq a}^k r_{ja}(w) \quad (7.12)$$

After calculating these metrics, they would be stored in the database alongside the transcript and displayed to the student on the front end, as shown in Fig. 7.5. With these metrics, students could now gather information about how they are participating and semantically in relation to other students

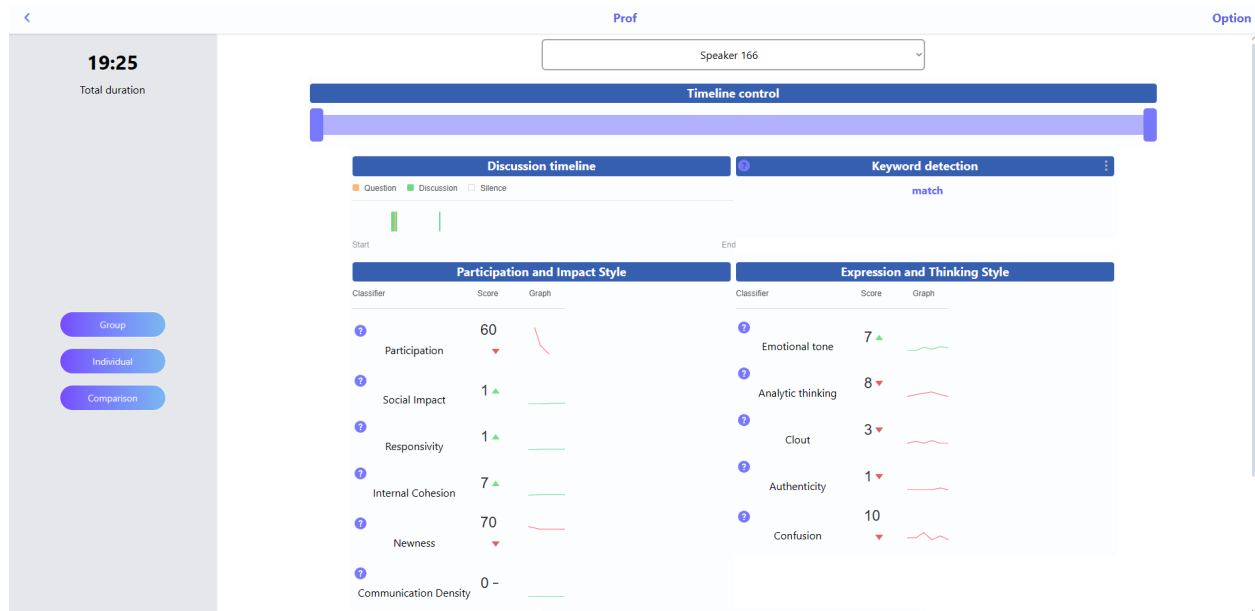


Figure 7.5: Participation and Impact Style

7.3.2 Video Features

A lot of insight can be gained from audio alone, but using a single modality limits the range of features that can be detected and the flexibility of their detection. Notably, incorporating additional modalities can improve the system's accuracy. This led to the addition of video processing, as many of the aspects of a collaboration's dynamics can be detected from video. The video can be beneficial for validating the detection of emotional states and for detecting students' joint attention throughout a collaboration. In this section, I'll go through how these video features were implemented.

An overview of the video processing is shown in Figure 7.6. The video recording was initially set up via a remote connection to the computer in the lab. Students can join the same way as before, but can choose to participate with audio, video, or video with cartoonification. Participants' video data is captured in 10-second intervals and converted into frames. These frames can then be further processed to detect visual indicators of collaboration, such as facial sentiment analysis, joint attention, and others.

Person and Object Detection

To detect these indicators, the system must identify all objects in a given frame, thereby determining which individuals to track and what may be related to their collaborations. The system uses the YOLOv4 Convolutional Neural Network (CNN) object detection algorithm [158] to perform this detection. This model primarily segments an image into different object areas using a part of the model trained on the COCO dataset. This detection further supports tracking people's Contributions, similar to speaker tracking, but with visual data, enabling the visualization of who is actively speaking. This also adds to understanding the **Context** by detailing who is participating.

The object areas in this detection can then be identified based on a part of the model trained on the ImageNet dataset. Once these objects are detected, it becomes much easier to track the pixels within each object while ignoring minor changes due to object motion. With this, we can identify the active participants in the collaboration and any potential resources they may be using.

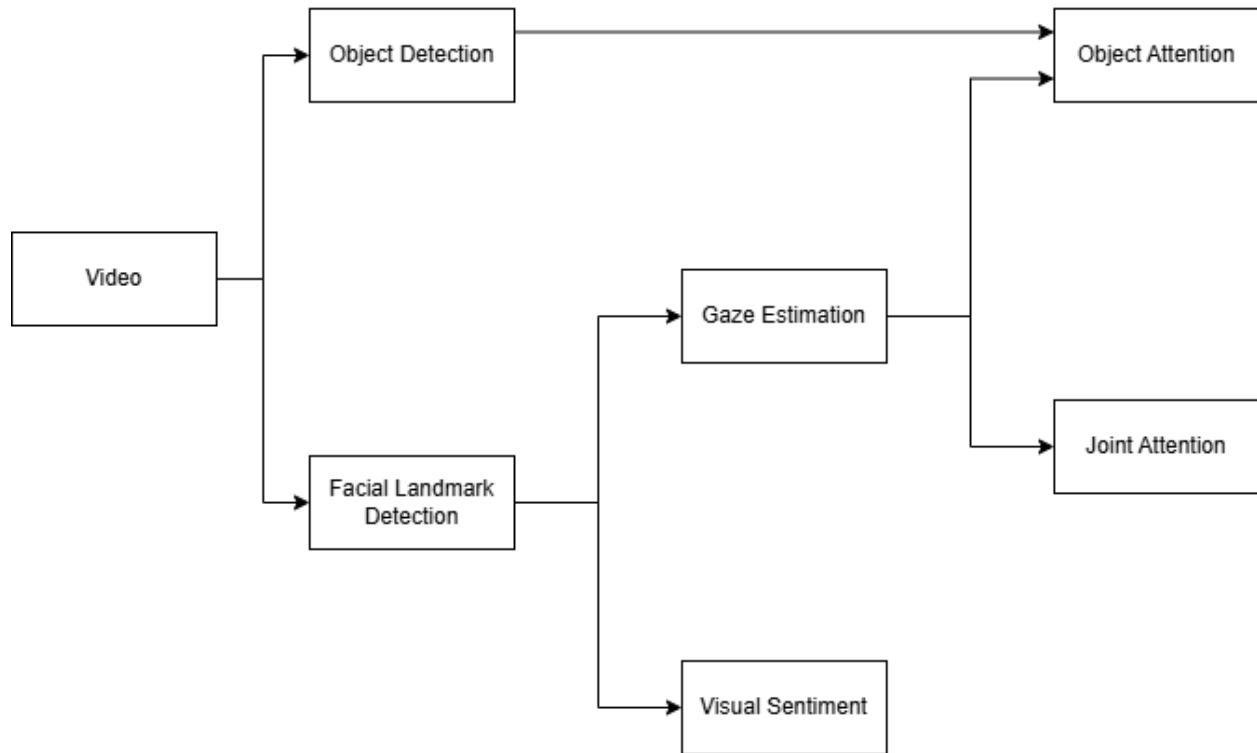


Figure 7.6: Map of video data flow

Facial Sentiment Analysis

By detecting students in the collaboration, the system can identify visual indicators of collaboration. One of these indicators is students' sentiment, which helps understand the **Climate** of a collaboration. Understanding visual sentiment will also deepen our understanding of **Communication**, enabling us to examine people's reactions to communication and the emotions someone may be trying to convey. To detect this, the system first needs to identify students' faces. The previous object detection detected the whole person, not just the head. Hence, another detection algorithm is necessary to isolate participants' heads for further sentiment analysis. The system used YOLOv5 to perform this head detection on people within the collaboration. Once the entire body of people is detected, the pixels within the detection bounding box are fed into YOLOv5 to reduce the computational cost of face detection. The participants' faces can then be used to detect their emotions, which are then passed to Dlib's emotion detection model [159].

Joint Attention

Another visual indicator of collaboration that the system can detect is joint attention. This is the number of collaborators focusing on the same object. This can enable a more detailed analysis of **Communication** by revealing with whom someone is speaking and what they may be discussing. This provides further **Context** into what is happening within a collaboration. In addition to the previously detected student faces, this included other pose information, such as students' gaze direction. Gaze direction can be used to infer what students are looking at. With this, the system can determine when two students' gaze is focused on the same object, as detected by the previous object-detection algorithm. This allows the system to measure students' joint attention throughout a discussion.

Cartoonification



Figure 7.7: Cartoonification Sample

The implementation of the facial cartoonification feature in the BLINC platform de-identifies participants' faces, rendering them distinctly different from their original appearance while retaining essential facial features, as shown in Figure 7.7. Within each frame, facial detection and

landmark prediction are performed utilizing a pre-trained machine-learning model based on the Histogram of Oriented Gradients from Dlib[159]. To achieve the desired cartoonification effect, image-based style transfer is subsequently applied to the predicted facial landmarks. The VToonify facilitates this transformation using a pre-trained machine-learning model [160].

The real-time streaming of these transformed frames to the dashboard completes the process, allowing for continuous monitoring of the de-identified facial representations during collaborative discussions on the BLINC platform. This meticulous methodology improves participants' privacy while maintaining the analysis. This is important to support a healthy **Climate**, as some students experienced intense negative feelings in the presence of recording devices, and to maintain a sense of privacy.

