UC Irvine

UC Irvine Electronic Theses and Dissertations

Title

Gender Gaps and Inclusivity in STEM: Perspectives From Online Courses and Collaborative Teams

Permalink

https://escholarship.org/uc/item/1f78h010

Author

Lin, Yiwen

Publication Date

2024

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA,

IRVINE

Gender Gaps and Inclusivity in STEM: Perspectives From Online Courses and Collaborative Teams

DISSERTATION

submitted in partial satisfaction of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in Education

by

Yiwen Lin

Dissertation Committee:

Assistant Professor Nia Nixon, Chair

Professor Di Xu

Professor Lindsey Richland

2024

© 2024 Yiwen Lin

TABLE OF CONTENTS

LIST OF FIGURES ..... **iv**

LIST OF TABLES ..... **v**

ACKNOWLEDGEMENTS ..... **vi**

VITA ..... **vii**

ABSTRACT OF THE DISSERTATION ..... **xii**

Introduction **1**

Theoretical Framework **7**

Study 1: STEM Pathways in a Global Online Course: Are Male and Female Learners Motivated the Same? **11**

Introduction 11
Current Study 15
Method 16
Results 22
Discussion 24

Study 2: Harnessing the Equalizing Power of Groups: Gender-Specific Effects of Group Experience on STEM Belonging **28**

Introduction 28
Current Study 33
Method 34
Results 39
Quantitative Outcomes *39*
Qualitative outcomes *41*
Discussion 46

Study 3: Unpacking Collaborative Problem-Solving Interactions: the Role of Gender Composition for STEM teams **50**

Introduction 50
Current Study 55
Method 62
Results 74
Individual level outcomes *74*
Team level outcomes *81*
Discussion 85

Contributions to the Field **89**

Implications 91
Future Studies 93

References **98**

Appendix A: Inventory of Psychological Measures - Sense of Belonging, STEM Identity, and Self-Efficacy **115**

Appendix B. Knowledge Test Items **118**

Appendix C. Full list of country of origin in Neuromatch **119**

Appendix D. Codebook for student feedback **120**

Appendix E. Sample CPS Transcript **121**

Appendix F. Exclusion criteria for text processing.....122

LIST OF FIGURES

		Page
Figure 1	Conceptualization of Multilevel Factors Influencing Gender Discrepancies in STEM Outcomes	11
Figure 2	Unconstrained structural model of the effect of belonging, field identity, and self-efficacy on persistence separately for women and men	24
Figure 3	The interaction environment of a team playing Physics Playground	67
Figure 4	A screenshot of the team trying to solve a problem	67
Figure 5	Gender differences in Single-gender and Mixed-gender teams	77-78
Figure 6	Gender differences in Female-majority and Male-majority teams	80-81
Figure 7	Participation Balance in Single-gender and Mixed-gender teams	83
Figure 8	Participation Balance in Single-gender and Mixed-gender teams	85

LIST OF TABLES

		Page
Table 1	Sample Bivariate Correlations, Means, and Standard Deviations	23
Table 2	Summary of regression models	40
Table 3	Relationship between Group Experience and Field Belonging, Moderated by Gender	42
Table 4	Overall view analysis plan corresponding to each research question	61-63
Table 5	Distribution of male and female students in CPS groups by gender composition	64
Table 6	Distribution of teams in the CPS task	65
Table 7	Descriptive statistics of participants by gender grouping	68-69
Table 8	Summary and definitions of outcome variables at individual and group level	72-73
Table 9	Means and standard deviations of dependent variables at individual level with different gender compositions	75-76
Table 10	Summary of results at individual level	81-82
Table 11	Means and standard deviations of dependent variables at team level with single and mixed gender compositions	83
Table 12	Means and standard deviations of dependent variables at team level of all compositions	85
Table 13	Summary of results at group level	86

ACKNOWLEDGEMENTS

I would like to express the deepest gratitude to my advisor, Nia Nixon, whose unwavering support, compassionate mentorship, and boundless enthusiasm for research and life have been indispensable throughout my journey over the past six years. She believed in my potential when I couldn't see it myself, and I am eternally grateful for the nurturing environment she cultivated, which allowed my intellectual and personal identity to flourish.

I would like to extend my appreciation to my committee members, their thoughtful feedback and fresh perspectives have significantly enhanced this dissertation. I would also like to thank my current labmates, former colleagues and mentors at the University of Michigan, and collaborators in the learning analytics community, for engaging me in critical thinking and improving the quality of my research.

To my friends and peers at UCI, I am deeply grateful for your love and support. I feel incredibly fortunate to have been part of such a vibrant and diverse community, where I formed some of my most profound and meaningful friendships. Our interactions over tea, food, routine workouts, and countless stimulating conversations have enriched my body and soul. You have been my anchor both within and outside the academic environment, and have been instrumental in shaping who I am today.

I would like to thank my community of friends outside of academia, for allowing me to explore intriguing questions about life choices and find passion in everyday life. They have turned the challenges of the pandemic into a meaningful and memorable chapter of my life.

I would also like to acknowledge the beautiful Southern California landscapes, weather, and culture. These elements provided an environment for resilience and much-needed escapes during the ups and downs of this PhD journey.

Finally, my parents, who have unconditionally supported me every step of the way. Their selfless dedication has been my stable foundation, giving me the courage to reach this point and the confidence to move forward. I could not have come this far without their love and support.

I thank the Association for Computing Machinery for permission to include Study1, which was published in the proceeding of the Eleventh ACM Conference on Learning @ Scale. I thank the National Institute of Health for partially supporting this dissertation work (Grant 5UC2NS128361-02).

VITA

Education

University of California, Irvine	School of Education PhD, Concentration: Teaching, Learning, and Educational Improvement	2019 – 2024 Irvine, CA
University of Michigan-Ann Arbor	Bachelor of Arts in Psychology Bachelor of Arts in International Studies	2015 – 2018 Ann Arbor, MI

Research Experience

Language and Learning Analytics Lab
UC Irvine
09/2019 - Present

- Research equity in STEM and computer-mediated communication, analyze text data to examine collaborative processes in online teams, and develop strategies to teach collaboration skills and expand diversity and inclusion in higher education.

Advisor: Nia Nixon

Teacher Discourse
UC Irvine
06/2022 – 02/2023

- Conducted literature review on existing research practices around teacher-student interactions, synthesized theories and analytical methods to model teacher discourse.

Advisor: Qiujie Li, Di Xu

STEM Learning Lab
UC Irvine
03/2020 – 06/2022

- Assisted planning and organizing co-design sessions, conducted qualitative analysis to gain insights on culturally relevant designs for STEM learning, and co-written learning science publications.

Advisor: Andres Bustamante, June Ahn

Educational Technology Collective
University of Michigan-Ann Arbor
09/2018 – 06/2019

- Conducted two studies on gender differences in teams and effects of gender composition on collaborative dynamics, supported research on leveraging crowdsourced effort from learners to make course improvements.

Advisor: Christopher Brooks, Nia Dowell

Emotion and Self-control Lab University of Michigan-Ann Arbor
2017 – 2018

- Analyzed Twitter data to explore emotional expression in public discourse and detect linguistic cues associated with stress and wellbeing.

Advisor: Ariana Orvell

Language and Literacy Lab University of Michigan-Ann Arbor
Summer 2016

- Carried out observations on parent-child interactions in informal learning environment (i.e. science museum), and conducted experiments with children on their language development.

Culture and Self-Insight Lab University of Michigan-Ann Arbor 2015 – 2017	• Helped construct survey instruments to study cultural variations in educational decisions, and cultural factors in psychological experiences.

Professional Experience

United Nations Foundation	Applied behavioral science frameworks to find cost effective solutions to the global challenges in clean energy conversion and market shift towards a more sustainable model. Conducted 2 user studies to improve the usability of two existing designs.	06/ 2023 – 09/2023
International Academic Peer Advisor, University of Michigan-Ann Arbor	Strategized curriculum plans for incoming international and transfer students, advised individual learning paths based on students' career goals.	08/ 2017

Peer-Reviewed Publications

Journal

1. Nixon, N., Lin, Y., & Snow, L. (2024). Catalyzing Equity in STEM Teams: Harnessing Generative AI for Inclusion and Diversity. *Policy Insights from the Behavioral and Brain Sciences*, 23727322231220356.

2. Lin, Y., & Nixon, N. (2024). Transitioning to Online Instructions and COVID-19 Response: A View from Mining Emergent College Students Discourse in Online Discussion Forum. International Journal of Artificial Intelligence in Education, 1-26.

3. Belgrave, A.B., Bermudez, V.N., Chew, P., Lin, Y., Ahn, J., & Bustamante, A.S. (2021) Community Engaged Design: A Systematic Approach for Designing Culturally-Situated STEM Enrichment Activities. *Journal of Applied Developmental Psychology*

4. Dowell, N. M., Lin, Y., Godfrey, A., & Brooks, C. (2020). Exploring the Relationship between Emergent Sociocognitive Roles, Collaborative Problem-Solving Skills, and Outcomes: A Group Communication Analysis. *Journal of Learning Analytics*, *7*(1), 38-57.

5. (under review) Lin, Y., Cavazos, J. G., Nixon, N., Popov, V., (2024) Navigating the Melting Pot: Understanding Communication Dynamics in Culturally Diverse STEM Teams. Journal of Computer-Assisted Learning.

Conference Proceedings (Peer-Reviewed)

6. Lin, Y., Nixon, N. (2024 July). STEM Pathways in a Global Online Course: Are Male and Female Learners Motivated the Same? In Proceedings of the Eleventh ACM Conference on

Learning @ Scale (L@S '24). Association for Computing Machinery, New York, NY, USA, 243–249. https://doi.org/10.1145/3657604.3662045

7. Samadi, M. A., Jaquay, S., Lin, Y., Tajik, E., Park, S., & Nixon, N. (2024, March). Minds and Machines Unite: Deciphering Social and Cognitive Dynamics in Collaborative Problem Solving with AI. In *Proceedings of the 14th Learning Analytics and Knowledge Conference* (pp. 885-891).

8. Chopra, H., Lin, Y., Samadi, M. A., Cavazos, J. G., Yu, R., Jaquay, S., & Nixon, N. (2023). Semantic Topic Chains for Modeling Temporality of Themes in Online Student Discussion Forums. *International Educational Data Mining Society*.

9. Samadi M. A., Cavazos, J. G., Chopra H., Lin, Y., Dowell, N. (2022, July) Exploring Cultural Diversity and Collaborative Team Communication through a Dynamical Systems Lens. *International Educational Data Mining Society.*

10. Lin, Y., Dowell, N., & Godfrey, A. (2021, April). Skills Matter: Modeling the relationship between decision-making processes and collaborative problem-solving skills during Hidden Profile Tasks. In *LAK21: 11th International Learning Analytics and Knowledge Conference* (pp. 428-437).

11. Chew, P., Lin, Y., Bermudez, V. N., Bustamante, A. S., & Ahn, J. (2021). Understanding goals, pedagogical frameworks, and relationships in community-based participatory design. In *Proceedings of the 15th International Conference of the Learning Sciences-ICLS 2021.*. International Society of the Learning Sciences.

12. Singh, A., Brooks, C., Lin, Y., & Li, W. (2021, June). What's In It for the Learners? Evidence from a Randomized Field Experiment on Learnersourcing Questions in a MOOC. In *Proceedings of the Eighth ACM Conference on Learning@ Scale* (pp. 221-233) [Best Paper Award].

13. Lin, Y., Yu, R., & Dowell, N. (2020, July). LIWCS the same, not the same: Gendered linguistic signals of performance and experience in online STEM courses. In *International Conference on Artificial Intelligence in Education* (pp. 333-345). Springer, Cham.

14. Lin, Y., & Dowell, N. M. (2020, April). Does size matter? modeling student interactions in online collaborative discussions. Paper in the proceeding of the *Annual meeting of American Education Research Association*, San Francisco, CA.

15. Dowell, N., Lin, Y., Godfrey, A., & Brooks, C. (2019, June). Promoting inclusivity through time-dynamic discourse analysis in digitally-mediated collaborative learning. In *International Conference on Artificial Intelligence in Education* (pp. 207-219). Springer, Cham. [Best Paper Award]

16. Lin, Y., Dowell, N., Godfrey, A., Choi, H., & Brooks, C. (2019, March). Modeling gender dynamics in intra and interpersonal interactions during online collaborative learning. In *Proceedings of the 9th International Conference on learning analytics & knowledge* (pp. 431-435).

17. Lin, Y., & Dowell, N. M. (2019, July). Does gender really matter?: Exploring differences in emerging discourse styles during digitally mediated collaborative interactions. *Society for Text and Discourse*, New York, NY.

18. Lin, Y. & Dowell, NM. (2020, April). Does Size Matter? Modeling Student Interactions in Online Collaborative Discussions. Paper presented in SIG-Online Teaching and Learning at the annual meeting of the American Educational Research Association, San Francisco, CA.

Posters

1. Chopra, H., Lin, Y., Samadi, M. A., Cavazos, J. G., Yu, R., Jaquay, S., & Nixon, N. (2022, July). Modeling Student Discourse in Online Discussion Forums Using Semantic Similarity Based Topic Chains. In *Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners' and Doctoral Consortium: 23rd International Conference, AIED 2022, Durham, UK*, July 27–31, 2022, Proceedings, Part II (pp. 453-457). Cham: Springer International Publishing.

2. Lin, Y., Dowel, N. (April 2020) Gender Differences in Online Collaborative Discourse Overtime. *AERA Satellite Conference on Educational Data Science*, Stanford, CA. Editors: Daniel McFarland, Ben Domingue, and Zachary Pardos.

3. Lin, Y., Dowel, N. (October, 2023). STEMing from Teams: Effects of Small Group Interactions on Women's Sense of Belonging and Persistence in Online STEM course. Invited poster presentation at LEAD Graduate School & Research Network, Herrenberg, Germany

Workshop Papers

4. Lin, Y., Collaborative Analytics Workshop, the 13th International Conference on Learning Analytics & Knowledge. (March, 2023)

Invited Talks

1. Invited Panelist, *Empowering Learners in the Age of AI (ELAI) virtual conference* (October, 2023), Advancing the Science of Collaboration through AI. Panel Chair: Nia Nixon.
2. Invited Panelist, at *the 32rd Annual Meeting of the Society for Text and Discourse* (July 2022), on Promoting Research on Diversity, Equity, and Inclusion. Moderator: Joe Magliano.
3. Guest Lecturer, Topic: Measuring 21st-century skills (March, 2024). In 21st-century Literacy.
4. Guest Lecturer, Topic: Visualization in Learning Analytics (March, 2024). In Learning Analytics.
5. Guest Lecturer, Topic: Multimodal Analytics (May, 2023). In Learning Analytics, UCI.
6. Guest Lecturer, Topic: Unpacking Collaboration (May, 2023). In 21st-century Literacy.
7. Guest Lecturer, Topic: Collaborative Learning (Nov, 2021). In 21st-century Literacy, UCI.