7.3.3 User Interface

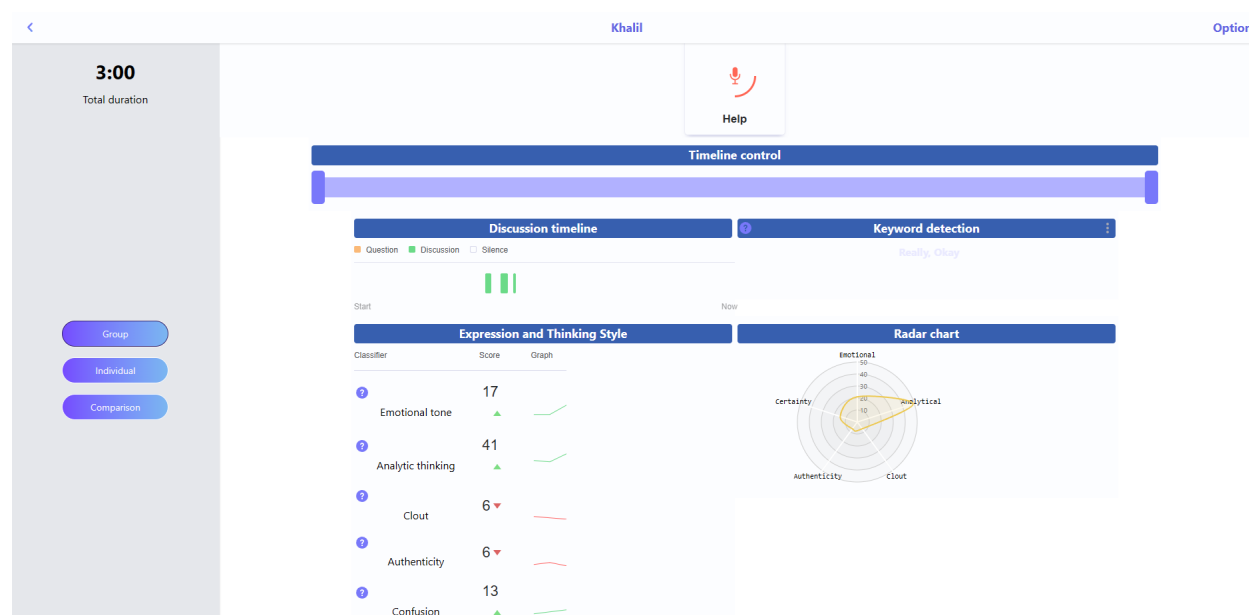


Figure 7.8: BYOD Live Connection Sample

To improve student interactivity, a crucial initial change was to allow Bring-Your-Own-Device (BYOD) connections to view their data interactively, enabling students to manage their collaborations more actively. In my initial user study, I showed students their data after the discussion to help them understand what was being collected. However, students were unable to view their data

in real time while collaborating. A workaround was to sign in students with the account used to set up the discussion and select their device for them. The problem was that students would also have access to the management of the other devices and could easily get distracted by different data or even view other students' discussions without permission. This led to the development of the student view when connected from a non-ReSpeaker device.

The student's view is the same as the teacher's, with minor differences. For instance, a student can only see their device's data, whereas the teacher can view data from all active devices on the account. Students can also ask for help through the system, which will notify the teacher who set up the discussion in the multi-device view. Students can also control when they end a conversation if they know that recording is no longer necessary and desire privacy. Additionally, students can temporarily pause processing to provide short-term privacy. This led to other vital changes to support students in managing their collaborations from their devices rather than from the ReSpeakers, as my focus shifted from the ReSpeakers to the variety of user devices students may have.

Reactive UI



Figure 7.9: BLINC Data Display Desktop vs Mobile

The need to adapt to students using their own devices grew during COVID-19. With students not physically present in classrooms, I began to focus the design on BYOD support. This furthered the goal of providing a student-facing UI, since most user devices already offered one. A problem arose as the number of devices attempting to connect increased substantially, necessitating a more flexible UI. This involved using ReactJS to create a reactive UI that could scale and adapt its interactivity to the device on which it was used. The reactive UI will support the different **Contexts** I've needed to adapt to.

The original system was implemented in Angular and can be adjusted to accommodate different device screen sizes. Still, issues arose when using Angular, as many undergraduate students who attempted to work on the project had little confidence in TypeScript and Angular, despite their similarities to React. This made it challenging to recruit and onboard students to aid in the system's development, a task that became necessary after Sony's participation concluded. Recruitment could allow me to delegate some of the more minor yet essential changes. At the same time, I could focus on more complex tasks such as adding new analytics, integrating new features, supporting new modalities, and deploying the system to a scalable architecture. This forced a rewrite of Sony's Angular front-end to a ReactJS front-end to enable better long-term support, onboarding, and reactive design. As a result, much of the original functionality was maintained.

The initial system was designed for mobile devices, so many elements were arranged in a single column. When viewed on a desktop, most of the screen was white space, which can be good in moderation, but this was limiting, given how much scrolling was necessary to see the added features and the available space. Now, on the desktop, the elements will attempt to fill most of the screen, leaving white space in areas to provide spacing between groups of elements. This means on most modern widescreen monitors, no scrolling is necessary to view all of the data. By implementing an adjustable UI design, our system supported a wider range of devices smoothly. This was one of the earliest methods for improving interactivity.

Compare and Contrast

In addition to improving the design's flexibility, a new method for comparing students' data was introduced. Many students sought context on what their values meant and how they compared to those of their peers during their collaborations. Similarly, in prior studies, comparing and contrasting information about collaborators improved students' collaborative outcomes. This strongly suggested that the design needed a way for students to compare their data with their collaborators' and other students'. This was one of the essential reasons for accurately tracking the speakers throughout a collaboration. This will allow students to compare their types of **Contributions** within their **Context**.

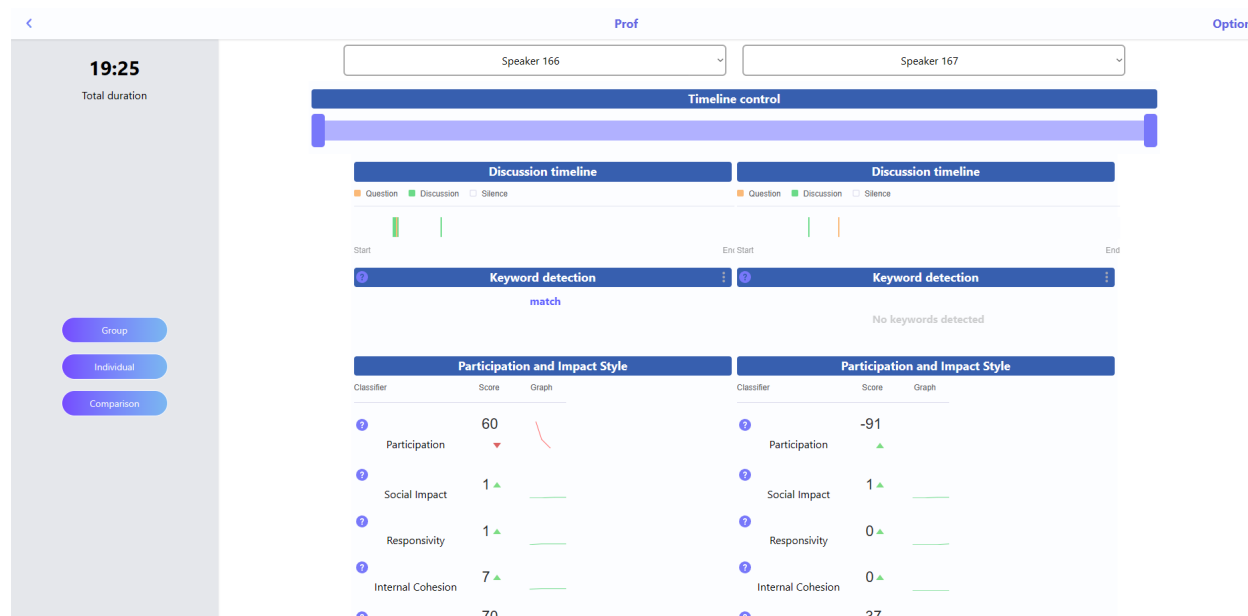


Figure 7.10: Compare and Contrast

With the speakers accurately labeled, a new interface was added that displays each speaker's metrics side by side, as shown in Fig. 7.10. Speakers can be selected from the drop-down, and the timeline can be moved to change the current audio set, as with the individual and group interfaces. This added a third button to the device metric interface, allowing students to toggle between the group, individual, and comparison data. This interface will enable students to compare themselves in real time and quickly after the discussion.

Post Discussion Meta-Survey

During the surveys, students requested ways to peer-review other students. One way to do this was to administer post-discussion surveys that asked students about various aspects of the collaboration. After the discussion ends, a pop-up appears, as shown in Fig. 7.11, with a QR code that links to a survey students can complete. The results of these surveys could be shared with students to provide them with their peer reviews. These views could also be shown over time to reveal any trends. The survey questions are a significant factor in determining the insights. The right questions enable students to **Constructively** think about their collaborations. The survey can also ask questions around students' **Compatibility** and how they managed **Conflicts**.

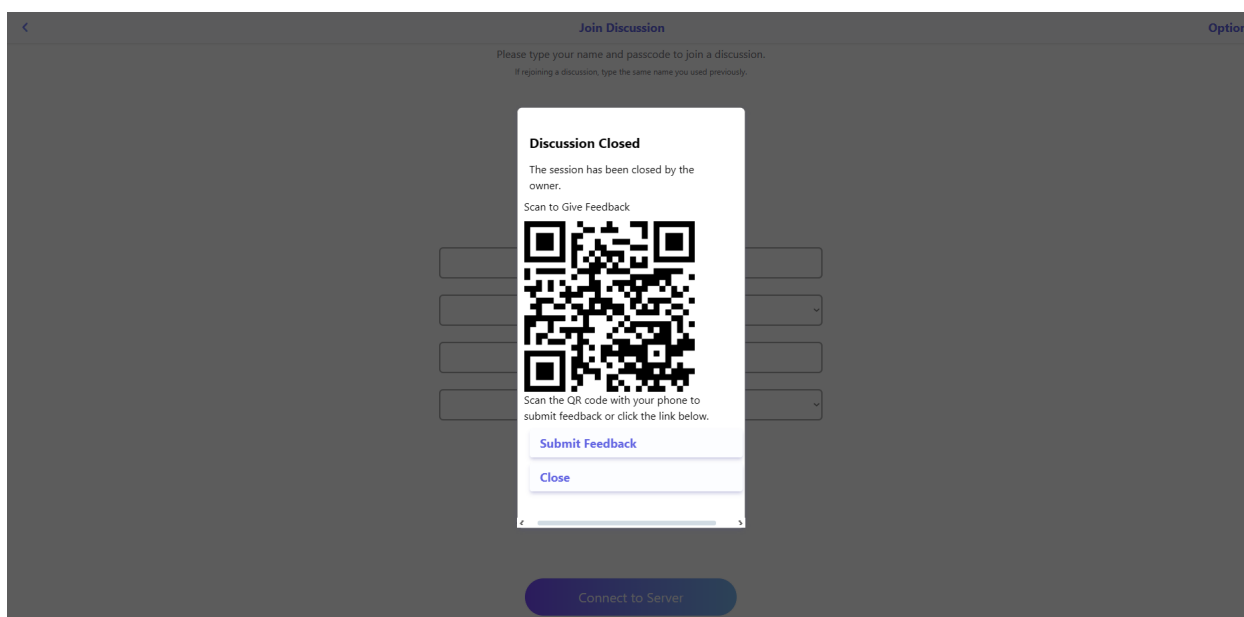


Figure 7.11: Post Discussion QR Code popup

The survey questions were designed to provide meta-cognitive insights into their collaborations. The goal is that, by providing these questions, students will be forced to reflect on their cognitive processes regarding overall group collaboration and on how their team's processes have unfolded. Group meta-cognition [161] includes four dimensions: knowledge of cognition, planning, monitoring, and evaluating. Knowledge of cognition encompasses the awareness of group learning strategies and the effective use of information. Planning includes selecting the most effec-

tive approach and managing the workload. Monitoring involves consistently checking for errors and reevaluating strategies. Evaluating consists of recognizing the group's capacity to make informed judgments about teamwork, methods, and outcomes.

The format of the questions is essential in prompting students to consider these meta-cognitive areas. Experts validated these questions to ensure clarity. For each question, students indicate their level of agreement with the given statement. By asking meta-cognitive questions, we can help students reflect on their understanding of cognitive collaborative processes and how they regulate them. These questions included:

1. We know our strengths as learners
2. We know how to select relevant information
3. We know how to use the material
4. We know how to organize new information
5. We know how to connect new information with prior knowledge
6. We plan the activities
7. We determine what the task requires
8. We select the appropriate tools
9. We identify the strategies depending on the task
10. We organize our time depending on the task
11. We modify our work according to other group participants' suggestions
12. We ask questions to check our understanding
13. We check our approach to improve our outcomes
14. We improve our work with group processes
15. We detect and correct errors
16. We make judgments on the difficulty of the task
17. We make judgments on the workload
18. We make judgments on the instruments

19. We make judgments on our learning outcomes
20. We make judgments on the teamwork process

The figure shows a 'Session Feedback' form with 19 questions. The questions are as follows:

1. We know our strengths as learners
2. We know how to select relevant information
3. We know how to use the material
4. We know how to connect new information with prior knowledge
5. We plan the activities
6. We determine what the task requires
7. We select the appropriate tools
8. We identify the strategies depending on the task
9. We organize our time depending on the task
10. We modify our work according to other group participants' suggestions
11. We ask questions to check our understanding
12. We check our approach to improve our outcomes
13. We improve our work with group processes
14. We detect and correct errors
14. We detect and correct errors
15. We make judgments on the difficulty of the task
16. We make judgments on the workload
17. We make judgments on the instruments
18. We make judgments on our learning outcomes
19. We make judgments on the teamwork process

Each question is followed by five buttons: 'Strongly Disagree', 'Disagree', 'Neutral', 'Agree', and 'Strongly Agree'. A 'Submit Feedback' button is at the bottom right.

Figure 7.12: Post Discussion Questions

7.4 Performance Evaluation

This section provides an evaluation of the system's performance in a live-deployed system.

7.4.1 Experimental Setup

Parameter	Value
Instance	c7i.xlarge
vCPU cores	4
Threads per core	2 Threads
CPU clock	3.2 GHz
RAM	8 GB
Audio Sample Rate	16 kHz

Table 7.2: Hardware Parameters

To evaluate the system, it was deployed to an AWS server. The hardware parameters are detailed in table 7.2. This was a c7i.xlarge instance with 8GB of RAM and four vCPUs at 3.2 GHz,

with two threads per core. An AWS t4g.large MySQL database was used to store the data, while an ElastiCache Redis was used for caching. The audio would be converted to 16 kHz and processed. The audio processor, discussion capture server, and NGINX web server were all run on the same instance.

Test	Devices	Speakers per Device	Test Length
1	20 Devices	1 Speaker	28:44
2	1 Device	4 Speakers	16:26
3	4 Devices	4 Speakers	19:25

Table 7.3: Testing Parameters

Three live tests were planned to evaluate the system’s performance. The first test was a stress test with 20 devices, each recording a single speaker, to test multiple devices simultaneously. This was done during a remote lab meeting. A second test used a single device and four speakers; it was conducted during a collaboration among four students in a summer CS class. The final test used four devices, each recording four speakers, to stress-test speaker tracking; this was conducted during a fall CS class. The testing parameters are listed in Table 7.3.

Test	Average Processing Latency	Top CPU Utilization
1	0.366 seconds	18.1%
2	0.107 seconds	13.2%
3	0.658 seconds	8.09%

Table 7.4: Testing Results

7.4.2 Processing Latency

This was a measure of the time required for the system to send the processed data back after receiving an audio signal. Across all tests, we observe the system maintaining a latency under 1 second. This is accompanied by a slight increase in system latency with the number of devices, and a more significant increase due to speaker tracking.

7.4.3 Hardware Usage

The CPU usage never exceeds a quarter of the system's resources, indicating significant CPU overhead. Although this meant that a smaller instance might be possible, when loading the models and starting a processing session, the system would crash on instances with less than 8GB of RAM. The models are loaded only once, regardless of the number of connected devices; we also observe in Table 7.4 that increasing the number of devices or speakers increases CPU usage.

7.5 Discussion and Limitations

The system test was significant for evaluating its performance and the viability of supporting multiple groups simultaneously. This allowed me to discuss some of the evaluation's implications and to determine possible future directions for continuing the system's development.

To ensure the system can provide near-real-time feedback, processing latency must be minimized. The tests show that the system can maintain a processing latency under 1 second, but speaker tracking increases latency across multiple devices. This means there will need to be a limit on the number of speakers that can be used, given the number of connected devices, or it may be helpful to implement a different speaker-tracking algorithm.

The system did not use the majority of the multi-core CPU architecture, and reducing system RAM had catastrophic effects on speaker tracking. This would cause the system to crash when a device attempts to join and load all the models needed to process the data. It would be possible to decouple the audio processing server and the discussion capture server, as the audio processing server consumes most of the RAM when loading audio processing models into memory. This could allow for two distinct instances, each designed for a server's purpose, rather than a single combined server. This could mean an AWS memory-optimized server for audio processing and a general-purpose server for the web application. Any video processing would also require an additional server, not only because of the limited memory on the instances, but also because it requires a GPU.

Although the video had been integrated into the system, supporting multiple devices simultaneously was difficult due to cartoonification, leading to significant processing latency. This is important to note, as incorporating the video may require reducing its frequency to allow processing to scale more effectively with the other system components.

The system evaluation shows how well the system performs across the provided setups. A system that accounts for students' perspectives can run smoothly across multiple devices and groups while also providing real-time collaborative feedback. This system illustrates how to integrate students' views into the design to prioritize the development of their collaboration literacy. Further testing over extended periods should be conducted to measure and compare the system's collaboration literacy development.

7.6 Conclusion

Throughout this chapter, I detailed a set of design goals informed by analyses from previous chapters and prior work in automated collaboration development. These are student-centered design goals that could be used in other collaboration analytics systems. These design goals guided the development of the BLINC collaboration literacy tool and enabled the system to align with students' perspectives. This is an essential step in ensuring that an educational tool is as helpful as possible for students, and I subsequently ran a live deployment of the design for testing.

During testing, the system supported 20 devices with individual speakers and four devices with four speakers each, maintaining near-real-time performance within 1 s. These tests allowed analysis of the system's performance. In these tests, the system incurred significant CPU overhead, and RAM usage exceeded its limit when the system was too small. The cartoonification also had minor issues across multiple devices, and further work is needed to improve real-time performance. Despite these issues, the system's promising audio performance demonstrated its ability to integrate students' perspectives while providing tailored collaborative support.