Teaching

Teaching Assistant, University of California, Irvine

Advanced Fieldwork After School	Spring 2024
Theories of Development and Learning Applied to Education	Winter 2024
(Masters) Instructional Design & Education Technology for the Elementary Classroom	Summer 2022
(Masters) Children's Sense-Making in Science	Summer 2022
Theories of Development and Learning Applied to Education	Spring 2022
American History and Education Policy	Winter 2022
Introduction to Education	Fall 2022
21st-century Literacy	Spring 2021
21st-century Literacy	Fall 2020

Academic Service

Reviewer

International Journal of Artificial Intelligence in Education	2024
British Journal of Educational Technology	2024
International Learning Analytics and Knowledge Conference	2020 - 2024
Journal of Artificial Intelligence in Education	2023
Journal of Learning Analytics	2021

Student Volunteer

International Learning Analytics and Knowledge Conference	2020 - 2021
International Conference on Artificial Intelligence in Education	2022
University of California School of Education PhD Recruitment	2022, 2023

Student Mentor
3 Master students
8 Undergraduate students

Professional Training

Inclusive Teaching Training 2024
Certificate: Promoting Inclusion & Accessibility (DEI-A) Culture in the Classroom, Inclusive Teaching Institute (ITI)

Public Speaking Program 2023
Certificate: Activate to Captivate

Learning Analytics Summer Institute 2019
Society for Learning Analytics Research

ABSTRACT OF THE DISSERTATION

Gender Gaps and Inclusivity in STEM: Perspectives From Online Courses and Collaborative Teams

by

Yiwen Lin

Doctor of Philosophy in Education

University of California, Irvine, 2024

Professor Nia Nixon, Chair

Gender disparities have been a longstanding issue in STEM. This dissertation investigates gender disparities within the realm of global online STEM courses and STEM learning teams. The objective of this research is to elucidate how personal factors such as a sense of belonging, self-efficacy, and identity, along with situational factors like team experience and composition, may be associated with various outcomes. The insights garnered from this study aim to illuminate future interventions for fostering equity and inclusion in STEM learning environments.

This dissertation comprises three studies, each addressing a different facet of the complex issue of gender imbalance in STEM. Study 1 offers an updated perspective on gender gaps in innovative online learning formats, uncovering the distinct relationships between individual factors such as a sense of belonging, STEM identity, and self-efficacy in relation to STEM persistence intentions. Study 2 investigates the role of group experience on learners' sense of belonging and the moderating role of gender. Additional factors related to the course and peers

were examined to highlight areas for designing for learning experience. Study 3 evaluates gender disparities in collaborative STEM teams, examining how computer-supported communication patterns differ across genders, and vary based on team compositions. Taken together, these three studies offer significant implications for future research, practices in online teaching and learning, and endeavors toward a more equitable and inclusive STEM environment.

Introduction

Diversity and equity in Science, Technology, Engineering and Mathematics (STEM) related disciplines are driving forces of innovation and scientific discoveries in society (Smith-Doerr, Alegria, & Sacco, 2017). Historically, women have been excluded from these disciplines and face discrimination in professional and educational environments (Sonnert & Holton, 1995). Despite significant progress in diversifying STEM workforce, gender disparities are still evident in key fields such as computer science, physics and engineering (Smith, 2011; Hammond et al., 2020; Fry et al., 2021). A recent report from the National Science Foundation suggests that women make up about a third of the STEM workforce by 2022 (Beede et al., 2011; Burke et al., 2022). Given that engineering and technology industries offer some of the highest paying jobs, persistent gender imbalance can exacerbate gender inequality in income and financial opportunities. In the education sector, while women earn more than half of the social and behavioral science degrees at all educational levels, women earn disproportionately fewer shares of degrees in physics, mathematics, and computer science (National Science Foundation, Science & Engineering Statistics, 2023).

Over the past few decades, researchers have sought to identify the person and situational factors that account for these gender disparities. Literature in psychological science identifies a few key factors contributing to women's interest and experience in STEM. A male-dominant environment often subjects women to heightened stereotype threats, adversely affecting their engagement and learning outcomes (Shapiro and Williams, 2012; Casad et al., 2019; Van Veelen et al, 2019). Conversely, increasing the number of women in these settings can reduce perceived threats, bolster women's STEM identity and encourage more active female participation. Murphy et al., (2007) found that women in STEM are particularly responsive to situational cues,

suggesting that their sense of belonging and willingness to engage are significantly influenced by the numerical representation of their gender in the environment. Research on gender equity in STEM posits the importance of gender representation in countering the salient disparity in students' self concept, increasing women's participation in STEM, and influencing women's educational choices (Dasgupta and Stout, 2014; Sax et al., 2015; Lavy and Sand, 2015). Similarly, Good et al. (2012) determined that greater female representation in STEM is associated with women feeling accepted and valued members of the academic community. Fischer (2017) further explores how classroom composition affects men and women differently, revealing that while a high-achieving classroom dynamic positively impacts men, it can be detrimental to women's self-efficacy. These findings emphasize the need to scrutinize how men and women react to situational factors such as representation and dynamics in STEM learning environments, to foster a more inclusive and supportive atmosphere for all.

In light of this, collaborative learning emerges as a critical strategy for expanding women's participation in STEM. Collaborative problem-solving promotes complex cognitive reasoning and critical thinking, fostering positive attitudes and developing essential self-concepts towards science (Johnson et al., 1994; Stout and Tamer, 2016). Within STEM teams, Cheryan et al (2009) highlighted that team environments and behaviors signaling STEM stereotypes can significantly influence women's interest and sense of belonging, often deterring them from these fields. By cultivating a collaborative and inclusive learning environment and ensuring equitable participation of women and other minoritized groups, we can unlock the full potential of diversity, thereby driving innovation and productivity (Kozlowski,et al., 2022; Love et al., 2023). Consequently, it is imperative to understand the patterns of interaction and communication behaviors within CPS teams and how these dynamics vary across different team environments.

This understanding can inform effective strategies to foster equitable and inclusive participation in STEM.

Dissertation Overview

This dissertation aims to deepen our understanding of the factors contributing to gender gaps in STEM learning. The goal of this research is to elucidate these gaps in the context of online courses and virtual learning teams, thereby informing future interventions to foster equity and inclusion in STEM learning environments. The learning science community emphasizes the importance of considering the context in which learning occurs. In line with that, this research strives to evaluate the generalizability and applicability of previous findings in the new digital learning era. It seeks to balance scalable solutions with the need for contextualized support (Shibani et al., 2019). Through this work, I aim to contribute toward the endeavors of creating more equitable and inclusive STEM education.

The dissertation comprises three distinct studies, each tackling different facets of this complex issue. Study 1 focuses on individual factors such as belonging, STEM identity, and self-efficacy, investigating how these factors relate differently to persistence outcomes for men and women. Studies 2 and 3 shift focus to situational and contextual factors, exploring the impact of group-learning experiences and gender compositions within teams on students' sense of belonging and communication behaviors. The three studies tackle the issues of gender gaps from different angles. Study 1 takes a holistic approach by surveying students' persistent intentions in STEM fields and identifying key drivers for both genders. Building on Study 1, Studies 2 and 3 delve into a micro social environment to examine gender gaps in the context of learning groups,

which are essential components in online STEM education for fostering social connectedness and community.

The interdisciplinary nature of this research necessitates drawing upon theories and methodologies from learning sciences, psychology, and information science to achieve a comprehensive understanding of existing practices and insights. By updating existing trends and introducing novel perspectives for potential interventions, this study aims to provoke critical reflection on the distinct challenges and opportunities in advancing STEM equity in today's educational landscape. In Study 1 and 2, I engage with a new cohort of globally distributed and culturally diverse learners characterized by their motivation, self-regulation, and adeptness in upskilling. Departing from traditional learner archetypes focused on sequential course completion and degrees, these learners seek additional skills and knowledge essential for gaining a competitive edge in the fast evolving labor market. Our investigation delves into the factors influencing educational decisions among this group of learners. Study 3 explores an emerging pedagogical approach: synchronous collaborative problem-solving in virtual teams. With online courses becoming more mainstream following the pandemic, educators are finding ways to mitigate the social isolation often associated with asynchronous courses. As virtual conferencing tools and online learning platforms continue to develop with more advanced capabilities, the integration of synchronous collaborative components into future online STEM courses is only inevitable. Thus, it is crucial to investigate gender dynamics within these contexts to foster an inclusive learning environment effectively. Taken together, this dissertation study assesses gender disparities within these specific contexts to gain accurate insights and guiding recommendations tailored to these scenarios for future learning.

In **Study 1**, I explored whether gender disparities previously identified in research could be observed within a global online learning community. Additionally, I examined how personal factors, such as belonging, STEM identity, self-efficacy influence intentions to persist in STEM fields to different extents for both men and women. The findings offer a complex picture: while some align with established literature, others challenge prevailing assumptions about gender differences. Persistent disparities were evident in areas such as perceived belonging and academic readiness, although no significant gender differences were found in persistence intentions. Interestingly, women reported higher levels of STEM identity and self-efficacy compared to men. The study also reaffirmed existing research showing strong correlations between a sense of belonging, STEM identity, and STEM persistence, highlighting the importance of interventions aimed at fostering a strong sense of belonging among learners. Contrary to previous studies (e.g., Lewis et al., 2019), which suggested a stronger effect for women, I found a more positive association between belonging and persistence intent among men. Self-efficacy, however, emerged as a significant predictor of persistence intentions exclusively for women. These nuanced findings provide valuable insights into the diverse factors influencing learners' interest in pursuing STEM fields, offering guidance for effectively tailored interventions in large-scale learning environments to promote equity.

In **Study 2**, building on the critical role of a sense of belonging for STEM persistence established in Study 1, I expanded the focus from personal factors to examine the influence of a situational factor—group experience—and its relationship with belonging. The results indicate a positive relationship between group experience and learners' sense of belonging in their field, even after controlling for variables such as prior knowledge, gender, first language, education level, and country of residence. Furthermore, the findings suggest that gender moderates this

relationship, with women's sense of belonging being more significantly influenced by their group experience compared to their male peers. This underscores the necessity for targeted interventions that consider the varying impacts on different learner subgroups. I also aimed to identify specific elements within the course design that contribute to learners' experiences, by qualitatively analyzing students' open-ended feedback in an online STEM program. I found emergent themes in students. The findings imply that future research should delve into the specific group dynamics that foster a sense of belonging and explore aspects of inclusivity to enhance the learning environment for all participants.

In **Study 3**, with a similar objective to Study 2 in examining situational factors affecting equity in STEM, I focused on STEM teams to investigate the connection between group compositions and gendered interactions. The premise for this study assumes that the male dominance observed in the broader underrepresentation in STEM environments carries over into micro social environments, such that the gender ratio in small teams may have explicit or implicit effects on individual interactions. Combined with the cultural narrative that women tend to exhibit a more collaborative and supportive communication style, I anticipated observing varying degrees of gender differences in collaborative problem-solving (CPS) interactions across gender compositions. To test this hypothesis, I analyzed triads of university students engaged in a synchronous physics CPS activity. I defined two levels of gender compositions: the first level distinguishes between mixed-gender and single-gender teams, while the second level further categorizes groups into all-male teams, male-majority teams (2 males, 1 female), female-majority teams (2 females, 1 male), and all-female teams. I conducted various sets of ANCOVA to compare communication behaviors among different gender groups, characterized by factors such as turns, word count, proportional turns, and sentiment at the individual level, while

controlling for demographic variables. Additionally, I examined group-level communication frequencies, balance, and sentiment using separate sets of ANOVA.

Theoretical Framework

This dissertation is anchored in the epistemological foundations of Constructivism and Bronfenbrenner's Ecological Model. Constructivist learning theories posit that learners construct their own understanding and knowledge of the world through personal experiences and reflective processes. This framework offers a basis for understanding why psychological experiences are important to women's engagement in STEM, influencing their motivation to either persist in or withdraw from these fields. The core principles of constructivist theories—such as active learning, collaborative knowledge construction, and social interaction—align with the group-based approach used in the current research context, particularly for Study 2 and 3. In this approach, smaller groups are formed to facilitate students' understanding through discussions, the sharing of perspectives, and collaborative learning activities.

Vygotsky's Socio-cultural Theory provides a robust framework for evaluating how online group experiences influence learners' sense of belonging, particularly within the diverse socio-cultural contexts represented in our dataset. Vygotsky's emphasis on the importance of social interactions and cultural influences in learning helps us understand the complexities that arise from the varied cultural and disciplinary backgrounds of our globally represented sample of online learners. Moreover, this theory is instrumental in addressing how preconceived gender stereotypes in STEM, embedded within different societies, impact students' sense of belonging. By leveraging Vygotsky's ideas, we can gain deeper insights into the gender differences in STEM learning and how these differences are shaped by the socio-cultural environments of online learners worldwide.

Additionally, Vygotsky's theoretical perspectives lend themselves nicely to understanding the impact of micro social environments, such as online groups, on psychological learning outcomes like STEM belonging. According to Vygotsky, learning is inherently a social process, and the interactions within these micro social environments significantly shape cognitive and affective outcomes of individual participants in the group. An individual's experience in a group can be influenced by social dynamics, varying levels of engagement, and group cohesion and communication. These factors further contribute to the formation of a learning environment that can either support or hinder a student's sense of belonging in STEM. Using this framework, we can dissect the nuances in how group experience, influenced by social structures and cultural dynamics of online groups, uniquely affects male and female students' sense of belonging in STEM fields. For instance, men and women might experience different expectations and participation patterns within these groups, leading to disparities in how supported and included they feel. Men and women often demonstrate different communication styles and interpret communications from their peers differently, resulting in varied impacts on their sense of belonging. Through this lens, we hypothesize for this study that in a short-term, intensive online learning program, women may be more sensitive to group dynamics, meaning their sense of belonging in STEM could be more easily influenced by temporary group experiences. In contrast, men's sense of belonging in the field might be more stable and less affected by short-term group interactions.

Bronfenbrenner's ecological model is also highly relevant because it emphasizes the multiple layers of environment that influence an individual's development. In the context of this research, the microsystem is particularly relevant as it involves direct interactions among teammates. The hypothesis is that these team-based experiences can significantly influence

students' sense of belonging and motivation, which are crucial for their academic success in STEM fields. Bronfenbrenner's model helps frame this inquiry by providing a structured way to analyze how immediate interactions within teams (microsystem) may interact with and impact students' belonging within their field of study (mesosystem). For instance, positive experiences within teams can enhance a student's engagement and persistence in STEM by fostering a supportive and collaborative environment, while negative experiences might have the opposite effect. Figure 1 depicts the conceptualization of multilevel factors influencing gender discrepancies in STEM Outcomes. Although not directly assessed in this dissertation, the influence of the macrosystem—encompassing broader cultural and societal narratives and stereotypes—can significantly impact women's persistence and sense of belonging in STEM fields. If a society has a strong discriminatory structure that conveys the biased message that scientific careers are for men, this exclusionary environments can be reinforced by organizations and institutions with skewed gender ratios. In addition, the dynamics and social climate in small groups can significantly impact learners' perceptions of acceptance and belonging as they interact with their peers. Positive interactions within groups may serve as a buffer against the negative effects of stereotype threats for women and marginalized students.

Figure 1. *Conceptualization of Multilevel Factors Influencing Gender Discrepancies in STEM Outcomes*



Study 1: STEM Pathways in a Global Online Course: Are Male and Female Learners Motivated the Same?

Introduction

The "Leaky Pipeline" problem has long been a significant challenge in efforts to increase diversity in STEM. This metaphor describes the phenomenon where students drop out of the STEM learning trajectory at various stages: students who initially express interest in science careers may reconsider when applying to colleges and universities, opting for other areas of study. Historically, this issue has disproportionately affected women and other minoritized students compared to men (Alper, 1993; Clark Blickenstaff, 2005). Women graduating with a college STEM degree are more likely to shift their careers towards non-STEM fields (Botella et al., 2019). As a result, we have disproportionately fewer women scientists, engineers, and other technical positions, which is a colosso loss of talents for the scientific communities. Increasing women's interest in STEM and retaining them within these fields has become a critical mission in addressing the gender imbalance in STEM disciplines.