Chapter 8

CONCLUSION

Engaging students directly in their collaboration literacy development is necessary to maximize it. By centering the student perspective in the system's design, I can ensure the system supports students at their current level of collaboration literacy. This dissertation explored the student perspective, and a thematic analysis was conducted to develop design goals for a collaboration literacy development system called BLINC. An implementation of BLINC was tested to assess the design's viability. The main contributions of this dissertation are the development and application of design goals for collaboration analytics that center on students and are supported by other well-studied design goals in LA. This is accompanied by a detailed understanding of students' views on collaboration and on being supported by LA tools in this context.

I started this work by designing a prototype system to measure different aspects of collaboration. This was intended to indicate the feasibility of further system development. In this work, I focused on developing collaborative analytics, while Sony, in a partnership, focused on the user interface design. A second study was then conducted to dissect students' conceptualizations of collaboration. This allowed the development of the 7 C's of collaboration—**Climate, Communication, Compatibility, Conflict, Context, Contribution, and Constructive**—which were used in thematic analysis throughout the dissertation. This allowed future feedback to be conceptualized in this way, ensuring that features support students' understanding of collaboration.

Further student surveys were conducted to understand the interactions students sought from a collaboration analytics tool, and another to assess students' ethical and privacy concerns. The analysis of these surveys provided the basis for developing student-centered design goals for BLINC and other collaboration literacy development tools. These goals included: **Data Transparency, Student Autonomy, Prioritizing Learning over Assessment, Positive Reinforcement, Student Customization, and Setting Flexibility, along with some feature-specific goals.** With the system

design detailed, I performed a performance evaluation to assess its viability and limitations. This demonstrates that the audio analysis performed consistently across multiple devices and speakers, indicating that aligning the system with students doesn't sacrifice performance.

These findings present a new set of design goals for the LA system, which aims to be student-centered. In addition, a prototype tool to support students' collaboration literacy development has shown promising results. Future student-centered LA can use and apply the design goals to align their system with students. At the same time, this prototype can be tested to improve collaboration literacy over a longer time frame. Furthermore, these design goals can help incorporate new metrics or modalities into the system and should be continually reviewed and iterated upon to adapt to a constantly changing world.

8.1 Future Work

This work can take several directions in the future. This dissertation has sought to understand students' conceptualization of collaboration using the 7 Cs of collaboration, to measure aspects of collaboration, and to provide feedback about collaborations. Still, more work is needed to clearly differentiate between novice and expert collaborators. Being able to identify levels of collaboration literacy mastery actively would help demonstrate how the system supports the development of collaboration literacy. This would also mean a formalization of the specific skills that entail collaboration literacy. This could serve as a guide for measuring mastery.

Another direction for future research is the feedback provided to students. This dissertation has sought to make multimodal feedback available to students that actively and constructively engages them in their collaborations, such as having them complete surveys or compare different sessions or speakers. In following the ICAP framework, interactive feedback is the best type of feedback for learners. The development of LLMs and AI agents presents an opportunity to leverage them, along with the calculated collaboration analytics, to provide interactive feedback on collaborations. Much work would need to be done to validate the feedback provided and to determine the types of interactions that integrating with an AI agent could enable, to ensure the safety of use with

students.

A third direction is the integration of additional modalities and the improvement of video integration. The video presents different ways to analyze existing metrics and to enable new metrics. Importantly, adding more modalities can provide additional means to capture all collaborators' contributions, since not all collaboration is verbal. In addition to video data, other modalities could be examined to support holistic analysis of students' collaborations. Importantly, the integration of these modalities must be balanced against the processing complexity they introduce to the system, and they must be carefully adapted to the system.

A final direction for future work could focus on the effect of errors on students and their effectiveness. While students rarely noticed errors, it could be jarring when they did. At times, the errors would distract students entirely from the collaboration. Even so, this didn't limit most students' overall engagement with the system metrics, as most transcripts were representative. Future work could examine the effects of perceived errors and further analyze students' perceptions of them. Overall, this work opens multiple research directions for future work, and these four are among the strongest, based on my dissertation.

REFERENCES

- [1] Government of India, *The digital personal data protection act, 2023*, tex.howpublished: Published in The Gazette of India, 2023.
- [2] U. D. of Education, *Family educational rights and privacy act (FERPA)*, tex.howpublished: 20 U.S.C. § 1232g; 34 CFR Part 99, 1974.
- [3] E. Parliament and Council, *General data protection regulation (GDPR)*, Number: 2016/679 tex.entrytype: regulation, 2016.
- [4] Brazilian Government, *Lei Geral de Proteção de Dados Pessoais*, tex.howpublished: Sanctioned by Law No. 13.709 of August 14, 2018. Official text available from the Planalto website, Aug. 2018.
- [5] M. Cukurova, R. Luckin, E. Millán, and M. Mavrikis, “The NISPI framework: Analysing collaborative problem-solving from students’ physical interactions,” *Computers & Education*, vol. 116, pp. 93–109, Jan. 2018.
- [6] J. Roschelle and S. D. Teasley, “The Construction of Shared Knowledge in Collaborative Problem Solving,” in *Computer Supported Collaborative Learning*, C. O’Malley, Ed., ser. NATO ASI Series, Berlin, Heidelberg: Springer, 1995, pp. 69–97, ISBN: 978-3-642-85098-1.
- [7] M. Worsley and X. Ochoa, “Towards Collaboration Literacy Development through Multimodal Learning Analytics,” 2020.
- [8] OECD, *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic, Financial Literacy and Collaborative Problem Solving (PISA)*. OECD, Aug. 2017, ISBN: 978-92-64-28184-4 978-92-64-28182-0.
- [9] D. Wood, J. S. Bruner, and G. Ross, “THE ROLE OF TUTORING IN PROBLEM SOLVING*,” *Journal of Child Psychology and Psychiatry*, vol. 17, no. 2, pp. 89–100, Apr. 1976.
- [10] D. J. Deming, “The Value of Soft Skills in the Labor Market,” *NBER Reporter*, no. 4, pp. 7–11, Dec. 2017, Publisher: National Bureau of Economic Research, Inc.
- [11] M. Binkley *et al.*, “Defining Twenty-First Century Skills,” in *Assessment and Teaching of 21st Century Skills*, P. Griffin, B. McGaw, and E. Care, Eds., Dordrecht: Springer Netherlands, 2012, pp. 17–66, ISBN: 978-94-007-2324-5.

- [12] L. Wilson, S. Ho, and R. H. Brookes, “Student perceptions of teamwork within assessment tasks in undergraduate science degrees,” *Assessment & Evaluation in Higher Education*, vol. 43, no. 5, pp. 786–799, Aug. 2018, Publisher: Routledge.
- [13] E. R. Lai, “Collaboration: A literature review,” *Pearson Publisher*. Retrieved November, vol. 11, p. 2016, 2011.
- [14] S. Rojas-Drummond and N. Mercer, “Scaffolding the development of effective collaboration and learning,” *International Journal of Educational Research*, vol. 39, no. 1, pp. 99–111, Jan. 2003.
- [15] S. K. D’Mello, N. Duran, A. Michaels, and A. E. B. Stewart, “Improving collaborative problem-solving skills via automated feedback and scaffolding: A quasi-experimental study with CPSCoach 2.0,” *User Modeling and User-Adapted Interaction*, Feb. 2024.
- [16] M. Hernández-de-Menéndez, R. Morales-Menendez, C. A. Escobar, and R. A. Ramírez Mendoza, “Learning analytics: State of the art,” *International Journal on Interactive Design and Manufacturing (IJIDeM)*, vol. 16, no. 3, pp. 1209–1230, Sep. 2022.
- [17] S. Praharaj, M. Scheffel, H. Drachsler, and M. Specht, “Literature Review on Co-Located Collaboration Modeling Using Multimodal Learning Analytics—Can We Go the Whole Nine Yards?” *IEEE Transactions on Learning Technologies*, vol. 14, no. 3, pp. 367–385, Jun. 2021, Conference Name: IEEE Transactions on Learning Technologies.
- [18] M. Boothe, C. Yu, A. Lewis, and X. Ochoa, “Towards a Pragmatic and Theory-Driven Framework for Multimodal Collaboration Feedback,” in *LAK22: 12th International Learning Analytics and Knowledge Conference*, ser. LAK22, New York, NY, USA: Association for Computing Machinery, Mar. 2022, pp. 507–513, ISBN: 978-1-4503-9573-1.
- [19] H. Javed and N. Jamali, *Group Dynamics: Survey of Existing Multimodal Models and Considerations for Social Mediation*, arXiv:2306.17374 [cs], Jun. 2023.
- [20] V. Echeverria, M. Wong-Villacres, X. Ochoa, and K. Chiluiza, “An Exploratory Evaluation of a Collaboration Feedback Report,” in *LAK22: 12th International Learning Analytics and Knowledge Conference*, ser. LAK22, New York, NY, USA: Association for Computing Machinery, Mar. 2022, pp. 478–484, ISBN: 978-1-4503-9573-1.
- [21] A. Kruse and R. Pongsajapan, “Student-Centered Learning Analytics,” *CNDLS thought papers*, vol. 1, no. 9, pp. 98–112, 2012.
- [22] S. Blömeke, J.-E. Gustafsson, and R. J. Shavelson, “Beyond Dichotomies,” *Zeitschrift für Psychologie*, vol. 223, no. 1, pp. 3–13, Jan. 2015, Publisher: Hogrefe Publishing.