Supporting women in their pursuit of STEM careers requires a comprehensive understanding of the factors influencing their educational choices, including what deters and what attracts them. Existing literature emphasizes the essential role of psychological mechanisms—such as a sense of belonging, identity, and self-efficacy beliefs—in shaping learners' aspirations to persist along the STEM trajectory and remain resilient in challenging STEM environments (Rosenthal et al., 2013; Casad et al., 2019). Women continue to encounter barriers to academic and career advancement, including subtle messages and stereotypes that signal low competency, as well as fewer opportunities for quality support and mentorship. These barriers reinforce the stereotype that women do not belong in or align with the science identity.

Understanding how these factors influence women's STEM learning pathways is crucial for developing effective interventions that support their retention and success in STEM fields.

Sense of belonging

Good et al. (2012) identify a **sense of belonging** as one's feelings of membership or acceptance in a particular domain. Sense of belonging is a driving factor to motivation and career interest (Smith et al., 2013), academic achievement (Walton et al., 2015), performance (Edwards et al., 2022) and intent to pursue STEM fields (Rainey et al., 2013). Female students were consistently found to have a lower sense of belonging in STEM, which compromises their STEM attitudes relative to career interests (Xu and Lastrapes, 2021). Importantly, sense of belonging is found to have a stronger link to women's persistent intent and actual decision to persist (Lewis et al., 2017). This suggests supporting women's sense of belonging can be an effective way to level the gender differences in persistence.

STEM Identity

Educational research identifies "science identity" as a pivotal solution, as it creates opportunities for girls and young women to envision themselves as physicists, engineers, or computer scientists and enabling them to be actively engaged in such fields (Fines-Neuschild et al., 2024). Females traditionally report greater difficulty feeling identified with their field of study, or seeing themselves as scientists in male-dominated environments. Interventions that target fostering women's STEM identity, such as assigning female mentors and increasing representation in groups (Niler et al., 2020), have been found to be effective in promoting deeper levels of interest, more positive experiences, and persistence intents in STEM (Hernandez et al.,

2017; Dowel et al., 2021). Developing a STEM identity compatible for women also predicts higher belonging and self-efficacy (Clark et al., 2016).

Self-Efficacy

Self-efficacy refers to beliefs about one's ability to plan for and execute steps necessary for future success (Bandura, 1986). Women are likely to report lower self-efficacy than men (Stout et al., 2011), given the stereotypical cultural narrative that women are less competent (Miller et al., 2015). The gender differences in self-efficacy further explain women's lower STEM persistence.

A New Age of Learning

To effectively enhance the retention of female talent in STEM, it is essential to understand how promising measures work within a new context. Prior to implementing any interventions or initiatives, it is essential to thoroughly comprehend the student population and the specific environment. Recent shifts in the educational landscape necessitate a renewed perspective on gender equality and equity in STEM.

One significant change has been the rise in prominence of online credential programs offered by prestigious universities. These programs are appealing due to their lower costs, flexibility, and scalability. Studies indicate that women's enrollment in online STEM programs has steadily increased over the years (Joyner & Duncan, 2024). Additionally, significant events such as the global COVID-19 pandemic have accelerated the adoption of various online learning formats, further popularizing online education. Simultaneously, there has been a growing interest in interdisciplinary STEM education. Scholarly communities have initiated programs

specializing in interdisciplinary areas, recognizing that integrated STEM education can generate powerful knowledge to address global issues and drive significant advancements (Kurup et al., 2021; Williams, & Roth, 2021). Research also suggests that women are increasingly engaged in interdisciplinary research and collaborative work (Van Rijnsoever & Hessels, 2011). These interdisciplinary STEM programs have the potential to mitigate structural power imbalances between the more gender-divided "hard" natural sciences and "softer" disciplines, creating a more welcoming environment for women (Light et al., 2022). Combining scientific knowledge with computational approaches and emphasize real-world applications and versatile skills for employment, can be also particularly appealing to women and other marginalized groups.

Recent studies have demonstrated the potential for online platforms to broaden equitable participation and increase diversity in STEM (Brooks et al., 2020). Additionally, online programs have been shown effective in producing similar student learning outcomes as traditional in-person instruction at substantially lower costs (Chirikov et al., 2020). However, little research has been conducted on short-term, intensive, domain-specific online learning programs, such as boot camps or summer programs, which are increasingly attracting self-motivated learners seeking to strengthen their knowledge and skills. As the future of work demands the ability to reskill and upskill in response to rapidly changing industries, understanding these programs becomes increasingly imperative (Li, 2022). Most existing evidence regarding gendered psychological processes and their differentiated effects on persistence is based on studies of traditional STEM programs at accredited institutions, often within the U.S. or specific countries. While gender gaps in persistence have also been identified in informal online STEM courses such as MOOCs (Kizilcec & Saltarelli, 2019), it remains unclear whether these findings generalize to more specialized STEM programs. Additionally,

gender differences may manifest in nuanced ways, even when overall learning outcomes appear equal (Lin et al., 2019). Therefore, it is essential to examine learner characteristics and gender differences in these new learning contexts. This understanding is crucial for refining strategies and interventions that promote more equitable environments and support the retention of diverse talent in STEM.

Current Study

The current study focuses on neuroscience learners participating in an emergent and rapidly growing online program. This program offers a high-quality, short-term, intensive computational neuroscience curriculum to learners worldwide. More details about this context will be discussed in the next section. In this study, I aim to explore gender differences in key psychological factors, specifically sense of belonging, STEM identity, and self-efficacy. Additionally, I will examine two other variables: learners' persistence intent and academic readiness, the latter of which assesses learners' understanding of basic concepts in computing and math to gauge their preparedness for the program. The research questions guiding this study are as follows:

RQ1. Do we find gender disparities in sense of belonging, STEM identity, self-efficacy, and the outcome variables of STEM persistence intent and academic readiness, in line with established patterns in the literature?

RQ2. Do belonging, STEM identity, self-efficacy demonstrate distinct associations with persistence intent and academic readiness among men and women?

Method

Study Context

Data was collected from students enrolled in an online Computational Neuroscience in Neuromatch Academy during the summer of 2021. The program was launched in 2020 to offer accessible, high-quality science education for students around the world who want to acquire computational skills to advance their careers. The pedagogical content was contextualized to teach students how to apply computational skills to answer scientific questions in neuroscience. The course was conducted in a flipped classroom format over the course of three weeks, with an intensive curriculum covering topics of machine learning, dynamical systems, and stochastic processes and their application in neuroscience in each week (van Viegen et al., 2021). The course was developed in response to the growing need for neuroscientists to possess coding skills. These computational abilities are often implicitly expected to handle the increasing size and complexity of neuroscience data. However, coding courses are rarely required or offered in traditional neuroscience degrees (Juavinett, 2022). It is crucial that coding content be made relatable to neuroscience and the student experience. Additionally, given the well-documented differences in the perception of computational science based on students' gender and racial identities, a discipline-specific introduction can better motivate students to overcome their fears and engage fully.

Students from any background at any level were welcome to apply, but they were expected to have some familiarity with Python programming (given that a majority of the assignments and course activities relied on Python notebooks), core math concepts, and exposure to neuroscience. Students were invited to review the materials before applying to ensure the

course is a good fit for them. All material are open-source and students can find previous course content and schedules available online. Applicants who did not have any related experience in those areas were rejected at the initial application process. While the program is relatively low cost compared to similar training programs on the market, there was an enrollment fee that was adjusted to the cost of living standards in different countries. Scholarships were offered to students who displayed financial needs to ensure a high level of inclusivity and universal accessibility.

Participants

Over 1600 students enrolled in the program. We acquired students' consent to participate in research and share their data. Students who were accepted and enrolled in the course were invited to complete an online survey prior to the start of the program. All responses were anonymized for secondary data analysis. Survey participation was entirely voluntary, no incentive was given out and students could skip items they did not wish to answer. We excluded a few invalid responses that exceeded the scale of the item, and omitted students who had missing responses on all survey items. Given our motivation in understanding STEM persistence interest, we restricted our sample to undergraduate and graduate students who are preparing for a STEM career. However, we intend to conduct future analysis that features unconventional learners in the dataset (i.e. learners who are further along in their STEM career), such as industry professionals, and learners who already hold academic employment such as postdoctoral scholars or assistant professors. Eventually, a total number of 1124 responses were included for further analysis.

Within the current sample, 45% of the students were female (N = 515). 65% of the students were currently pursuing a graduate degree (N = 734), and the remainder 35% were undergraduate students (N = 390). The sample represents an internationally diverse and geographically dispersed demographic. A total of 88 countries of origin were represented in the sample, the top five most represented countries were China (N = 228), India (N = 159), the United States (N = 96), Iran (N = 56), and Germany (N = 51). The full list of countries represented can be found in Appendix. The majority of students reside outside of the U.S. (N = 967), and 13.9% of the enrolled students reported that they were living in the U.S. at the time of enrollment. The racial and ethnic makeup of the sample was 42% Asian, 27.6% Caucasian, 8.4% of Middle Eastern or North African origin, 8.1% of Hispanic, Latinx, or Spanish origin, 7.6% self-described mixed race, 4% Black or African American, and the rest declined to answer.

Measures

Full details of the instruments used to measure the following variables are provided in Appendix A.

Field Belonging

Field belonging was measured using a scale from (Lewis et al., 2019), originally adapted from (Cheryan et al, 2009) for assessing an individual's sense of belonging in the field of study. The survey consists of five items: "I feel like I really belong in my field of study," "I feel accepted by other students in my field of study, " "I feel accepted by the instructors in my field of study," "I regret choosing my field of study (reverse scored)," "I feel like I have a lot in common with other students in my field of study." Each item used a scale ranging from 1 = strongly disagree to 7 = strongly agree. For the reverse coded items, we converted the score to

make them consistent with the rest of the scale. For the sample in this study, Cronbach's value was 0.75, indicating a moderately reliable measure of STEM sense of belonging.

STEM Identity

STEM identity was measured using an instrument from (Lewis et al., 2019), which was originally adapted from (Schmader, 2002). The survey consists of four items that ask students to reflect on how they identify with their field of study. The specific items include: "Being a student in my field of study is an important part of my self-image," "Being a student in my field of study is unimportant to my sense of what kind of person I am," "Being a student in my field of study has very little to do with how I feel about myself," "Being a student in my field of study is an important reflection of who I am." Each item used a scale ranging from 1 = strongly disagree to 7 = strongly agree. For the reverse coded items, we converted the score to make them consistent with the rest of the scale. For the sample in this study, Cronbach's value was 0.77, indicating a moderately reliable measure of STEM identity.

Self-Efficacy

Students' self-efficacy was assessed with a ten item survey adapted from (Farr et al., 2011). Students rate how good are they at each statement, based on a 100 point scale, where 100 is extremely good at. Sample items include: "how good are you at finishing your homework assignments by deadlines?", "how good are you at studying even when there are other things to do?", and "how good are you at finding strategies to succeed in class?" For the sample in this study, Cronbach's value was 0.85, indicating a reliable measure of self-efficacy.

STEM Persistence Intent

To measure STEM persistence intent, we adapted a scale from (Schmader et al., 2004). The following items were used to assess students' interests in persisting in STEM: "How likely is it that you will pursue graduate study related to your major?" "How likely is it that your eventual career after graduation will directly pertain to mathematics or science?" "How often do you think about changing your major?""How likely is it that you will change your major? " Students rated these prompts based on a scale ranging from 1 = Not all likely to 7 = Very Likely. For the two reverse coded items, we converted the score to make them consistent with the rest of the scale. For the sample in this study, Cronbach's value was 0.68, indicating a moderately reliable measure of self-efficacy.

Academic Readiness

Academic readiness was assessed with a knowledge test consisting of 12 questions which was distributed after students finished the questionnaire. The test covered math concepts (i.e. algebra, calculus) and basic computational modeling concepts(i.e. Bayes rules, dimensionality reduction) that were necessary for the course. Students receive one point for a correct answer or zero point for an incorrect answer. We calculated a test score by accumulating the correct answers and dividing that by the total number of questions. We then converted the raw score into percentages relative to a 100% scale for the analysis.

Analytical Approach

We used a multivariate analysis of variance (MANCOVA) to test whether there are gender differences across a sense of belonging in STEM, STEM identity, self-efficacy, academic readiness, and STEM persistent interest. MANCOVA is appropriate for determining if there are significant differences among groups (i.e., gender) for multiple dependent variables, while

considering the interrelationships among the dependent variables. To examine the hypothesized links between belonging, STEM identity, and persistence after controlling for self-efficacy, we conducted a structural invariance test for men and women. Path models were fit to the data using a robust maximum likelihood (MLR) approach. MLR performs well across different sample sizes and accounts for a degree of non-normality present in the outcome variables. Model estimation was implemented using the R lavaan package (Rosseel, 2012). Model-fit was evaluated using indices included $\chi 2$ statistic, comparative fit index (CFI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR). A path model is considered a good fit to the data when CFI is close to or greater than 0.90, RMSEA is close to or less than 0.08, and SRMR is close to or less than 0.08 (Kline, 2005).

First, we used structural equation modeling to establish the baseline latent variable measurement models. After establishing an initial measurement model, we conducted invariance testing to compare the hypothesized measurement model of belonging and STEM identity, and the covariate of self-efficacy between using increasingly more restrictive models. We constructed a baseline model where all parameters were freely estimated for women and men for model comparisons; subsequently, a constrained model where all factor loadings were set to be equal for women and men was tested to establish metric invariance; The difference in $\chi 2$ statistics and CFI were computed between each step to determine whether or not the more constrained model resulted in a significantly worse fit.

Results

Correlations among the study variables are shown in Table 1. Results from MANCOVA suggested that the overall difference in these variables within the gender group was statistically significant $F(5, 1118) = 8.464$, $p < .01$. We did not observe differences between males and females in students' STEM persistent interest, hypothesis 1a was not supported. We did find evidence for hypothesis 2a, consistent with previous literature, female students reported a lower sense of belonging in STEM ($p < .01$) compared to men. For hypothesis 1c, we found opposite to what we expected, women reported significantly higher STEM identification ($p < .01$) compared to male students. Hypothesis 1d was partially supported, as female students were found performing worse on knowledge tests ($p < .01$), but reported significantly higher self-efficacy ($p = .01$) compared to men.

Table1

Sample Bivariate Correlations, Means, and Standard Deviations.

Variables	1. Field belonging	2. Field identity	3. Self-efficacy	4. STEM persistence	5. Knowledge score
1. Field belonging	-	.28***	.39***	.42***	-.12***
2. Field identity	.23***	-	.13***	.20***	.01
3. Self-efficacy	.40***	.12***	-	.24***	-.10**
4. STEM persistence	.46***	.25***	.16***	-	.01
5. Knowledge score	.01	.06	-.12***	.05	-
Total sample (n= 1124)	5.51 (.87)	5.16(1.21)	7.84(.99)	5.85(.99)	.39(.19)
Male (n= 609)	5.56(.87)	5.03(1.2)	7.73(1.01)	5.83(1.01)	.42(.21)
Female (n= 515)	5.461(.87)	5.32(1.13)	7.92(.97)	5.88(.97)	.37(.18)

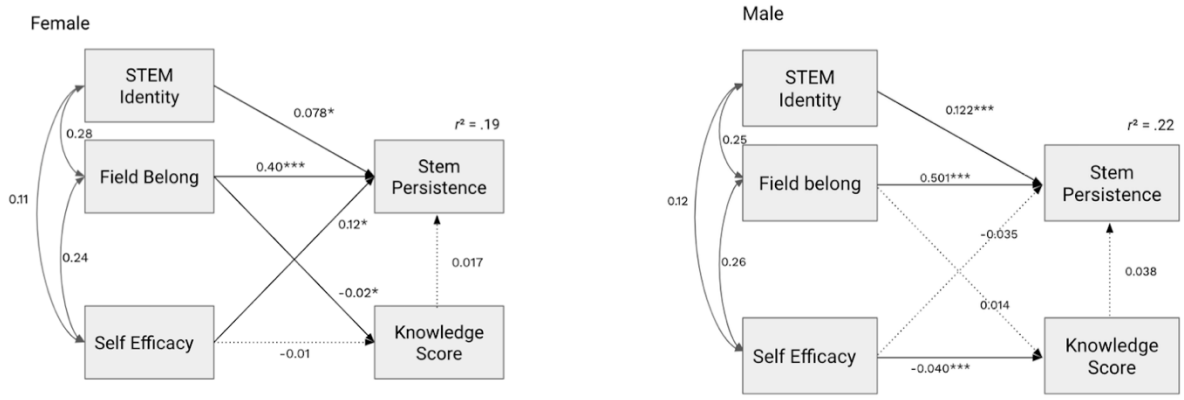
Note. Bivariate correlations for female students are above the diagonal of the matrix and the correlations for male students are below the diagonal. Standard deviations are in parentheses following means. Self-efficacy is represented as transformation scores with 0 representing the minimum possible score and 10 representing the maximum possible score. Exam scores are reported as mean % correct.
*$p < .05$, **$p < .01$, ***$p < .001$.

Next, we tested our hypothesis on the links between sense of belonging, STEM identity, self-efficacy and intention to persist as well as academic readiness measured by knowledge score

with path analysis. Our structural model showed excellent fit to the data, $\chi 2(2)= .08$, $p < .01$, RMSEA= .037, 90% CI [.000, .078], CFI = .99, SRMR = .015. Figure 1 presents the standardized path values along with individual paths estimated from the unconstrained model fitted for men and women separately. The model showed that sense of belonging is the strongest predictor of intention to persist for both women and men. Contrary to our expectation, this effect was stronger for men than for women. STEM identity was also a significant predictor for persistence, and similarly it was stronger for men than women. As hypothesis 2a and 2b posit that belonging and identity would have a stronger positive effect for women, the results suggest the contrary and thus only partially supported these hypotheses. We found that self-efficacy was indeed a significant predictor of intent to persist, although for women only and not for men. Lastly, results were opposite to hypothesis 2d since we found that self-efficacy was not predictive of female students' performance on knowledge test, and rather it is negatively linked to male students' performance.

Figure 2

Unconstrained structural model of the effect of belonging, field identity, and self-efficacy on persistence separately for women and men.



Note. The coefficients are standardized parameter estimates. Solid black lines indicate significant paths, and dashed grey lines indicate non-significant paths. For simplicity, the individual

indicators of the latent variables and the measurement residuals are not depicted.
*$p < .05$, **$p < .01$, ***$p < .001$.

Discussion

The present study delves into the realm of gender disparities within STEM by examining the psychological experiences of participants in a brief yet intensive online computational neuroscience program. This research contributes empirical evidence to the expanding body of literature dedicated to understanding gender inequalities in STEM fields. By focusing on previously understudied actors and contexts, our aim is to provide fresh insights into this critical issue. We spotlight learners engaged in an interdisciplinary neuroscience program, characterized by its global diversity. Notably, our investigation extends beyond the confines of traditional accredited higher education institutions, a departure from the predominantly location-bound context of previous research on this topic.

Our findings provide a nuanced perspective on gender differences in belonging and persistence, aligning only in part with prior research. While we confirmed the presence of a gender gap in the sense of belonging, intriguingly, our study revealed no significant gender disparities in STEM persistence intent. This observation may be attributed to the notable rise in female representation within the field of neuroscience, as evidenced by a recent report (Ramos et al., 2017). Over the past three decades, the number of women in neuroscience has steadily grown, with female graduates accounting for more than 50% of undergraduate neuroscience program enrollment by 2015, a percentage likely even higher in 2021, the year of our study. This increase in female representation may have positively influenced the motivation of women pursuing degrees within this domain. However, it's important to consider that participants in our

program were likely inherently self-motivated, driven by a desire to enhance their skills and knowledge, which could explain the observed gender parity in persistence intent.

We found that women reported significantly lower sense of belonging compared to men. However, women also reported that they felt more identified with the field. This contrast may suggest men and women may have different perceptions internally regarding their self-image, and outwardly regarding their environment. Specifically, women's higher sense of identification may come from being driven by certain causes to study the subject. While STEM identity and belonging are strongly associated, females may experience lower acceptance by the field than men influenced by academic achievement and validation from instructors and peers. Another potential explanation is that in an interdisciplinary field, it might be inherently difficult to discern which field students felt belonged to. For instance, a student may feel comfortable with neuroscience but more concerned about being a computationally skilled person. In line with previous literature, our results also highlight a sense of belonging as a strong predictor of learners' STEM persistence intent, except, interestingly, the effect is stronger for men. This suggests that creating an inclusive environment that cultivates a sense of belonging for all students can foster overall retention of students in the field. Many previous interventions target specifically at elevating women's belonging with a goal to lower their dropout rates. However, the future narrative might shift towards how we can create environments and online learning communities that enhance a collective sense of belonging.

We found that self-efficacy significantly contributes to women's resilience and persistence. Having a stronger belief in their ability to learn effectively may have compensated for the weaker links between persistence and belonging and STEM identity for women. This might explain why even though women reported lower belonging, there were no evident

differences in STEM persistence. Noticeably, men's self-efficacy was negatively related to their performance on knowledge test. This indicates a potential mismatch between beliefs about their competency for learning and their actual conceptual knowledge amongst male learners. Again, given the interdisciplinary nature, it is possible that male students who felt confident in their learning abilities derived from validations in their existing disciplinary background, in fact were underprepared for math and computational concepts which were tested on the knowledge test. While the link between self-efficacy and test performance was not evident for women, this might suggest women had a more realistic belief regarding their ability and their actual academic readiness. We did find that women were less prepared and less familiar with key concepts than men. However, students' performance on the test was not associated with persistence.

There are several limitations to this study. First, we did not capture learners' existing disciplinary backgrounds. While we assume the majority of learners came from a neuroscience background, as the program was advertised heavily to neuroscience communities, it is possible that some students are trained in adjacent disciplines (e.g. biology, psychology) and enroll in the program with the goal to pivot to neuroscience. Students could also be trained in computer science, with an interest in applying skills to neuroscience specifically. This made it difficult to pinpoint whether the learner's report on field belonging and identity was more relative to the field of science or computing, which have different gender representations. Second, we did not further distinguish undergraduate and graduate students within the sample. It is possible that the gender gaps might look different at different degree levels. Our next step would be further dissecting these separately for undergraduate and graduate students. Third, we have relatively small R-squared value for both models. We can attribute 19% of the variability in the women's STEM persistence to belonging, STEM identity and efficacy, and only 22% of men's persistence

intent variable can be explained by the independent variables included in the model. This implies that other factors that contributed to students' interest in pursuing STEM were not fully accounted for. For instance, one of those factors could be cultural variation. Since learners in our study were internationally dispersed, their own social, cultural, and academic environment may influence their perception of their identity and belonging. For instance, Mau et al. (2020) found that there is a greater gender difference in STEM learning experience and self-efficacy of students from the collectivist culture than students from the individualist culture. Another study suggest that non-native English speakers may experience more barriers to social-belonging and achieving academic success in STEM (LaCosse et al., 2020). Future studies would be needed to further investigate the effects of culture, language, and their roles on belonging and identity within a demographically diverse STEM learning community.

In moving towards a more equitable future, applications of AI can facilitate the process and provide opportunities for interventions. This learning community, for instance, leverages algorithmic sorting to group students in peer learning groups. It is therefore possible to take into account of these psychological factors to build a more complex model that optimize student experience, beyond interventions focusing on surface level features such as demographics. Future studies could examine whether grouping female learners with lower belonging and self-efficacy with peers who are higher on those could influence persistence intent. Additionally, ongoing assessments and evaluations are needed to uncover more nuanced ways learners experience different learning programs. We contend that interventions in STEM ought not to be deficiency or disparity oriented, strengthening what students are "less of". Rather, we should be more invested in thinking about how we might provide contextualized support that prioritizes improvement in both learning outcomes and experience, for learners of diverse backgrounds.

Study 2: Harnessing the Equalizing Power of Groups: Gender-Specific Effects of Group Experience on STEM Belonging

Introduction

Group work is a prevalent pedagogical strategy in higher education institutions. Employers expect college students to engage in group learning in an academic environment to become better prepared to work in teams in professional settings (Archer & Davison, 2008; Lowden et al., 2011). Group learning experiences help develop essential soft skills, such as communication, collaboration, and problem-solving, which are crucial for success in STEM careers (Schweingruber et al., 2012). In online courses, group work is commonly implemented to increase learner engagement. Given that students in remote settings are more likely to experience feelings of isolation, group-based learning is an ideal tool to help foster a sense of community and facilitate peer interactions, thereby enhancing social presence and social belonging. This is particularly crucial for STEM students, who often face unique challenges related to the rigorous nature of their disciplines.

Collaborative learning teams can serve as a crucial platform for fostering diversity and advancing equity within STEM disciplines (Nixon et al., 2024). This is particularly important towards the effort of promoting equity across STEM fields. In the United States, the disparity in STEM retention remains pronounced among students from underrepresented groups (i.e., women and underrepresented racial minorities), with approximately three-fourths of minority students leaving STEM disciplines at the undergraduate level (Wilson et al., 2015). At a global level, the participation of women in STEM still lags behind that of men, and despite performing better on tests, women who study STEM are less likely to enter into STEM careers and exit these careers

earlier than male peers (Hammond et al., 2023). Across different societies, the gender gap is even more significant in countries with biases due to gender stereotypes or cultural norms that influence female behavior and the family environment of girls and women (Ortiz-Martínez et al., 2023). To reach into the untapped potential of women in STEM, a host of educational interventions have been designed and implemented to model best practices to increase female participation, belonging, and retention in STEM disciplines (Dasgupta, 2014; Wang & Degol, 2017; Cheryan et al., 2017). Amongst these practices, most notable being paired mentoring (Estrada et al., 2012), affirmation of scientific identity (Stout et al., 2011), and building close peer relationships and network (Dennehy & Dasgupta, 2017). These strategies have shown promises in creating inclusive and supportive environments that empower women to succeed in STEM fields. Furthermore, they have potential for successful integration into online learning courses that utilize group-based pedagogies.

Group-based learning can serve as an important venue for expanding equitable practices in STEM. Psychologically safe environments can be nurtured through group interactions, enhancing the willingness of minority students to engage and contribute, leading to better learning outcomes (Edmondson, 1999; Sachmpazidi et al., 2021). Group learning experiences have been shown to foster a sense of community, increase student motivation, and improve comprehension of complex concepts (Johnson, Johnson, & Smith, 2007). These benefits are particularly significant for traditionally underrepresented STEM students, who may face additional barriers in their educational journeys. Research indicates that positive learning experiences in group settings can be especially beneficial for these students by providing supportive networks and promoting a sense of belonging (Chang et al., 2008). It is crucial to

expand our understanding on how learning experience in groups is linked to a sense of belonging for women in STEM.

Sense of belonging

Research has shown that a sense of belonging significantly impacts students' academic success and persistence in STEM fields. Strayhorn (2011) found that a strong sense of belonging is associated with higher academic performance, retention, and persistence, particularly for marginalized groups. In Lin & Nixon (2024), we demonstrated the strong link between belonging and STEM persistence, emphasizing the necessity to further examine what might hinder or foster learners' belonging. A study by Walton and Cohen (2011) demonstrated that students' perceptions and interpretations of their learning experiences are especially critical to the academic performance and well-being of minority students. For female students who encounter stereotype threats in STEM fields, their sense of belonging could be sensitive to their learning experiences.

Learning experience

Intricately connected to learners' sense of belonging in STEM fields, learning experiences play a crucial role in learners' motivation to engage in online courses. The literature emerging out of the learning sciences has increasingly focused on understanding learning experiences, recognizing their role, particularly in STEM disciplines, where a positive experience might be more pivotal to underrepresented students' success. This reflects a broader shift towards optimizing online learning environments to meet the diverse needs of students beyond performance-oriented outcomes (Sanger, 2020; Johnson, 2015; Kumi-Yeboah et al., 2020). Learning experiences have been studied both as an outcome and as a contributing factor to

learning gains. For instance, Means, Bakia, and Murphy (2014) highlight that the quality of the learning experience significantly impacts student outcomes. Similarly, Song et al (2004) discuss the pedagogical challenges and opportunities inherent in online learning environments, suggesting that the design of these experiences is pivotal to educational success. The field reaches consensus that well-structured and engaging online learning experiences are essential for fostering meaningful learning gains (Puzziferro, 2008; Gunawardena & McIsaac, 2004).

Sun and Chen (2016) present a comprehensive review of effective practices in online education, elucidating how positive learning experiences can enhance student engagement and achievement. By contrast, negative experiences can demotivate students and deter them from pursuing STEM career trajectories. For example, Maltese and Tai (2011) found that negative experiences in high school science classes were associated with a decreased likelihood of students earning a STEM degree. Their research highlights the importance of positive and supportive educational environments in fostering STEM persistence. Similarly, Goff et al. (2019) demonstrated that informal science and math experiences significantly influence undergraduate students' STEM identities and career intentions. Their study emphasizes the role of enriching and affirming experiences in building students' confidence and interest in STEM fields. On the other hand, studies have shown that negative learning experiences can have detrimental effects on students' motivation and career aspirations in STEM. In this context, for example, Hughes and Pace (2003) found that poor learning experiences and lack of support can lead to decreased motivation and higher dropout rates among STEM students. Similarly, Rainey et al. (2018) highlighted that when women and students of color experience a "chilly" environment where they feel unwelcome or unsupported, their motivation and engagement can suffer, leading to lower persistence rates.

Learning experiences within more immediate and specific environments, such as teams, may also have a significant impact on students. Murphy et al. (2007) found that women in STEM are more responsive to situational cues, suggesting that their sense of belonging and desire to engage are more significantly influenced by the features of a given setting. Therefore, it is reasonable to hypothesize that women's sense of belonging may be more sensitive to certain signals within their learning groups. Within team-based learning research, Dasgupta et al (2015) have found that female peers in small groups within STEM classes motivate female students' participation and career aspirations in engineering. This research suggests that the experience of being a minority in small groups can have a negative impact on female students' sense of belonging, confidence, performance, and satisfaction. The representation of gender in the learning environment can signal who naturally belongs in STEM and is likely to succeed, and who might be perceived as a misfit (Dasgupta et al., 2015). Recently, a study by Whitcomb et al. (2023) highlights that perceived recognition and effective peer interactions are crucial for fostering a positive learning environment, which in turn bolsters students' sense of belonging and academic performance in physics. Mickelson et al.(2020) reveals that a supportive peer network and inclusive classroom climate are vital for retaining female students in STEM majors, as these factors directly influence their sense of belonging and identity as STEM professionals .

These findings collectively suggest that fostering a collaborative and inclusive environment in STEM learning teams can improve female students' sense of belonging and retention in STEM fields. This motivated the current this study to investigate the direct correlation between students' learning experiences in groups and their sense of belonging in an intensive online STEM course.

Current Study

Implementing successful group-based learning at scale in online courses entails significant costs and complexities, particularly when dealing with geographically dispersed learners who vary in educational backgrounds, experience with the subject, cultural contexts, and linguistic capabilities. Students may be motivated by different factors, and their experiences can be shaped by their expectations and interactions with others. Given that online courses are prone to social isolation, facilitating social connectedness and cultivating a sense of belonging requires a thorough understanding of how students are impacted by groups, and what affects students' learning experience.