- [23] X. Huang and X. Ochoa, “Charting the Development of Collaboration Skills Through Collaborative Learning Analytics Systems,” *Journal of Learning Analytics*, pp. 1–29, Mar. 2025.
- [24] K. J. Anderson, T. Dubiel, K. Tanaka, M. Worsley, C. Poultney, and S. Brenneman, “Chemistry Pods: A Mutlimodal Real Time and Retrospective Tool for the Classroom,” in *2019 International Conference on Multimodal Interaction*, Suzhou China: ACM, Oct. 2019, pp. 506–507, ISBN: 978-1-4503-6860-5.
- [25] M. Worsley, K. Anderson, N. Melo, and J. Jang, “Designing Analytics for Collaboration Literacy and Student Empowerment,” *Journal of Learning Analytics*, vol. 8, no. 1, pp. 30–48, Apr. 2021.
- [26] B. G. Glaser and A. L. Strauss, *The discovery of grounded theory: strategies for qualitative research*, 5. paperback print. New Brunswick: Aldine Transaction, 2010, ISBN: 978-0-202-30260-7.
- [27] K. Charmaz, *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis*. SAGE, Jan. 2006, Google-Books-ID: 2ThdBAAQBAJ, ISBN: 978-1-4462-0040-7.
- [28] J. Corbin and A. Strauss, *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. SAGE Publications, Nov. 2014, Google-Books-ID: hZ6kBQAAQBAJ, ISBN: 978-1-4833-1568-3.
- [29] S. D’Mello, A. Kappas, and J. Gratch, “The Affective Computing Approach to Affect Measurement,” *Emotion Review*, vol. 10, no. 2, pp. 174–183, Apr. 2018.
- [30] S. Oviatt, A. Cohen, and N. Weibel, “Multimodal learning analytics: Description of math data corpus for ICMI grand challenge workshop,” in *Proceedings of the 15th ACM on International conference on multimodal interaction*, Sydney Australia: ACM, Dec. 2013, pp. 563–568, ISBN: 978-1-4503-2129-7.
- [31] J. Zhou, K. Hang, S. Oviatt, K. Yu, and F. Chen, “Combining empirical and machine learning techniques to predict math expertise using pen signal features,” in *Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge*, Istanbul Turkey: ACM, Nov. 2014, pp. 29–36, ISBN: 978-1-4503-0488-7.
- [32] X. Ochoa, K. Chiluita, G. Méndez, G. Luzardo, B. Guamán, and J. Castells, “Expertise estimation based on simple multimodal features,” in *Proceedings of the 15th ACM on International conference on multimodal interaction*, ser. ICMI ’13, New York, NY, USA: Association for Computing Machinery, Dec. 2013, pp. 583–590, ISBN: 978-1-4503-2129-7.
- [33] M. Worsley and P. Blikstein, “Using learning analytics to study cognitive disequilibrium in a complex learning environment,” in *Proceedings of the Fifth International Conference on*

Learning Analytics And Knowledge, Poughkeepsie New York: ACM, Mar. 2015, pp. 426–427, ISBN: 978-1-4503-3417-4.

- [34] M. Cukurova, M. Mavrikis, and R. Luckin, “Evidence-Centered Design and Its Application to Collaborative Problem Solving in Practice-based Learning Environments,” SRI International, Menlo Park, CA, Report 619738, Jan. 2017, Num Pages: 21 Publication Title: (Seventh Framework Programme of the European Commission: GA No 619738 619738). SRI International: Menlo Park, CA.
- [35] B. Schneider and R. Pea, “Does Seeing One Another’s Gaze Affect Group Dialogue? A Computational Approach,” *Journal of Learning Analytics*, vol. 2, no. 2, pp. 107–133, Dec. 2015.
- [36] B. Schneider, K. Sharma, S. Cuendet, G. Zufferey, P. Dillenbourg, and R. Pea, “Leveraging mobile eye-trackers to capture joint visual attention in co-located collaborative learning groups,” *International Journal of Computer-Supported Collaborative Learning*, vol. 13, no. 3, pp. 241–261, Sep. 2018.
- [37] M. Worsley and P. Blikstein, “Leveraging multimodal learning analytics to differentiate student learning strategies,” in *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*, Poughkeepsie New York: ACM, Mar. 2015, pp. 360–367, ISBN: 978-1-4503-3417-4.
- [38] R. Martinez-Maldonado *et al.*, “Supporting collaborative design activity in a multi-user digital design ecology,” *Computers in Human Behavior*, vol. 71, pp. 327–342, Jun. 2017.
- [39] R. Martinez-Maldonado, “A handheld classroom dashboard: Teachers’ perspectives on the use of real-time collaborative learning analytics,” *International Journal of Computer-Supported Collaborative Learning*, vol. 14, no. 3, pp. 383–411, Sep. 2019.
- [40] V. Echeverria, R. Martinez-Maldonado, and S. Buckingham Shum, “Towards Collaboration Translucence: Giving Meaning to Multimodal Group Data,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, Glasgow Scotland Uk: ACM, May 2019, pp. 1–16, ISBN: 978-1-4503-5970-2.
- [41] H. Vrzakova, M. J. Amon, A. Stewart, N. D. Duran, and S. K. D’Mello, “Focused or stuck together: Multimodal patterns reveal triads’ performance in collaborative problem solving,” in *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, Frankfurt Germany: ACM, Mar. 2020, pp. 295–304, ISBN: 978-1-4503-7712-6.
- [42] S. Oviatt and A. Cohen, “Written Activity, Representations and Fluency as Predictors of Domain Expertise in Mathematics,” in *Proceedings of the 16th International Conference on Multimodal Interaction*, ser. ICMI ’14, New York, NY, USA: Association for Computing Machinery, Nov. 2014, pp. 10–17, ISBN: 978-1-4503-2885-2.

- [43] A. Wise and M. Chiu, “Statistical Discourse Analysis of a Role-Based Online Discussion Forum: Patterns of Knowledge Construction,” *Hawaii International Conference on System Sciences*, vol. 0, pp. 3378–3386, Jan. 2012.
- [44] E. TSE, S. GREENBERG, C. SHEN, and C. FORLINES, “Multimodal multiplayer tabletop gaming,” *Comput. Entertain.*, vol. 5, no. 2, 12:12, Apr. 2007.
- [45] G. Reimers and A. Neovesky, “Student Focused Dashboards - An Analysis of Current Student Dashboards and What Students Really Want:” in *Proceedings of the 7th International Conference on Computer Supported Education*, Lisbon, Portugal: SCITEPRESS - Science, 2015, pp. 399–404, ISBN: 978-989-758-108-3.
- [46] S. S. J. Alhadad, “Visualizing Data to Support Judgement, Inference, and Decision Making in Learning Analytics: Insights from Cognitive Psychology and Visualization Science,” *Journal of Learning Analytics*, vol. 5, no. 2, pp. 60–85, Aug. 2018, Number: 2.
- [47] S. D. Teasley, “Student Facing Dashboards: One Size Fits All?” *Technology, Knowledge and Learning*, vol. 22, no. 3, pp. 377–384, Oct. 2017.
- [48] V. Echeverria, R. Martinez-Maldonado, S. B. Shum, K. Chiluiza, R. Granda, and C. Conati, “Exploratory versus explanatory visual learning analytics: Driving teachers’ attention through educational data storytelling,” *Journal of Learning Analytics*, vol. 5, no. 3, pp. 73–97, 2018, Publisher: Society for Learning Analytics Research.
- [49] R. Williams, *The Non-designer’s Design Book: Design and Typographic Principles for the Visual Novice*. Pearson Education, 2015, Google-Books-ID: pqHfBQAAQBAJ, ISBN: 978-0-13-396615-2.
- [50] J. Sweller, “Cognitive Load During Problem Solving: Effects on Learning,” *Cognitive Science*, vol. 12, no. 2, pp. 257–285, 1988, eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1207/s15516709>
- [51] R. Moreno and B. Park, “Cognitive Load Theory: Historical Development and Relation to Other Theories,” in *Cognitive Load Theory*, J. L. Plass, R. Moreno, and R. Brünken, Eds., 1st ed., Cambridge University Press, Apr. 2010, pp. 9–28, ISBN: 978-0-521-67758-5 978-0-521-86023-9 978-0-511-84474-4.
- [52] N. Cheng, W. Zhao, X. Xu, H. Liu, and J. Tao, “The influence of learning analytics dashboard information design on cognitive load and performance,” *Education and Information Technologies*, Apr. 2024.
- [53] J. L. Redifer, C. L. Bae, and Q. Zhao, “Self-efficacy and performance feedback: Impacts on cognitive load during creative thinking,” *Learning and Instruction*, vol. 71, p. 101 395, Feb. 2021.