This study aims to examine the relationship between group experience and learners' sense of belonging in the context of an online computational neuroscience course. Additionally, I investigated whether gender moderates this relationship. I then conducted a qualitative analysis to identify key factors affecting learners' experiences and their suggestions for future improvements. These findings have the potential to inform strategies for enhancing sense of belonging in diverse learning teams in STEM environments. The research questions for this study are framed as follows:

RQ1. What is the relationship between group learning experience and students' sense of belonging?

RQ2. How does the relationship between group learning experience and students' sense of belonging differ between male and female students?

RQ3. What factors, as identified through qualitative exploration, contribute to students' group learning experience?

Based on previous research, I hypothesize that group experience will be positively associated with students' sense of belonging (H1). I further hypothesize that the relationships between group experience and sense of belonging will be stronger for female students (H2). This is in line with previous research that suggests underrepresented students benefit more from active learning.

Method

Study Context

Data was collected from students enrolled in a Computational Neuroscience course in an online community called Neuromatch Academy during the summer of 2021. The basic program details have previously stated in Study 1. I hereby provide additional information only relevant to this study.

Neuromatch implemented a unique approach to foster a sense of social connection. Specifically, to foster a sense of social connection and increase students' active engagement, the program implemented group-based learning within the larger community. These smaller groups are referred to as "pods". This is an innovative aspect of the program, which is to create an interactive online classroom environment at scale with students from geographically dispersed locations. This pedagogical decision was driven by the prevalent phenomenon of disengagement across online courses and Massive Open Online Courses (MOOCs), and the lack of attention from instructors further lowering students' engagement and hindering their learning outcomes.

Each group consisted of about 10-15 students, and each group was assigned teaching assistants to facilitate tutorials and group discussions. The group assignment process was assisted by using a matching algorithm that prioritized students' location, respective time zones, and language preferences. This is because the online tutorial and group discussion was conducted synchronously on Zoom.

Students first watch the pre-recorded lesson videos (all teaching materials are openly accessible online where the schedule for each week and general daily structure is available here). After watching the lesson videos, students participate in a synchronous group discussion for 15 minutes, followed by two tutorials with their teaching assistants, and round out with a 10-minute group discussion at the end. Students are generally expected to have a video camera on and interact with their peers in group discussions. Students have the option to participate in additional research projects collaborating with peers. The program does not have any formal assessments and students were not graded. Participants receive a certificate at the end that verifies their completion of the program.

Participants

Participants reported their demographic of gender, career status, first language, race, country of origin and beyond in the survey distributed at the beginning of the course. Within the current sample, 45% of the students were female (N = 515). 65% of the students were currently pursuing a graduate degree (N = 734), and the remainder 35% were undergraduate students (N = 390). The sample represents an internationally diverse and geographically dispersed demographic. A total of 90 countries of origin were represented in the sample, the top five most represented countries were China (N = 228), India (N = 159), the United States (N = 96), Iran (N

= 56), and Germany (N = 51). The majority of students reside outside of the U.S. (N = 967), and 13.9% of the enrolled students reported that they were living in the U.S. at the time of enrollment. The racial and ethnic makeup of the sample was 42% Asian, 27.6% Caucasian, 8.4% of Middle Eastern or North African origin, 8.1% of Hispanic, Latinx, or Spanish origin, 7.6% self-described mixed race, 4% Black or African American, and the rest declined to answer. In accordance with Institutional Review Boards requirements, all our participants have agreed to participate in the research study and have their data collected.

Survey Collection

All responses were anonymized for secondary data analysis. Survey participation was voluntary, no incentive was given out and students could skip items they did not wish to answer. We excluded a few invalid responses that exceeded the scale of the item, and omitted students who had missing responses on all survey items. The dataset includes participants across different levels of expertise, including those who hold academic employment such as post-doctoral scholars or assistant professors, and who already have a STEM degree and work as industry professionals. Given our interest around STEM belonging in higher education, we restricted our sample to undergraduate and graduate students who are still preparing for a STEM career. The response rate dropped in the post-survey as learners experienced fatigue, resulting in $N = 444$ responses in the final analysis.

Field Belonging

Field Sense of Belonging was measured using a scale from Lewis et al., (2019), originally adapted from Cheryan et al., (2009) for assessing an individual's sense of belonging in the field of study. The survey consists of five items: "I feel like I really belong in my field of

study," "I feel accepted by other students in my field of study. " Full survey inventory is available in Appendix. Each item used a scale ranging from 1 = strongly disagree to 7 = strongly agree. For the reverse-coded items, we converted the score to make them consistent with the rest of the scale. For the sample in this study, (Cronbach's alpha = 0.75), indicating a moderately reliable measure of STEM sense of belonging.

Group experience

Students were asked to rate their learning experience in groups at the end of the program. The question asked students to rate "How was your experience interacting with your pod-peers?" with options ranging from "A-Excellent", "B-Very good", "C-Satisfactory", "D-Negative", and "E-Very negative". This is coded into 5 to 1 respectively for quantitative analysis. For qualitative analysis of what affected students' group experience, two open response questions were examined. First, students were followed up with the option to elaborate on their group experience after they rated it.

Additionally, I retrieved qualitative data from an open response in the post-course survey to get a direct reflection of what impacted students' learning experience. The question asks "Please let us know why your experience with your pod-peers was negative". Since only 7 students responded to this question, which is insufficient to make a representative conclusion, we decided to include open responses from another question to increase our data sample. Students provided more lengthy qualitative feedback in the question "Is there anything else at all you want us to know about your experiences with NMA in general?". We retrieved 216 responses from this question and focused on comments that are relevant to group work and inclusive experience (N = 45) from this question for further analysis.

Analytical Methods

Prior to exploring the relationship between group learning experience and students' sense of belonging, I first conducted a Spearman's correlation analysis between group experience and field belonging using pre- and post-survey data. Preliminary checks revealed that the variables did not meet the normal distribution assumptions required for Pearson's correlation. Consequently, I employed Spearman's correlation, as it is well-suited for assessing monotonic relationships between non-parametric data and ordinal variables.

To test the hypotheses related to the first research question, ordinary least squares regression was used to examine the relationship between group experience and field belonging (Model 1). Control variables included field belonging measured in the pre-course survey plus various individual demographic variables such as prior knowledge, first language, education level, and country of residence. Prior knowledge was assessed in the same manner as academic readiness in Study 1, using a knowledge test consisting of 12 items that evaluated learners' understanding of basic math and computation concepts (full details can be found in Appendix B). Other demographic variables were coded as binary variables: first language indicated whether English is the individual's first language (0 = no, 1 = yes), education level indicated whether learners were currently enrolled in an undergraduate or graduate program, and country of residence indicated whether learners resided in the United States or not.

To test whether male and female students' sense of belonging was differentially related to group experience (RQ2), gender was added as an interaction term in the moderated model (Model 3). Additional regression models were conducted with field self-efficacy and STEM identity as controls (Model 2 & 4). The addition of these models were driven by the theoretical

association between field belonging and STEM identity and self-efficacy. This is to test whether by adding STEM-related constructs, the results would alter significantly.

Table 2

Summary of regression models

Model	1	2	3	4
Independent variable	Group Experience	Group experience	Group experience*Female	Group experience*Female
Control variables	Demographics	Demographics + STEM attributes	Demographics	Demographics + STEM attributes

To explore the factors contributing to students' group learning experience (RQ3), we conducted an inductive thematic analysis to identify prominent themes in student feedback. The coding process involved three researchers and unfolded as follows: I initially reviewed the entire dataset and conceptualized a preliminary codebook based on this initial review. After discussions and revisions with the two other raters, the codebook was finalized. Full details of the inductive coding framework is provided in Appendix D. Once finalized, notable quotes were highlighted to serve as illustrative examples of the identified themes.

Results

Quantitative Outcomes

Based on the correlation analysis, we found a positive significant correlation between group experience and post belonging with rs=0.160, $p < .05$. This suggests that there is a significant positive association between group experience and field belonging in the post-survey data. Results presented in Table 3 further illustrate the relation between group experience and belonging.

From the regression model outcomes, Model 1 and 2 reveal that group experience predicts post-course field belonging β= 0.1617, $p < .01$ after controlling for student demographics and STEM attributes. This suggests that facilitating positive group experiences is important because they generally enhance students' sense of belonging across the board. The main effect coefficient for group experience was positive but not statistically significant in the interaction models. The positive interaction term in Model 3 & 4 suggests that the relationship between group experience and sense of belonging varies by gender. Specifically, the impact of group experience on belonging is more pronounced for females than for males. Our results show that the association between group experience and the sense of belonging increases by .33 units for females compared to males at a statistically significant level ($p < .01$). The interaction analysis also suggests that being female is initially associated with a significantly lower sense of belonging β = -1.55, $p < .01$. The model outcomes did not vary significantly with or without STEM attributes as controls, indicating that field self-efficacy and STEM identity had little influence on the relationship between group experience and field belonging.

Table 3
Relationship between Group Experience and Field Belonging, Moderated by Gender

	Field Belonging			
	(1)	(2)	(3)	(4)
Group experience (β1)	.16 (.05)***	.16 (.05)***	.019 (.063)	.018(.063)
Female (β2)	-.07 (.074)	-.07 (.074)	-1.55(.43)***	-1.56(.43)***
Group experience x Female (β3)			.33 (.10)***	.34(.10)***
N. of Obs	444	444	444	444
Subject demographic controls	Yes	Yes	Yes	Yes
Subject psychological attributes control	No	Yes	No	Yes
Adj. R2	.36	.37	.378	.384

Note. This table shows OLS regressions using field belonging measured in the post-course survey as a dependent variable. The regression model controlled for belonging measured in pre-course survey, prior knowledge, gender, first language, education level, country of residence (in or not in the U.S.). Column (2) and (4) additionally controls for the individual's psychological attributes around STEM learning, including field confidence, field identity measured in pre-course survey. Standard errors in parentheses.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Qualitative outcomes

Three main themes emerged from learners' qualitative responses about their learning experience (RQ3). Students pointed out two prominent areas that should be given more attention in order to improve group dynamics: 1) Engagement and Participation. 2) Skill and Knowledge Disparities. Additionally, students mentioned other barriers that are directly or indirectly associated with group dynamics. We discuss some of the implications along with the results.

Engagement and Participation

Participants reported that disengagement was the biggest barrier to group dynamics, which negatively impacted learners' experience. Students reported that a lack of engagement

from teammates and inactivity decreased their motivation to engage. The presence of "ghost members" had led to frustration and disappointment in other team members. The lack of active discussion and interaction also turned many group activities into a one-person show where the majority of students remained silent, making it more appealing for individuals to study alone rather than collaboratively in the group setting. One student said, "*I was quite disappointed that most students of my pod did not show up. As a result, we were only 3 people during the brainstorming of day two. I feel this limited the potential of project ideas we could come up with*." Another student mentioned that "*most of them were not interactive. I was energized at first but they put me off, they sometimes wouldn't even answer me. Wouldn't communicate how they want the session to go.*" This is a common challenge in online courses, particularly among geographically dispersed learners. He, Xu, and Kruck (2014) suggested that the lack of physical presence exacerbates issues related to accountability and active participation in online group work. Online group projects often suffer due to limited communication, leading to uneven workload distribution and decreased overall group efficacy. Martin and Bolliger (2018) found that students from different geographic locations face additional barriers such as varying time zones and cultural differences, which make synchronous collaboration more challenging. This might explain the disengagement of our participants as well, as they also face communication challenges that stem from cultural differences, time zones, and different goal orientations.

Skill and Knowledge Disparities

The second most prominent area that influenced learner experience and feelings of inclusivity was group composition. Specifically, the diversity of expertise levels within a group emerged as a notable area for attention. Participants expressed divergent opinions about whether groups should have a matching level of expertise and knowledge backgrounds. Some suggested

that students should be matched at a similar expertise level because the knowledge gap contributed to a less inclusive and less efficient environment. One student believed that putting students at varying levels of previous experience in one group poses challenges for both sides to engage in meaningful ways. "... *While some of my peers in my pod had a good foundation in neuro-data analysis, python, or research in general, others were completely new to all of these topics. I think this led to the inexperienced students being overwhelmed at times and the more experienced students not being able to engage in the material in a deeper way…I felt like skill level could have been matched a bit better*." Another student recognized that while there are benefits to mixing novice and advanced students, a more similar knowledge level could reduce the intellectual discomfort of individuals and increase group cohesion. The student further expressed the desire to have the option to choose a different group that might be a better fit for them. "*While being in a mixed group may help them learn from peers, sometimes, it might lead to them lagging behind as their doubts can get lost/ diluted in group of more advanced students. In order to optimise 'utility', I suggest to take into consideration the willingness of students to switch pods/ groups in case they are not intellectually comfortable with the learning approach of the group…I believe I could have used this course better if I were in a less advanced group or a group where initiative is taken from TA's end to explain/ engage in inclusive discussions to give better intuitions*"

Conversely, some students did not perceive barriers to belonging within the groups of varying levels of expertise. We found that the intellectual discomfort of novices might be mitigated by other teammate's openness and inclusiveness. One student expressed a bolstered sense of belonging despite being in a group with experts: "*I was placed in a pod of post-docs,research scientists and PhD's and my knowledge was very limited but nevertheless I felt as*

if we all belonged together and that we were all learning together. This sense of oneness and happiness is what that has made me not give up on NMA and work towards it. It also built my confidence levels." This suggests that positive group dynamics can increase academic motivation and boost learners' confidence.

Additional Insights

Students identified various factors that directly or indirectly influenced their group learning experience. Notably, many students pointed out the intensity of the coursework as overwhelming, which could diminish their own motivation or their teammates' motivation to engage in group activities after being exhausted by other learning tasks. One student specifically mentioned experiencing fatigue from additional Zoom meetings. This is a difficult balance to strike, as the course team should keep the course sufficiently challenging is important to ensure that students make meaningful learning progress. However, given the knowledge disparities previously discussed, the rigorous coursework could heighten the stress and anxiety particularly felt by novice learners within group settings. This suggests that breaking down the material into more manageable segments for those feeling overwhelmed, or covering fewer topics per day, might be an effective approach to mitigate these issues.

While men and women shared similar concerns regarding skill disparity and group engagement, workload management surfaced predominantly by women. Addressing gender-specific nuances can help in creating a more balanced and effective learning environment for all participants. One learner surfaced a unique point about the invisibility of personal circumstances outside of learning. One participant pointed out that as a woman and a parent with a small toddler, the tight course schedule and strict policies were incompatible with the demands of parenting, and left her little time to attend to responsibilities such as washing, tidying, cleaning,

cooking, grocery shopping, etc. As a result, the learner felt "*really heavily the roadblocks young parents/women encounter in science/career.*" This situation highlights a complex but unfortunately common challenge that could hinder non-traditional learners, particularly young mothers, from advancing their science careers. The participant pointed out that a lack of detailed course schedule information made it difficult for her to plan childcare and other responsibilities effectively. Regarding group work, it also adds stress and uncertainty as the group meeting time can be dependent on other teammates' schedules. This highlights the need for teaching assistants and course facilitators to be cognizant of the unique needs of non-traditional learners and be proactive in offering resources or suggestions to assist them. This might include offering a more flexible absence policy and potential make-up assignments, and making sure there is productive coordination amongst group members. These adjustments might help alleviate stress and improve the overall learning experience, making the course more inclusive and accessible to all participants. underscores the importance of understanding and accommodating the diverse personal circumstances of students.

Several other participants noted that language barriers impacted their perceived group effectiveness. Although all participants chose English as the primary language for group communication, varying levels of language proficiency can create challenges in exchanging ideas and collaborating effectively. In the pre-course survey, 34.3% of respondents indicated language barriers as a significant concern, second only to peer programming, which was chosen by 44.3% of participants. As a result, implementing some measures could enhance group dynamics and overall effectiveness. For instance, group formation based on language preference should continue to be implemented. Additionally, teaching assistants can conduct structured ice-breaker activities aimed at building rapport and understanding among group members, which can

not only enhance social connections but also help ease initial communication barriers. Lastly, the course team could provide resources such as online blogs or videos on communication skills and effective group collaboration, helping participants develop strategies to overcome perceived language-related challenges.

Discussion

The findings from this study underscore the importance of online group experiences in the educational context. Our results support the hypothesis that group experiences are positively associated with field belonging. Furthermore, the reduced significance of the main effect of group experience when the interaction term is included suggests a potential moderating effect of gender. The interaction model outcomes reveal that the beneficial effects of group experience are not uniform across all students but are influenced by gender. Specifically, the positive effect of group experience on sense of belonging is more pronounced for female students than for male students. This suggests that while females generally report a significantly lower sense of belonging, a positive group experience can potentially counteract this negative association, enhancing females' sense of belonging to a greater extent. Overall, these findings highlight the value of group experiences in fostering a sense of belonging and their critical role in promoting gender equity in STEM fields.

The differentiated effects of group experiences observed in this study suggest that implementing belonging interventions and inclusive strategies in small group settings could be particularly effective for female students in STEM. For instance, previous studies have indicated that incorporating discussions about contributions from underrepresented groups in STEM, including women, can positively influence female students' self-perception and sense of

belonging (Singer et al., 2020). This implies that targeted group-based interventions could play a crucial role in enhancing the educational experience and retention of female students in STEM fields. Future research could examine whether such discussions carried out in a small group setting would benefit female students more, compared to when an intervention is applied to the entire class or cohort. This may have important implications for large-scale online courses to determine resources and efforts when facilitating group interactions.

We provided additional qualitative insights to illuminate the personal and interpersonal dynamics that contributed to students' learning experience. We found that attrition and imbalanced participation remain a persistent challenge in online group work. Additionally, group interactions on top of intense coursework can become a stressor when not facilitated well. Addressing these issues requires a deliberate effort to create interactive and inclusive learning environments that foster community and active participation. In practice, this could mean more frequent check-ins with the group and collecting feedback from students about team dynamics. Additionally, while diversity can spark creativity and ideas, it also presents unique challenges for groups that have varying levels of prior skills and knowledge. This is particularly true for the context of a short-term intensive course like the one we investigated. The learner feedback indicates that stratifying groups by expertise can be helpful, although the exact range of. As Neuromatch academic performs group matching with the assistance of artificial intelligence, prior knowledge can be a factor to be considered when sorting future learners in teams.

These findings offer practical implications for facilitating an inclusive learning environment. Providing additional warm-ups for learners who lack solid foundations in certain topics in advance may reduce the friction in group work due to knowledge gaps. Our findings also demonstrated the crucial role of characteristics of teammates, which could either buffer or

exacerbate the challenges knowledge discrepancies add to the groups. As a result, providing collaboration guidelines to enhance group cohesion and collaborative behavior could be beneficial. Specifically, for the nature of computational neuroscience courses and other computation heavy courses alike, providing scaffolding and guidelines on how to code collaboratively and distribute work in collaborative teams could foster more organized teamwork. Lastly, while it presents unique challenges in managing a large number of synchronous groups, mechanisms that allow students to opt out of inactive groups or automatically switch non-responsive students as observers might increase learners' agency.

Our study has a few limitations. First, we found that women's sense of belonging is more sensitive to their group experience than men. However, our model could not explain what specific factors in the team contributed to men and women's experiences that subsequently predicted a sense of belonging. Our qualitative findings indicate that situational factors within teams, such as team composition, expertise levels, and the relative engagement of teammates, can influence learners' group experiences. Future studies that directly investigate these factors using quantitative data could provide more robust evidence regarding the elements most critical for fostering an inclusive and supportive environment. Second, although teaching assistants were assigned to small groups, we did not account for the potential impact of individual TAs' characteristics on the learners' group experience. Despite adhering to structured pedagogical content, TAs' personalities and teaching styles may influence group dynamics and learners' experiences. Second, we did not conduct follow-up assessments to determine whether the effects of group experiences on field belonging were sustained after the program ended. We recommend conducting follow-up surveys to examine any potential fade-out effects. Third, students in this study were intrinsically motivated, which contributed to a high level of sense of belonging

overall. They identified additional factors beyond group dynamics— specifically, the structure and intensity of the course content—that influenced their sense of belonging and feelings of inclusivity. To better understand which aspects of group activities contribute to learners' experiences and sense of belonging, we suggest implementing weekly surveys. Lastly, regarding data collection and study design, we acknowledge that the demanding learning schedule may cause students to experience survey fatigue. Therefore, we should carefully consider the length and frequency of survey design when conducting future research within an authentic learning environment. Potential incentives could be provided to increase the participation rate.

Future research should consider tracking students' sense of belonging and experiences over time. For instance, weekly surveys can be distributed to examine whether the relationship between group experience and belonging fluctuates. In this study, we focused on the link between group experience and a sense of belonging within a specific context. Future studies should investigate whether students' sense of belonging to the group mediates this relationship. Additionally, conducting follow-up surveys after one month and six months could strengthen the validity of the field belonging measure by observing any changes in students' responses over time.

Study 3: Unpacking Collaborative Problem-Solving Interactions: the Role of Gender Composition for STEM teams

Introduction

Effective and productive communication is essential for the success of virtual team collaborations (Griffin, McGaw, & Care, 2012; von Davier & Halpin, 2013). Teamwork is inherently social, requiring the exchange of ideas and a shared understanding of the problem at hand to be successful. Demographic characteristics such as gender can significantly influence this process. Previous research has highlighted distinct roles and communication patterns based on gender, emphasizing the importance of addressing individual differences to create more inclusive online learning environments (Roscoe et al., 2022; Nixon et al., 2024). This focus on inclusivity is particularly important for STEM disciplines, where gender gaps in participation driven by systemic discrimination and structural biases persist. Women often encounter more interruptions and their ideas are challenged and considered less compelling (Blair-Loy et al., 2017; Aguillon et al., 2020). Women in male dominant environments are also more vulnerable to devaluation of competence, and pathologization of woman's personal character and gender (Kim and Meister, 2023). This highlights the critical need for gender representation in educational and team settings, a concerted effort to ensure inclusive and equitable participation in collaboration.

Gender Difference in Computer-Mediated Communication

There is longstanding evidence of gender differences in communication. Traditional gender roles suggest that men's speech tends to be more impersonal, authoritative, and assertive, whereas women's speech is typically more expressive, encouraging, and cooperative (Belenky et al., 1986; Baxter-Magolda, 1992). These patterns are evident in both face-to-face interactions

and computer-mediated environments (Aries, 1987; Tannen, 1990; Koh & Lim, 2012). Early studies focused on gender differences in communication portrayed distinct patterns for men and women. Blum's study (1999) on post-secondary students similarly found that male messages tended to be more confrontational, autonomous, certain, abstract, arrogant, or controlling, whereas female messages were more empathetic, often mentioned self, family, or spouse, and had a cooperative tone. The study also indicated that these communication patterns were more pronounced in online learning environments compared to face-to-face settings, suggesting that such patterns were more tolerated online. Herring (1996) corroborated the findings and suggested that the difference is rooted in power dynamics and biases, and extends to interactions within computer-mediated environments.

Gender differences are also found in affective expressions in online group communication. Kucuktunc et al. (2012) suggest that women express more positive emotions compared to men in a large online community. A large-scale meta-analysis by Gallus and Bhatia (2020) found that women contribute higher-valenced and more affective content in collaborative knowledge production compared to men. Other studies found men often establish dominance over women through hostility in their tones and interruptions (Jacobi and Schweers, 2017; Dupas et al., 2021; Miller and Sutherland, 2023). These gendered communication styles can signify power differences, leading to real-world power differentials between men and women across the private-public continuum, and in both digital and non-digital spheres (Judith 2004; Alvarez et al., 2021).

Understanding gender differences in communication within online environments has significant implications for learning. Gendered communication can indicate various learning processes and outcomes. For instance, Lin et al. (2020) identified linguistic signals that predict

differing academic performance and a sense of belonging for male and female students in introductory STEM courses. Likewise, Hendricks et al. (2024) found a relationship between gendered language and final grades in STEM courses. Gender differences in collaborative problem-solving (CPS) interactions directly influence levels of participation and engagement. Lin et al. (2019) revealed that while women participated less frequently in group settings, they took up roles that facilitated more meaningful interactions during synchronous collaborative activities. Ongoing research investigation into gendered communication in CPS settings is essential for recognizing diverse interaction patterns and advancing equitable practices in STEM fields.

Effects of Gender Composition on Team Communication

Group composition has been extensively researched as a factor influencing collaboration dynamics, with gender being one of the critical considerations in both practice and research (Prinsen et al., 2007). The gender ratio within a team can significantly impact the effectiveness and productivity of STEM groups. While there is a general consensus that increased gender diversity in teams boosts innovation and creativity (Smith-Doerr, Alegria, & Sacco, 2017), bringing in different perspectives also means more divergent ideas and potential conflicts during collaboration. There has been substantial debate regarding the role of gender pairing, and studies on the effectiveness of team compositions are inconsistent in STEM literature.

Multiple studies suggest the benefits of having a higher proportion of women on STEM teams. For instance, Dasgupta et al. (2015) suggest that creating small groups with high proportions of women in male-dominated STEM fields is an effective way to keep women engaged and motivated. They found that female peers in small engineering teams positively

influenced women's motivation, verbal participation, and career aspirations, particularly among first-year students. Similarly, Lin et al. (2020) reported that gender pairings significantly affect CPS outcomes. Specifically, females in same-gender groups were more attentive to the benefits of social learning and demonstrated better performance compared to mixed-gender groups, which, in turn, exhibited more diverse and divergent discussions.

In exploring the mechanisms by which female representation positively impacts team success, Woolley et al. (2010) introduced the term "collective intelligence," defined as an individual's ability to read nonverbal cues and make accurate inferences about what others are feeling or thinking. They argue that the positive effect of having more women on a team can be partially attributed to the higher levels of social sensitivity exhibited by women. Additionally, they suggest that groups with a higher proportion of women demonstrate greater equality in conversational turn-taking, which enables members to be more responsive to one another and better utilize the knowledge and skills of all members (Bear and Woolley, 2011). Supporting these findings, Dowell et al. (2019) provided quantitative evidence indicating that women engage in significantly more productive communication patterns, not only in female-majority groups but also in male-majority groups, compared to their male peers. These insights suggest that greater incorporation of women in teams fosters a more inclusive environment and enhances overall communication and team performance.

However, the effects of gender diversity on team communications remain inconclusive. Some scholars argue that homogeneity may be more advantageous for cohesion and productive communication. Research by Jiang et al. (2017) and Yuan et al. (2019) suggests that single-gender groups more easily reach a decision and exhibit higher responsiveness, particularly among female teams. Additionally, Tomai et al. (2014) contend that single-gender composition

in small groups can promote counter-stereotypical communication styles, with males in same-gender groups displaying more social-emotional communication and females demonstrating more task-oriented communication. In mixed gender teams, gendered communication is more equivocal and contingent upon a variety of contextual factors. In some cases, females post fewer messages and are less likely to continue posting if their messages go unanswered in a mixed gender discussion group (Herring, 2000). In other cases, when gender ratio reaches a balance, groups display enhanced collaboration in group work processes, which are subsequently associated with less social loafing behaviors and more equitable contributions to the group work (Takeda and Homberg, 2014).

Despite increased female representation in STEM and intentional efforts to expand female participation, persistent gender disparities in team communication can still remain a challenge to equitable teamwork. Love et al. (2023) connected multiple sources of data including field observations, social network surveys, and in-depth interviews to reveal that scientific teams often reproduce existing unequal hierarchies and stereotypical gendered interaction patterns. They suggest these behaviors are explained by the process of gender socialization, which creates ingrained perceptions regarding leadership, status, and dominance in interactions (West and Zimmerman, 1987; Butler, 1988). These imbalances can further marginalize women's experiences and limit their contributions, ultimately driving them away from STEM (Williams, Phillips, & Hall, 2016).

Overall, these studies highlight the complex dynamics of gender composition in collaborative problem-solving teams. While increasing female representation can enhance team collaboration and productivity, persistent gender disparities and communication hierarchies continue to challenge the collaborative problem-solving teams in STEM environments.

Therefore, further empirical research is needed to understand and address these intricate gender differences in team communication and the crucial role of contextual factors in advancing collaborative efforts within STEM teams.

Current Study

This study investigates the relationship between gender composition and team communication within an online game-based physics problem-solving environment. By analyzing computer-supported CPS dialogues of triads, I strive to understand the nature of student interaction contributions within teams in an authentic STEM learning task. To achieve these objectives, I compared gender differences across different team compositions. At the first level, I compare single-gender teams and mixed-gender teams. At the second level, I examine individual and team outcomes across specific team conditions: all-male teams; male-majority teams (2 male 1 female); female-majority teams (2 female 1 male); all-female teams.

The first outcome I focused on was collaborative problem-solving (CPS) participation, measured by word count, turn-taking, the proportional turns individuals take up in their given group. The operationalization was motivated by prior research. Previously, the measurement of participation in small group collaborations has been based on various communication metrics, such as word count, number of turns, and sentiment analysis (Hardt et al., 2023). When audio and speech data are available, indicators like speaking time and turn-taking are commonly used to assess participation (MacLaren et al., 2020; Praharaj et al., 2022). Nielsen et al. (2010) quantified participation by counting interactions and evaluating the equality of these interactions through the standard deviation of exchanged words within the group. Feng et al. (2023) adopted a more comprehensive approach, considering both verbal and nonverbal

interactions and analyzing the density and connectivity of interactions in groups with different gender compositions using social network analysis.

Additionally, I investigate individuals' affective expressions during collaboration to identify differences in speech styles that may influence group dynamics. We propose the following research questions and hypotheses:

RQ1.Does gender composition relate to individual students' participation in CPS teams?

RQ2.Does gender composition relate to individual students' affective communication in CPS teams?

RQ3. Does gender composition affect CPS teams' communication frequency and balance?

RQ4. Does gender composition affect CPS teams' affective communications?

Hypotheses

Hypotheses for RQ1

To examine the relationship between gender composition and individual students' CPS participation, we focus on three communication behaviors further described in the methods section: Word Count, Turns, and Proportional Turn-taking. We hypothesize that gender differences in CPS participation will appear differently when we compare male and female students in single-gender teams than those in mixed-gender groups. The first hypothesis is that gender differences in CPS participation will be more pronounced in mixed-gender groups (***H1***). This result is expected because when students only interact with peers of the same gender, their

participation level may be more similar given similar communication styles. This hypothesis assumes that when students are in a group with the same gender as their own, gendered communication does not influence participation discrepancies. But when we put male and female students in mixed-gender groups, it invokes potential contrast in individuals' communication behavior and contrast in perceived power dynamics, resulting in more evident gender differences in CPS participation. Since there is different gender composition in mixed-gender groups (e.g. female-majority groups, and male-majority groups), this hypothesis can be accompanied by the following sub-hypotheses:

H1-1. No significant correlation will be found between gender and individual outcomes (i.e.Word Count, Turns, and Proportional Turn-taking) in single-gender teams.

H1-2. Significant correlation will be found between gender and individual outcomes (i.e.Word Count, Turns, and Proportional Turn-taking) in mixed-gender teams.