- [54] S. Oviatt, “Human-centered design meets cognitive load theory: Designing interfaces that help people think,” in *Proceedings of the 14th ACM international conference on Multimedia*, Santa Barbara CA USA: ACM, Oct. 2006, pp. 871–880, ISBN: 978-1-59593-447-5.
- [55] E. Carbone, “Students Behaving Badly in Large Classes,” *New Directions for Teaching and Learning*, vol. 1999, no. 77, pp. 35–43, 1999, eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/tl.77>
- [56] P. Dillenbourg, “Design for classroom orchestration,” *Computers & Education*, vol. 69, pp. 485–492, Nov. 2013.
- [57] G. Gibbs and A. Jenkins, Eds., *Teaching large classes in Higher Education: how to maintain quality with reduced resources*, Repr. London: Kogan Page, 1994, ISBN: 978-0-7494-0600-4.
- [58] A. van Leeuwen, J. Janssen, G. Erkens, and M. Brekelmans, “Teacher regulation of cognitive activities during student collaboration: Effects of learning analytics,” *Computers & Education*, vol. 90, pp. 80–94, Dec. 2015.
- [59] P. Griffin and E. Care, Eds., *Assessment and Teaching of 21st Century Skills: Methods and Approach*. Dordrecht: Springer Netherlands, 2015, ISBN: 978-94-017-9394-0 978-94-017-9395-7.
- [60] E. Lai, K. DiCerbo, and P. Foltz, “Skills for Today: What We Know about Teaching and Assessing Collaboration,” Pearson, Tech. Rep., 2017, Publication Title: Pearson ERIC Number: ED576950.
- [61] C. Sun, V. J. Shute, A. Stewart, J. Yonehiro, N. Duran, and S. D’Mello, “Towards a generalized competency model of collaborative problem solving,” *Computers & Education*, vol. 143, p. 103 672, Jan. 2020.
- [62] L. M. Greenstein, *Assessing 21st Century Skills: A Guide to Evaluating Mastery and Authentic Learning*. Corwin Press, Jul. 2012, Google-Books-ID: uHWu_pPEPiUC, ISBN: 978-1-4522-8415-6.
- [63] J. Andrews-Todd, Y. Jiang, J. Steinberg, S. L. Pugh, and S. K. D’Mello, “Investigating collaborative problem solving skills and outcomes across computer-based tasks,” *Computers & Education*, vol. 207, p. 104 928, Dec. 2023.
- [64] X. Ochoa, V. Echeverria, G. Carrillo, V. Heredia, and K. Chiluiza, “Supporting Online Collaborative Work at Scale: A Mixed-Methods Study of a Learning Analytics Tool,” in *Proceedings of the Tenth ACM Conference on Learning @ Scale*, Copenhagen Denmark: ACM, Jul. 2023, pp. 237–247, ISBN: 979-8-4007-0025-5.
- [65] T. Breideband *et al.*, “The Community Builder (CoBi): Helping Students to Develop Better Small Group Collaborative Learning Skills,” in *Computer Supported Cooperative Work*

- and Social Computing*, Minneapolis MN USA: ACM, Oct. 2023, pp. 376–380, ISBN: 9798400701290.
- [66] A. Lewis, X. Ochoa, and R. Qamra, “Instructor-in-the-Loop Exploratory Analytics to Support Group Work,” in *LAK23: 13th International Learning Analytics and Knowledge Conference*, ser. LAK2023, New York, NY, USA: Association for Computing Machinery, Mar. 2023, pp. 284–292, ISBN: 978-1-4503-9865-7.
 - [67] Y. Chen, C. D. Andrews, C. E. Hmelo-Silver, and C. D’Angelo, “Coding schemes as lenses on collaborative learning,” *Information and Learning Sciences*, vol. 121, no. 1/2, pp. 1–18, Dec. 2019.
 - [68] J. Andrews-Todd and D. Kerr, “Application of Ontologies for Assessing Collaborative Problem Solving Skills,” *International Journal of Testing*, vol. 19, pp. 172–187, Apr. 2019.
 - [69] M. Boothe Jr, C. Yu, and X. Ochoa, *Bridging the Gap Between Theory and Tool: A Pragmatic Framework for Multimodal Collaboration Feedback* (LAK21), 11th ed. International Conference on Learning Analytics & Knowledge, Apr. 2021.
 - [70] H. Chai, T. Hu, and L. Wu, “Computer-based assessment of collaborative problem solving skills: A systematic review of empirical research,” *Educational Research Review*, vol. 43, p. 100591, May 2024.
 - [71] E. Gouli, A. Gogoulou, M. Grigoriadou, and M. Samarakou, “Supporting and Guiding Learners’ Collaboration through a Structured Adaptive Communication Tool,” Aug. 2023.
 - [72] N. M. M. Dowell, T. M. Nixon, and A. C. Graesser, “Group communication analysis: A computational linguistics approach for detecting sociocognitive roles in multiparty interactions,” *Behavior Research Methods*, vol. 51, no. 3, pp. 1007–1041, Jun. 2019.
 - [73] T. K. Landauer, D. S. McNamara, S. Dennis, and W. Kintsch, *Handbook of Latent Semantic Analysis*. Psychology Press, Feb. 2007, Google-Books-ID: JbzCzPvzpmQC, ISBN: 978-1-135-60328-1.
 - [74] Q. Zhou, W. Suraworachet, and M. Cukurova, “Harnessing Transparent Learning Analytics for Individualized Support through Auto-detection of Engagement in Face-to-Face Collaborative Learning,” *Collaborative Learning*, 2024.
 - [75] M. Bradford, I. Khebour, N. Blanchard, and N. Krishnaswamy, “Automatic Detection of Collaborative States in Small Groups Using Multimodal Features,” in *Artificial Intelligence in Education*, N. Wang, G. Rebolledo-Mendez, N. Matsuda, O. C. Santos, and V. Dimitrova, Eds., Cham: Springer Nature Switzerland, 2023, pp. 767–773, ISBN: 978-3-031-36272-9.

- [76] X. Li, L. Yan, L. Zhao, R. Martinez-Maldonado, and D. Gasevic, “CVPE: A Computer Vision Approach for Scalable and Privacy-Preserving Socio-spatial, Multimodal Learning Analytics,” in *LAK23: 13th International Learning Analytics and Knowledge Conference*, ser. LAK2023, New York, NY, USA: Association for Computing Machinery, Mar. 2023, pp. 175–185, ISBN: 978-1-4503-9865-7.
- [77] I. Culjak, D. Abram, T. Pribanic, H. Dzapo, and M. Cifrek, “A brief introduction to OpenCV,” in *2012 Proceedings of the 35th International Convention MIPRO*, May 2012, pp. 1725–1730.
- [78] Q. Zhou, W. Suraworachet, O. Celiktutan, and M. Cukurova, “What Does Shared Understanding in Students’ Face-to-Face Collaborative Learning Gaze Behaviours “Look Like”?” In *Artificial Intelligence in Education*, M. M. Rodrigo, N. Matsuda, A. I. Cristea, and V. Dimitrova, Eds., vol. 13355, Series Title: Lecture Notes in Computer Science, Cham: Springer International Publishing, 2022, pp. 588–593, ISBN: 978-3-031-11643-8 978-3-031-11644-5.
- [79] S. Strohkorb, I. Leite, N. Warren, and B. Scassellati, “Classification of Children’s Social Dominance in Group Interactions with Robots,” in *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, ser. ICMI ’15, New York, NY, USA: Association for Computing Machinery, Nov. 2015, pp. 227–234, ISBN: 978-1-4503-3912-4.
- [80] D. B. Jayagopi, H. Hung, C. Yeo, and D. Gatica-Perez, “Modeling Dominance in Group Conversations Using Nonverbal Activity Cues,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 3, pp. 501–513, Mar. 2009.
- [81] O. Aran and D. Gatica-Perez, “Fusing Audio-Visual Nonverbal Cues to Detect Dominant People in Group Conversations,” in *2010 20th International Conference on Pattern Recognition*, Place: Istanbul, Turkey, IEEE, Aug. 2010, pp. 3687–3690, ISBN: 978-1-4244-7542-1.
- [82] J. K. Burgoon, M. L. Johnson, and P. T. Koch, “The nature and measurement of interpersonal dominance,” *Communication Monographs*, vol. 65, no. 4, pp. 308–335, Dec. 1998.
- [83] J. Costa, M. F. Jung, M. Czerwinski, F. Guimbretière, T. Le, and T. Choudhury, “Regulating Feelings During Interpersonal Conflicts by Changing Voice Self-perception,” in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, Montreal QC Canada: ACM, Apr. 2018, pp. 1–13, ISBN: 978-1-4503-5620-6.
- [84] V. Vonikakis, Y. Yazici, V. D. Nguyen, and S. Winkler, “Group happiness assessment using geometric features and dataset balancing,” in *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, ser. ICMI ’16, New York, NY, USA: Association for Computing Machinery, Oct. 2016, pp. 479–486, ISBN: 978-1-4503-4556-9.

- [85] Wenxuan Mou, O. Celiktutan, and H. Gunes, “Group-level arousal and valence recognition in static images: Face, body and context,” in *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, Ljubljana: IEEE, May 2015, pp. 1–6, ISBN: 978-1-4799-6026-2.
- [86] G. Sharma, S. Ghosh, and A. Dhall, “Automatic Group Level Affect and Cohesion Prediction in Videos,” in *2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, Sep. 2019, pp. 161–167.
- [87] N. H. Frijda, *The Emotions*. Cambridge University Press, 1986, ISBN: 978-0-521-31600-2.
- [88] E. A. Clark *et al.*, “The Facial Action Coding System for Characterization of Human Affective Response to Consumer Product-Based Stimuli: A Systematic Review,” *Frontiers in Psychology*, vol. 11, p. 920, May 2020.
- [89] Y. R. Tausczik and J. W. Pennebaker, “The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods,” *Journal of Language and Social Psychology*, vol. 29, no. 1, pp. 24–54, Mar. 2010.
- [90] H. Hung and D. Gatica-Perez, “Estimating Cohesion in Small Groups Using Audio-Visual Nonverbal Behavior,” *IEEE Transactions on Multimedia*, vol. 12, no. 6, pp. 563–575, Oct. 2010.
- [91] A. Carron, W. Widmeyer, and L. Brawley, “The Development of an Instrument to Assess Cohesion in Sport Teams: The Group Environment Questionnaire,” *Journal of Sport Psychology*, vol. 7, no. 3, pp. 244–266, Sep. 1985.
- [92] D. Gatica-Perez, L. McCowan, D. Zhang, and S. Bengio, “Detecting group interest-level in meetings,” in *Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, ISSN: 2379-190X, vol. 1, Mar. 2005, I/489–I/492 Vol. 1.
- [93] J. Cassell, A. J. Gill, and P. A. Tepper, “Coordination in conversation and rapport,” in *Proceedings of the Workshop on Embodied Language Processing - EmbodiedNLP '07*, Prague, Czech Republic: Association for Computational Linguistics, 2007, pp. 41–50.
- [94] A. Street, “Anmol Madan, Ron Caneel, Alex “Sandy” Pentland,”
- [95] N. Campbell, “Multimodal Processing of Discourse Information; The Effect of Synchrony,” in *2008 Second International Symposium on Universal Communication*, Dec. 2008, pp. 12–15.
- [96] S. Shen, P. Slovak, and M. F. Jung, ““Stop. I See a Conflict Happening.”: A Robot Mediator for Young Children’s Interpersonal Conflict Resolution,” in *Proceedings of the 2018*