H1-3. Female students in male-majority teams will associate with lower average Word Count, Turns, and Proportional Turn-taking compared to male students.

H1-4. Female students in female-majority teams will associate with average lower Word Count, but a similar level of Turns, and Proportional Turn-taking compared to male students.

To test these hypotheses for research question one, we need to examine gender differences under the following four conditions: (1) gender difference in single-gender teams; (2) gender difference in mixed-gender teams; (3) gender difference in male-majority groups; (4) gender difference in female-majority groups.

Hypotheses for RQ2

To examine the effect of gender composition on individual students' affective communication in CPS, we focus on the speaker's objective sentiment, and the speaker's relative sentiment within a group. For this research question, we hypothesize that female students will express more positive sentiments in mixed-gender groups (***H2***). Similar to the reasoning above, we hypothesize that gender differences in affective communication are more subsided in single-gender teams. When students are in single-gender groups, the level of relative emotional expression may be similar within a team. That is to say, the sentiment a male student exhibits relative to his CPS teammates in an all-male team is similar to that of a female student in an all-female team. However, given the literature on gendered communication style, we assume that females are likely to speak in more positive and supportive ones measured by their objective sentiment. In other words, the average speaker sentiment will be higher for females than males. Since there is different gender composition in mixed-gender groups (e.g. female-majority groups, and male-majority groups), this hypothesis can be further broken down into the following sub-hypotheses:

H2-1. No significant gender differences will be found in sentiment in students CPS communication (i.e., speaker sentiment and relative sentiment).

H2-2. Females will likely express more positive sentiments in mixed-gender teams (i.e., speaker sentiment and relative sentiment).

H2-3. Female students in male-majority groups will associate with more positive sentiments compared to their male peers in the group (relative sentiment), and more positive speech in general (speaker sentiment).

H2-4. Female students in female-majority groups will associate with more positive sentiment relative to their teammates compared to male students (relative sentiment), and more positive speech in general(speaker sentiment).

Hypotheses for RQ3

To examine the relationship between gender composition on CPS teams' communication, we focus on two team-level outcomes further described in the methods section: Participation Density and Participation Balance. We first hypothesize that mixed-gender groups would have higher participation density and lower participation balance. (***H3***) This result was expected because mixed-gender groups consisting of students with different cognitive styles would likely take more communication to talk through different ideas during problem-solving. Additionally, the different communication styles may bring about different power dynamics, resulting in imbalance in how much students contribute. Considering the previous work indicating men are more talkative, we hypothesize that the composition with a higher proportion of males will be associated with higher participation density. Combined with the previous work indicating women being more collaborative, we hypothesize that the composition with a higher proportion of females will be associated with a higher participation balance. This hypothesis is combined with the following sub-hypotheses:

H3-1. Mixed-gender teams will be associated with higher participation density and lower participation balance.

H3-2. Male-majority teams will have a higher Participation Density than female-majority teams. **H3-3.** Female-majority teams will have a higher Participation Balance than male-majority teams.

Hypotheses for RQ4

To examine the relationship between gender composition on CPS teams' affective communication, we focus on two team sentiments as an outcome. We hypothesize that single-gender teams would have more positive sentiment than mixed-gender teams (***H4***). We expect this result because conflicts are more likely to happen in mixed-gender teams due to different social and cognitive processes associated with gender. Since females are associated with higher levels of social sensitivity and are considered more adept at emotional expressions, we hypothesize that teams with a higher proportion of females will have more positive interactions. This hypothesis is further elaborated into the following sub-hypotheses:

H4-1. Single-gender teams will be associated with more positive team sentiment.

H4-2. Female-majority teams will have a higher Participation Balance than male-majority teams.

To test hypotheses for research questions one and two, we will examine gender differences under the following four conditions: (1) gender difference in single-gender teams; (2) gender difference in mixed-gender teams; (3) gender difference in male-majority groups; (4) gender difference in female-majority groups. To test hypotheses for research question three and four, we will compare team outcomes under the following conditions: (1) single-gender v.s. mixed-gender teams; (2) a post-hoc analysis on all composition pairings. Table 4 provides an overall view of the level of comparisons for each research question.

Table 4

Overall view analysis plan corresponding to each research question.

Level of Analysis	**Dependent Variable**	**Comparison Groups**	**Conditions**	**Corresponding RQ/Hypothesis**

Individual	Word Count, Turns, Proportional Turn-share	Male v.s. Female	Single-gender teams; Mixed-gender teams	RQ1(H1-1, H1-2)
Individual	Word Count, Turns, Proportional Turn-share	Male v.s. Female	Female-majority teams; male-majority teams	RQ1(H1-3, H1-4)
Individual	Speaker sentiment, relative sentiment	Male v.s. Female	Single-gender teams; Mixed-gender teams	RQ2 (H2-1, H2-2)
Individual	Speaker sentiment, Relative sentiment	Male v.s. Female	Female-majority teams; male-majority teams	RQ2(H2-3, H2-4)
Team	Participation Density, Participation Balance	Single-gender v.s. Mixed-gender	N/A	RQ3 (H3-1)
Team	Participation Density, Participation Balance	Female-majority teams vs. male-majority teams; all-female v.s. all-male; all-female vs. female-majority; all-female vs. male-majority; all-male vs. female-majority; all-male vs. male-majority	N/A	RQ3(H3-2, H3-3)
Team	Group sentiment	Single-gender v.s. Mixed-gender	N/A	RQ4 (H4-1)
		Female-majority teams vs. male-majority teams; all-female v.s. all-male; all-female vs. female-majority;	N/A	RQ4 (H4-2)

		all-female vs. male-majority; all-male vs. female-majority; all-male vs. male-majority		

Method

Study Context

The data were collected as part of a larger study on collaborative problem-solving (Eloy et al., 2019), but only the details pertinent to the present study are reported here. Previous publications have reported findings related to team diversity and CPS processes coded from verbal utterances on task performance (see Stewart et al., 2020; Sun et al., 2022). We obtained permission to use de-identified transcripts, demographic, and individual differences data for analyzing group communication. The data has been used towards assessing the relationship between collaborative problem-solving behaviors and solution outcomes (Sun et al., 2022), understanding how diversity in team makeup predicts outcomes in computer-mediated collaboration (Stewart et al., 2020), and linguistic style matching of collaborative peers in problem-solving tasks (Duran et al., 2024). We extend the previous studies by adding a unique perspective on gender nuances in team communication.

Participants

288 students from two large public universities in the United States (177 from University of Colorado Boulder, and 111 from Arizona State University). Students were assigned to triads based on scheduling constraints. Students are randomly assigned and gender composition was

not an intentional manipulation. A total of 96 teams were formed. The study consists of two parts, including an at-home survey that students completed on their own and an in-lab session where the collaboration task took place. Participants were compensated either with a $50 Amazon gift card (95.8%) or with 3.5 hours of course credit (4.2%). Students participated in a CPS task using Physics Playground, which is a highly engaging, two-dimensional educational game that aims to teach students basic Newtonian physics concepts. More details of the game are described in the next section. Amongst all the participants, 5 reported their gender identity as "Other". We decided to exclude the five participants and their associated teams in the main analysis. A follow-up analysis including Other gender as an inclusive term was conducted and can be found in appendix. We also excluded teams that do not consist of three people.

The final sample for analysis has 264 participants, with 42.7% males (N = 117) and 57.3% females (N = 157). All participants, forming 88 teams, participated in the first block. Three teams did not complete the second block, leaving 85 teams and 255 students participating in the second block. Table 5 demonstrated distribution of male and female students in groups across different gender compositions Since most students participated in both blocks, the total number of participants in both blocks yields 516. Table 6 displays the distribution of teams in the CPS task.

Table 5

Distribution of male and female students in CPS groups by gender composition.

	Gender composition				
Gender	All male	All female	Male Majority	Female Majority	Total
Male	42	-	114	62	218
Female	-	120	57	124	301
Total	42	120	171	186	516

Table 6
Distribution of teams in the CPS task

Gender composition	Block	Team *N*
All male	1	7
	2	7
All female	1	20
	2	20
Male dominant (2M1F)	1	29
	2	28
Female dominant (2F1M)	1	32
	2	30

CPS Task

Upon arrival in the lab, participants were individually assigned to one of three computer-enabled workstations equipped with a webcam. The workstations were either partitioned in the same room using dividers or were located in different rooms (depending on the school where data was collected). They all had video conferencing capabilities and screen sharing through Zoom. Participants communicated through the webcam and headset microphone to see and hear each other.

Physics Playground was used as the problem-solving environment. This is a two-dimensional educational game that aims to teach students basic Newtonian physics concepts (e.g., Newton's laws, energy transfer, and torque) and has been found to be highly engaging (Bosch et al., 2015; Miguel et al., 2014). In Physics Playground, students complete game levels by using the mouse to draw simple machines (i.e., ramps, levers, pendulums, and springboards) that navigate a green ball to a red balloon. All objects that students draw, as well as pre-existing agents in the levels, obey the laws of physics. Figure 2 depicts a team using a lever (pre-existing agent in game) and a weight (drawn by the team) to roll the green ball toward the red balloon. A

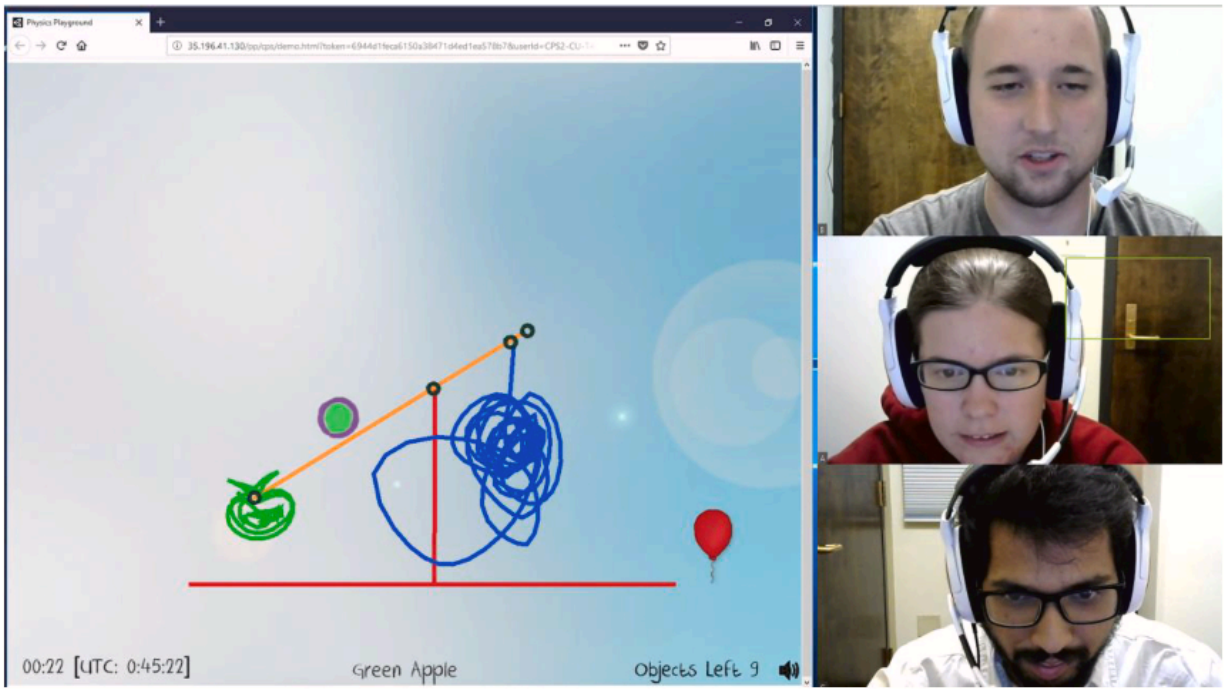
team earns a gold trophy when they successfully navigate the green ball to the red balloon using a few objects. If more objects are used, then a successful solution earns a silver trophy. Students could restart, exit, or change levels at any time during gameplay.

The team first participated in a 15-minute warm-up session. Students were instructed to familiarize themselves with their teammates and play a few levels together. The warm-up was considered a practice round and thus not included in this analysis. Two experimental blocks were subsequently conducted, totaling 30-min of collaborative gameplay. The experiment involves particular goal and physics concept manipulations that were counterbalanced across teams in a 2 ×2 (goal ×concept) within-subject design. Teams were randomly assigned the "energy can transfer" concept for one block, and the "properties of torque" concept for the other block. Each concept involves several levels of gameplay varied in terms of difficulty: nine levels related to "energy can transfer", and eight levels related to "properties of torque". Players could freely navigate through levels in their current playground and pick which levels to solve. There is an additional goal orientation manipulation. One goal orientation instructs teams to "solve as many levels as possible" to prioritize the quantity of levels solved. The other goal orientation instructs teams to "get as many gold coins as possible" to prioritize the quality of solutions. Prior to the start of each block, teams get one of two messages, then they will get the other messages in the second block.

During each block, one teammate was randomly assigned the role of controller and the other two were contributors. The controller was in charge of all mouse interactions within the game while others provided verbal contributions to the solutions. The role of controller was randomly rotated so each teammate served as controller for one block.

Figure 3

The interaction environment of a team playing Physics Playground.



Note. The triad used a lever and weight to navigate the green ball to the red balloon.

Figure 4

A screenshot of the team trying to solve a problem



CPS Discourse Data

Transcripts were created by transcribing audio collected during gameplay. Further details about audio processing can be found in Sun et al. (2022). The original transcript files were recorded at the game level. We aggregated interactions from all levels within a block, which adds up to a total of 15-minute for each block. A total of 48,133 lines of utterances were retrieved. (24,739 utterances were retrieved from the first block, and an additional 23,394 from the second block). An example of the transcript data can be found in Appendix D.

Survey Measures

Students were emailed a survey at least 24 hours prior to their scheduled lab session. The purpose of the survey was to assess individual difference measures, such as demographics, attitudes towards leadership and teamwork, and physics competency. The demographic questionnaire assessed the student's gender, race, age, first language, year in school, whether they are STEM majors, and how many formal physics courses they have taken. Table 7 provides the means and standard deviations of men and women across these variables.

Table 7
Descriptive statistics of participants by gender grouping

Gender composition	Gender	Age M (SD)	First language English M (SD)	School year M (SD)	STEM Major	Prior physics course	GPA
All male	Male	21.43 (3.24)	.71(.46)	2.63(1.31)	.76(.42)	1.61	3.46(.54)
All female	Female	20.91(3.56)	.79(.41)	3.05(1.70)	.77(.43)	.80	3.49(.48)
Male majority group	Male	23.72(5.54)	.80(.40)	3.02(1.55)	.79(.41)	1.43(1.12)	3.30(.51)
	Female	21.12(4.88)	.78(.42)	2.81(1.47)	.89(.31)	1.14(1.11)	3.48(.41)

Female majority group (2F1M)	Male	21.76(3.84)	.70(.46)	3.38(1.43))	.82(.39)	1.22(1.08)	3.36(.80)
	Female	22.51(6.27)	.69(.47)	2.70(1.47)	.86(.35)	.86(1.05)	3.49(.44)

Individual-level Outcome Variables

To quantify individual participation in CPS groups, we measured both the turn-taking frequency and word count for each participant. To contextualize these frequencies within each group, we also calculated the relative participation, representing the proportion of each individual's turns compared to the overall group activity.

To quantify individual sentiment, we applied the VADER (Valence Aware Dictionary and sEntiment Reasoner) Sentiment Analyzer within a Python environment. VADER is a lexicon and rule-based sentiment analysis tool designed to extract sentiment from text (Hutto & Gilbert, 2014). It utilizes a sentiment lexicon, a list of lexical features labeled according to their semantic orientation as either positive or negative, which has been validated to gold-standard quality by human evaluators.