- ACM/IEEE International Conference on Human-Robot Interaction*, Chicago IL USA: ACM, Feb. 2018, pp. 69–77, ISBN: 978-1-4503-4953-6.
- [97] M. F. Jung, N. Martelaro, and P. J. Hinds, “Using Robots to Moderate Team Conflict: The Case of Repairing Violations,” in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, Portland Oregon USA: ACM, Mar. 2015, pp. 229–236, ISBN: 978-1-4503-2883-8.
 - [98] C. L. Sidner, C. Lee, C. D. Kidd, N. Lesh, and C. Rich, “Explorations in engagement for humans and robots,” *Artificial Intelligence*, vol. 166, no. 1-2, pp. 140–164, Aug. 2005.
 - [99] S. M. Anzalone, S. Boucenna, S. Ivaldi, and M. Chetouani, “Evaluating the Engagement with Social Robots,” *International Journal of Social Robotics*, vol. 7, no. 4, pp. 465–478, Aug. 2015.
 - [100] J. Fan, L. C. Mion, L. Beuscher, A. Ullal, P. A. Newhouse, and N. Sarkar, “SAR-Connect: A Socially Assistive Robotic System to Support Activity and Social Engagement of Older Adults,” *IEEE Transactions on Robotics*, vol. 38, no. 2, pp. 1250–1269, Apr. 2022.
 - [101] H. Javed, R. Burns, M. Jeon, A. M. Howard, and C. H. Park, “A Robotic Framework to Facilitate Sensory Experiences for Children with Autism Spectrum Disorder: A Preliminary Study,” *ACM Transactions on Human-Robot Interaction*, vol. 9, no. 1, pp. 1–26, Mar. 2020.
 - [102] K. K. Sturmer, P. Bishop, and S. M. Lenhart, “Developing Collaboration Skills in Team Undergraduate Research Experiences,” *PRIMUS*, vol. 27, no. 3, pp. 370–388, Mar. 2017, Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/10511970.2016.1188432>.
 - [103] M. Kleinsmann, F. Deken, A. Dong, and K. Lauche, “Development of design collaboration skills,” *Journal of Engineering Design*, vol. 23, no. 7, pp. 485–506, Jul. 2012, Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/09544828.2011.619499>.
 - [104] T. D. Safarini, “Developing students’ collaboration skills through project-based learning in statistics,” *Journal of Physics: Conference Series*, vol. 1265, no. 1, p. 012 011, Jul. 2019, Publisher: IOP Publishing.
 - [105] S. Nahar, Suhendri, Zailani, and Hardivizon, “Improving Students’ Collaboration Thinking Skill under the Implementation of the Quantum Teaching Model,” *International Journal of Instruction*, vol. 15, no. 3, pp. 451–464, Jul. 2022, Publisher: Gate Association for Teaching and Education ERIC Number: EJ1355589.
 - [106] E. Tohani and I. Aulia, “Effects of 21st Century Learning on the Development of Critical Thinking, Creativity, Communication, and Collaboration Skills,” *Journal of Nonformal Education*, vol. 8, no. 1, pp. 46–53, Feb. 2022, Number: 1.

- [107] E. Jaleniauskienė, R. Leščinskij, and P. Jucevičienė, “Development of collaboration in the English for specific purposes courses: Opportunities and challenges,” <http://espeap.junis.ni.ac.rs/index.php> 2019, Accepted: 2023-09-18T20:18:51Z Artwork Medium: tekstas / txt Interview Medium: tekstas / txt Publisher: University of Niš.
- [108] R. Gratton, “Collaboration in students’ learning: The student experience,” *Support for Learning*, vol. 34, no. 3, pp. 254–276, 2019, _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-9604.12261>.
- [109] J. D. Bransford and National Research Council, Eds., *How people learn: brain, mind, experience, and school*, Expanded ed., 9. print. Washington, DC: National Academy Press, 2004, ISBN: 978-0-309-07036-2.
- [110] Committee on How People Learn II: The Science and Practice of Learning, Board on Behavioral, Cognitive, and Sensory Sciences, Board on Science Education, Division of Behavioral and Social Sciences and Education, and National Academies of Sciences, Engineering, and Medicine, *How People Learn II: Learners, Contexts, and Cultures*. Washington, D.C.: National Academies Press, Sep. 2018, Pages: 24783, ISBN: 978-0-309-45964-8.
- [111] V. J. Shute, “Focus on Formative Feedback,” *ETS Research Report Series*, vol. 2007, no. 1, pp. i–47, 2007, _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/j.2333-8504.2007.tb02053.x>.
- [112] M. T. Chi, S. A. Siler, H. Jeong, T. Yamauchi, and R. G. Hausmann, “Learning from human tutoring,” *Cognitive Science*, vol. 25, no. 4, pp. 471–533, 2001, _eprint: <https://onlinelibrary.wiley.com/doi/>
- [113] M. T. H. Chi and R. Wylie, “The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes,” *Educational Psychologist*, vol. 49, no. 4, pp. 219–243, Oct. 2014, Publisher: Routledge _eprint: <https://doi.org/10.1080/00461520.2014.965823>.
- [114] I. Molenaar, P. Sleegers, and C. van Boxtel, “Metacognitive scaffolding during collaborative learning: A promising combination,” *Metacognition and Learning*, vol. 9, no. 3, pp. 309–332, Dec. 2014.
- [115] L. Soares, A. Mendes, A. Gomes, and G. Fortes, “Fostering regulation of learning processes among programming students using computational scaffolding,” *International Journal of Computer-Supported Collaborative Learning*, Mar. 2023.
- [116] A. M. Santoso, P. R. Primandiri, S. Zubaidah, and M. Amin, “Improving student collaboration and critical thinking skills through ASICC model learning,” *Journal of Physics: Conference Series*, vol. 1806, no. 1, p. 012 174, Mar. 2021, Publisher: IOP Publishing.
- [117] P. Pazos, N. Magpili, Z. Zhou, and L. J. Rodriguez, “Developing Critical Collaboration Skills in Engineering Students: Results from an Empirical Study,” ISSN: 2153-5965, Jun. 2016, ISBN: 978-0-692-68565-5.

- [118] D. Hunter, “Assessing collaborative learning,” *British Journal of Music Education*, vol. 23, no. 1, pp. 75–89, Mar. 2006, Publisher: Cambridge University Press.
- [119] OpenAI *et al.*, *GPT-4 Technical Report*, arXiv:2303.08774 [cs], Mar. 2024.
- [120] M. I. Jordan and T. M. Mitchell, “Machine learning: Trends, perspectives, and prospects,” *Science*, vol. 349, no. 6245, pp. 255–260, Jul. 2015, Publisher: American Association for the Advancement of Science.
- [121] S. Poria, A. Gelbukh, E. Cambria, A. Hussain, and G.-B. Huang, “EmoSenticSpace: A novel framework for affective common-sense reasoning,” *Knowledge-Based Systems*, vol. 69, pp. 108–123, Oct. 2014.
- [122] A. U. Khan, A. Mazaheri, N. d. V. Lobo, and M. Shah, *MMFT-BERT: Multimodal Fusion Transformer with BERT Encodings for Visual Question Answering*, arXiv:2010.14095 [cs], Oct. 2020.
- [123] C. Cui *et al.*, “Deep multimodal fusion of image and non-image data in disease diagnosis and prognosis: A review,” *Progress in Biomedical Engineering*, vol. 5, no. 2, p. 022 001, Apr. 2023, Publisher: IOP Publishing.
- [124] S. K. Roy, A. Deria, D. Hong, B. Rasti, A. Plaza, and J. Chanussot, “Multimodal Fusion Transformer for Remote Sensing Image Classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–20, 2023.
- [125] S. Xiang *et al.*, “Deep multimodal representation learning for generalizable person re-identification,” *Machine Learning*, vol. 113, no. 4, pp. 1921–1939, Apr. 2024.
- [126] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*, arXiv:1810.04805 [cs], May 2019.
- [127] S. L. Pugh, A. Rao, A. E. Stewart, and S. K. D’Mello, “Do Speech-Based Collaboration Analytics Generalize Across Task Contexts?” In *LAK22: 12th International Learning Analytics and Knowledge Conference*, ser. LAK22, New York, NY, USA: Association for Computing Machinery, Mar. 2022, pp. 208–218, ISBN: 978-1-4503-9573-1.
- [128] G. Chen, L. Donahue, and R. Klimoski, “Training Undergraduates to Work in Organizational Teams,” *Academy of Management Learning & Education*, vol. 3, pp. 27–40, Mar. 2004.
- [129] D. McKinney and L. F. Denton, “Developing collaborative skills early in the CS curriculum in a laboratory environment,” *SIGCSE Bull.*, vol. 38, no. 1, pp. 138–142, Mar. 2006.

- [130] A. C. Edmondson and Z. Lei, “Psychological Safety: The History, Renaissance, and Future of an Interpersonal Construct,” *Annual Review of Organizational Psychology and Organizational Behavior*, vol. 1, no. 1, pp. 23–43, Mar. 2014.
- [131] M. Worsley, “(Dis)engagement matters: Identifying efficacious learning practices with multimodal learning analytics,” in *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, ser. LAK ’18, New York, NY, USA: Association for Computing Machinery, Mar. 2018, pp. 365–369, ISBN: 978-1-4503-6400-3.
- [132] K. Furuichi and M. Worsley, “Using Physiological Responses To Capture Unique Idea Creation In Team Collaborations,” in *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing*, Jersey City NJ USA: ACM, Oct. 2018, pp. 369–372, ISBN: 978-1-4503-6018-0.
- [133] M. Worsley and P. Blikstein, “What’s an Expert? Using Learning Analytics to Identify Emergent Markers of Expertise through Automated Speech, Sentiment and Sketch Analysis,” in *EDM*, 2011, pp. 235–240.
- [134] A. A. Marsh, H. A. Effenbein, and N. Ambady, “Nonverbal “Accents”: Cultural Differences in Facial Expressions of Emotion,” *Psychological Science*, vol. 14, no. 4, pp. 373–376, Jul. 2003, Publisher: SAGE Publications Inc.
- [135] K. Bachour, F. Kaplan, and P. Dillenbourg, “An Interactive Table for Supporting Participation Balance in Face-to-Face Collaborative Learning,” *IEEE Transactions on Learning Technologies*, vol. 3, no. 3, pp. 203–213, Jul. 2010.
- [136] N. Bassiou *et al.*, “Privacy-Preserving Speech Analytics for Automatic Assessment of Student Collaboration,” in *Interspeech 2016*, ISCA, Sep. 2016, pp. 888–892.
- [137] M. Cukurova, Q. Zhou, D. Spikol, and L. Landolfi, “Modelling collaborative problem-solving competence with transparent learning analytics: Is video data enough?” In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, Frankfurt Germany: ACM, Mar. 2020, pp. 270–275, ISBN: 978-1-4503-7712-6.
- [138] S. Grover, M. Bienkowski, A. Tamrakar, B. Siddiquie, D. Salter, and A. Divakaran, “Multi-modal analytics to study collaborative problem solving in pair programming,” in *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, ser. LAK ’16, New York, NY, USA: Association for Computing Machinery, Apr. 2016, pp. 516–517, ISBN: 978-1-4503-4190-5.
- [139] J. Boaler, D. Wiliam, and M. Brown, “Students’ Experiences of Ability Grouping—disaffection, polarisation and the construction of failure,” *British Educational Research Journal*, vol. 26, no. 5, pp. 631–648, 2000, eprint: <https://bera-journals.onlinelibrary.wiley.com/doi/pdf/10.1080/71365158>

- [140] T. Hoel and W. Chen, “Privacy and data protection in learning analytics should be motivated by an educational maxim—towards a proposal,” *Research and Practice in Technology Enhanced Learning*, vol. 13, no. 1, p. 20, 2018.
- [141] J. W. Pellegrino, “Assessment as a positive influence on 21st century teaching and learning: A systems approach to progress,” *Psicología Educativa*, vol. 20, no. 2, pp. 65–77, Dec. 2014.
- [142] P. Christou, “Thematic Analysis through Artificial Intelligence (AI),” *The Qualitative Report*, Feb. 2024.
- [143] S. U. Noble, *Algorithms of oppression: how search engines reinforce racism*. New York: New York university press, 2018, ISBN: 978-1-4798-3724-3 978-1-4798-4994-9.
- [144] M. Naeem, T. Smith, and L. Thomas, “Thematic Analysis and Artificial Intelligence: A Step-by-Step Process for Using ChatGPT in Thematic Analysis,” *International Journal of Qualitative Methods*, vol. 24, p. 16 094 069 251 333 886, Apr. 2025.
- [145] A. Turobov, D. Coyle, and V. Harding, *Using ChatGPT for Thematic Analysis*, arXiv:2405.08828 [cs], May 2024.
- [146] S. Dawson, “The impact of institutional surveillance technologies on student behaviour,” *Surveillance & Society*, vol. 4, no. 1/2, Sep. 2002.
- [147] A. E. B. Stewart *et al.*, “CPSCoach: The Design and Implementation of Intelligent Collaborative Problem Solving Feedback,” in *Artificial Intelligence in Education*, N. Wang, G. Rebolledo-Mendez, N. Matsuda, O. C. Santos, and V. Dimitrova, Eds., vol. 13916, Series Title: Lecture Notes in Computer Science, Cham: Springer Nature Switzerland, 2023, pp. 695–700, ISBN: 978-3-031-36271-2 978-3-031-36272-9.
- [148] D. Danks and A. J. London, “Algorithmic Bias in Autonomous Systems,” in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, Melbourne, Australia: International Joint Conferences on Artificial Intelligence Organization, Aug. 2017, pp. 4691–4697, ISBN: 978-0-9992411-0-3.
- [149] S. Slade and P. Prinsloo, “Learning Analytics: Ethical Issues and Dilemmas,” *American Behavioral Scientist*, vol. 57, no. 10, pp. 1510–1529, Oct. 2013, Publisher: SAGE Publications Inc.
- [150] D. Tzimas and S. Demetriadis, “Ethical issues in learning analytics: A review of the field,” *Educational Technology Research and Development*, vol. 69, no. 2, pp. 1101–1133, 2021, Publisher: Springer Nature.
- [151] A. Albrechtslund, “Online social networking as participatory surveillance,” *First Monday*, 2008.

- [152] S. Beattie, C. Woodley, and K. Souter, “Creepy Analytics and Learner Data Rights,” *AS-CILITE Publications*, pp. 421–425, Nov. 2014.
- [153] W. J. von Eschenbach, “Transparency and the Black Box Problem: Why We Do Not Trust AI,” *Philosophy & Technology*, vol. 34, no. 4, pp. 1607–1622, Dec. 2021.
- [154] H. A. Giroux, “Selling Out Higher Education,” *Policy Futures in Education*, vol. 1, no. 1, pp. 179–200, Mar. 2003.
- [155] T. Reeves, “Can Educational Research Be Both Rigorous and Relevant?,” 2011.
- [156] N. Dawalatabad, M. Ravanelli, F. Grondin, J. Thienpondt, B. Desplanques, and H. Na, “ECAPA-TDNN Embeddings for Speaker Diarization,” in *Interspeech 2021*, arXiv:2104.01466 [eess], Aug. 2021, pp. 3560–3564.
- [157] T. Giannakopoulos, “pyAudioAnalysis: An Open-Source Python Library for Audio Signal Analysis,” *PLOS ONE*, vol. 10, no. 12, e0144610, Dec. 2015, Publisher: Public Library of Science.
- [158] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, *You Only Look Once: Unified, Real-Time Object Detection*, arXiv:1506.02640 [cs], May 2016.
- [159] D. E. King, “Dlib-ml: A Machine Learning Toolkit,”
- [160] S. Yang, L. Jiang, Z. Liu, and C. C. Loy, *VToonify: Controllable High-Resolution Portrait Video Style Transfer*, arXiv:2209.11224 [cs], Sep. 2022.
- [161] M. Biasutti and S. Frate, “Group metacognition in online collaborative learning: Validity and reliability of the group metacognition scale (GMS),” *Educational Technology Research and Development*, vol. 66, no. 6, pp. 1321–1338, Dec. 2018.

Appendices

Appendix A

GEMINI THEMATIC CODING PROMPT

Seven major themes of Collaboration are Climate, Communication, Compatibility, Conflict, Context, Contribution, and Constructive.

Climate relates to the emotional and affective aspects of the collaboration. It involves creating a respectful and welcoming environment where students feel safe to share ideas without fear of being made fun of. A poor climate is described as an 'overall toxic atmosphere'. This includes being respectful, feeling comfortable and welcome, and maintaining a positive tone.

Communication refers to the quantity and quality of information shared among group members. Key aspects include active listening, ensuring everyone is on the same page about goals and expectations, and being open to compromise. This includes setting goals and expectations, listening, talking, and sharing concerns.

Compatibility focuses on how well group members' working and interaction styles complement each other. It includes having 'compatible work styles' and the benefits of team-building activities to foster a sense of community. This includes working with friends, creative vision, and complementary skills.

Conflict covers students' approaches to handling challenging or contentious situations that arise during group work. Students reported strategies like confronting the responsible person, simply 'dealing with it,' or involving an instructor as a mediator. This includes seeking external validation, confronting other collaborations, and adapting as problems arise.

Context refers to the who, why, and where of the collaboration. It includes the setting (in or out of class), the nature of the task, and whether students feel comfortable in that specific environment. This consists of the group members, how members are selected, students' interest in the activity, and concerns about data use.

Contribution highlights what each individual brings to the group, including both verbal con-

tributions (like ideas) and the effort put toward a group project. Students noted issues with both over-contribution (one person dominating) and under-contribution. This includes the balance of work, tracking accountability, balance in the discussion, engagement with the problem, and collaborations.

Constructive relates to the overall goals of the collaboration and the team's progress toward achieving them. A good collaboration is seen as mutually beneficial, where every member gains insights they wouldn't have had on their own. This includes the final product, learning goals, the efficiency of the collaboration, and the collective benefit of collaboration.

List which of these themes is detected in the response in a comma-separated list. If no themes are detected, write None. If the response is empty, write N/A