Several considerations guided our choice of VADER for this study. First, VADER is interpretable, empirically validated, and computationally efficient compared to machine learning approaches for sentiment analysis (Hilmy et al., 2019; Knöös and Rääf, 2021). Second, VADER distinguishes itself from other lexicon-based approaches, such as Linguistic Inquiry and Word Count (LIWC), by its superior ability to detect more nuanced sentiment expressions in short tokens. It further incorporates lexical features such as emoticons, sentiment-related acronyms and initialisms, as well as commonly used slang with sentiment value. In comparison to machine-learning classification methods that are often domain-specific, VADER is domain-

independent and performs well across various contexts (Bonta et al., 2019). Additionally, VADER is open-source and has been successfully applied to study learner interactions (Lin & Nixon, 2024). Although the application of sentiment analysis in the educational domain is less prevalent compared to social media research, various sentiment analysis tools have shown promise in automatically annotating online student interactions at scale (Ferreira-Mello et al., 2019; Zhou & Ye, 2023). Educational researchers have demonstrated the effectiveness of VADER in extracting sentiments in multi-party collaborative conversations (Dehbozorgi et al., 2020; Roberts et al., 2024). Consequently, we believe that VADER is an appropriate tool to characterize sentiment expressions for men and women during online CPS interactions in this study.

The levels of sentiment analysis are typically categorized into three types, namely Document level analysis, Sentence level analysis, Entity and Aspect level Analysis. In this study, we adopted sentence level analysis to provide more nuances in an individual's sentiments. For preprocessing, we removed irregular expressions and annotations for crosstalk. Additional excluded filler words can be found in appendix F. We retained annotations for laughter as this can indicate positive valence. With regards to pre-processing, while most studies perform stopword removal (e.g., "and," "the," "is", “he”) with social media texts, stopwords might actually contribute to the context and overall sentiment of a sentence in collaborative interactions, and removing them could potentially alter the sentiment analysis results. VADER is designed to handle a wide range of textual inputs, including those that contain common stopwords. It is also tested by Palomino and Aider (2022) that stopword removal did not heavily influence the model performance. Therefore, we decide to include stopwords in the analysis. .

VADER is both polarity-based and valence-based. It outputs sentiment scores into 4 classes of ‘Negative’, ‘Neutral’, ‘Positive’ and ‘Compound’ with values between -1 to 1. The compound value is the normalized value of the sum of valence scores of each word in the lexicon, adjusted according to the rules.

Team-level outcome measures

Participation Density. Participation density refers to a group's relative activity level compared to all other groups within a given block. It is calculated by dividing the total number of speaking turns of a group by the average number of speaking turns across all groups. Understanding participation density enables us to identify which groups are more communicative. Higher participation density can indicate more active thinking during CPS, greater articulation of ideas, and more frequent attempts to solve problems.

Participation Balance. Participation balance measures the extent to which speaking turns are equitably distributed among group members. This metric is important because an imbalance in participation within groups can suggest potential role differentiation. Notably, Sun et al. (2022) argue that while contributing ideas is crucial for CPS success, it is effective only when combined with active listening from others. Therefore, participation balance in a group may reflect whether all team members are giving each other space to contribute equally or if a division of roles into leaders and followers is occurring.

Group sentiment. We aggregated all utterances for each group and then conducted sentiment analysis based on these collective utterances. The sentiment score for each group represents the overall sentiment valence derived from all interactions. A score closer to 1 indicates more positive group interactions, while a score closer to -1 indicates more negative interactions.

Positive group sentiment may signal supportive and encouraging interactions during CPS activities. Additionally, positive sentiment can indicate a higher level of cohesion and a more open communication environment where members feel comfortable sharing ideas and feedback, which can enhance trust in collaboration. Moreover, positive sentiment can enhance motivation and engagement, contributing to overall team satisfaction and the likelihood of continued collaboration in future tasks.

Table 8

Summary and definitions of outcome variables at individual and group level

Individual-level outcomes	**Definition**
Total Turns	**Total Turns Taken** measures the total number of turns an individual has taken, calculated by aggregating the number of spoken turns at the individual level within a single block.
Word Count	The total number of words spoken, aggregated at the individual level within a single block.
Proportional Turn Share	Proportional Turn Share represents the number of speaking turns for each team member, expressed as a proportion of the group's total number of speaking turns.
Speaker Sentiment	Speaker sentiment is the sentiment score of an individual's speech. This is calculated at the utterance level and averaged out for an individual.
Relative Sentiment	Relative sentiment describes the sentiment of an individual relative to their groups.The measure is calculated by subtracting the speaker sentiment to their group's average sentiment. This variable is standardized to ensure that it has a mean of zero and a standard deviation of one. The transformation is performed so that the value is adjusted to the level of sentiments in the corpus.
Group-level outcomes	**Definition**
Participation Density	A group's relative active level compared to all other groups in a given block. Calculated as the total number of speaking turns of the given group, divided by the average turns of all groups within the same block.

Participation Balance	The standard deviation of speaking turns within the group. This variable represents the degree to which speaking turns are equitably distributed among group members.
Group Sentiment	Group level sentiment score represents the average sentiment of all utterances in a given group. This variable is calculated by first aggregating all utterances in a group. We applied sentiment analysis to generate a sentiment score for each utterance, then took the average of all utterances in a given group to yield the group sentiment score. (footnote: We processed group sentiment at the utterance level because processing at the entire group dialogue level provides little variations and returned an overly positive score for all groups.)

Statistical Analysis

For research questions one and two, the goal for our analysis is to reveal whether there are gender differences in an individuals' CPS participation and affective engagement, and how those differences vary across different gender compositions. To achieve this goal, we conducted Analysis of covariance (ANCOVA) tests to compare the differences between two gender groups (all-male vs. all-female) across individual CPS participation outcomes (i.e. Word Count, Turns, and Proportional Turn-share) and sentiment outcomes (i.e. speaker sentiment, relative sentiment). We conducted five separate ANCOVA tests with each five individual outcomes as a dependent variable. Following that, we conducted additional ANCOVA analysis to compare gender differences in mixed-gender teams (H1-3 & H1-4; H2-3 & H2-4).

The list of controlled variables in the ANCOVA tests include age, English as first language, school year, GPA, number of prior physics coursework, and a binary variable indicating whether students are in STEM majors. Given that we have multiple covariates, we applied a marginal approach by specifying type III sum of squares in the analysis, so that each term is adjusted for all other terms in the model, regardless of the order in which they were

entered. As opposed to type I tests, where gender is assessed first and likely to explain a large amount of the variance, this method computes the sum of squares for each term after accounting for the sum of squares for all other terms, providing a more comprehensive analysis in the presence of unbalanced data.

For research questions three and four, our goal is to examine the relationship between gender composition and team-level outcomes, namely participation density and participation balance. We first conducted an independent sample t-test to compare the mean differences between single-gender and mixed-gender groups across team-level outcomes (i.e. participation density, participation balance) (H3-1; H4-1). The Welch t-test was chosen for this comparison because it does not assume equal variances between the groups and is more robust when the sample sizes and variances are unequal. Then, we conducted an ANOVA to compare how specific gender composition groups differ across team-level outcomes. A Tukey post-hoc is combined with the analysis to compare the means of each group composition pairing (H3-2; H3-3; H4-2).

Results

Individual level outcomes

Male vs. Female in single-gender teams

The means and standard deviations of dependent variables with different gender compositions are presented in Table 9. To examine gender differences in CPS participation and affective communication in single-gender teams, separate analysis of covariance (ANCOVA) was conducted to determine the mean differences between gender groups across all five dependent variables, controlling for Age, English as a first language, School Year, GPA, experimental block, prior physics knowledge, and STEM Major status. The results indicated that females produced significantly fewer words than males, $F(1,123) = 4.89$, $p = .03$, $\eta^2 = .04$. No significant difference was found between genders for other individual outcomes. This result partially supports hypothesis H1-1 and supports hypothesis H2-1.

Table 9
Means and standard deviations of dependent variables at individual level with different gender compositions.

Gender composition	Gender	*N*	Word Count *M*(*SD*)	Turns *M*(*SD*)	Proportional turn-share *M*(*SD*)	Speaker Sentiment [a] *M*(*SD*)	Relative Sentiment [b] *M(SD*)
All male	Male	21	682.31(314.46)	92.50 (26.15)	.33(.10)	.23 (.07)	.00
All female	Female	60	543.00(297.91)	82.88 (36.95)	.33(.07)	.22 (.08)	.00
Male majority group (2M1F)	Male	87	599.97(290.68)	81.47 (28.93)	.35 (.09)	.20(.07)	-.08(.76)
	Female	84	484.91(268.79)	72.00 (32.56)	.31 (.10)	.22 (.07)	.14(.73)

Female majority group (2F1M)	Male	96	693.65(357.60)	92.98 (36.60)	.34(.06)	.20 (.06)	-.21 (.60)
	Female	90	559.77(217.59)	89.69 (33.66)	.33 (.09)	.22 (.06)	.14(.65)

[a] The value of this variable ranges from -1 (negative) to 1 (positive).
[b] This variable is standardized to ensure that it has a mean of zero and a standard deviation of one.

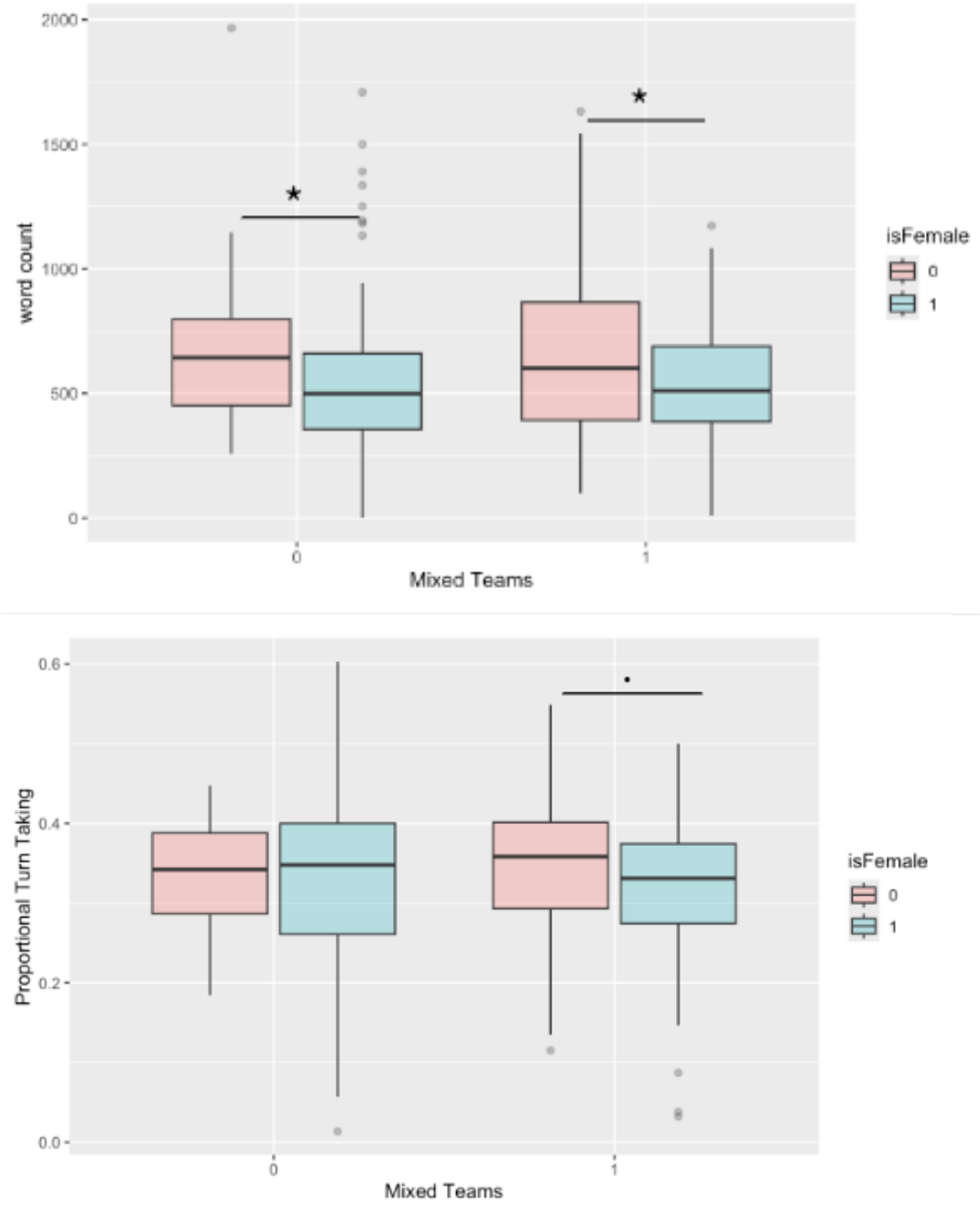
Male vs. Female in mixed-gender teams

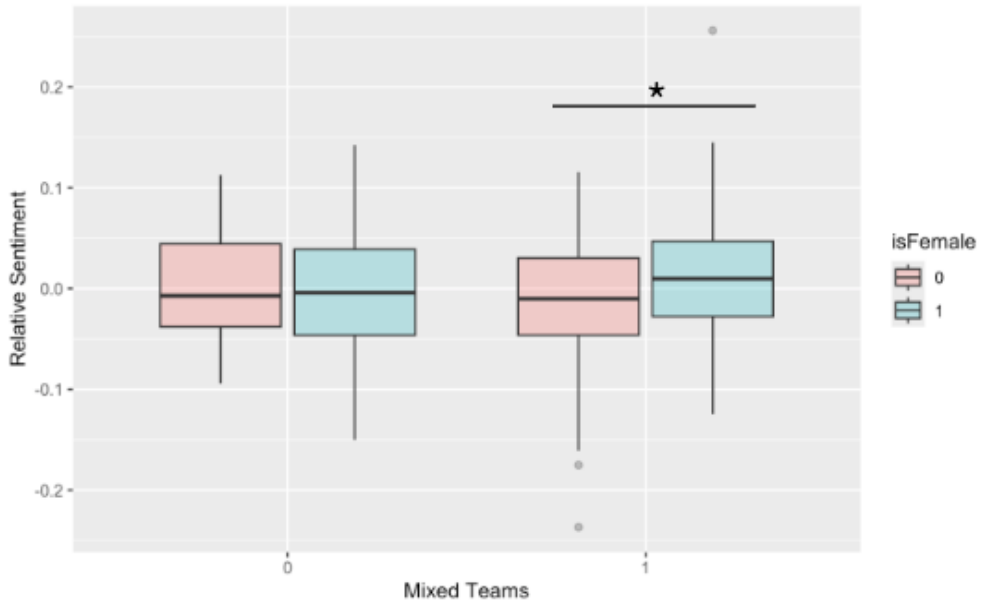
The same set of analysis above was repeated for mixed-gender teams to examine the mean differences between male and females in mixed teams. The results indicated that females produced significantly fewer words than males, $F(1,279) = 7.59$, $p < .01$, $\eta^2 = .036$. No significant difference was found between genders for turns, although females took fewer turns relative to their peers in mixed groups at a marginally significant level $F(1,279) = 3.34$, $p = .068$, $\eta^2 = .012$. As a result, hypothesis H1-2 is partially supported, suggesting that more divergence is found in the outcomes compared to single-gender teams.

We also found that females in mixed teams expressed more positive language on average, $F(1,279) = 5.35$, $p = .021$, $\eta^2 = .018$. When compared to their peers in the group, females expressed significantly more positive relative sentiments, $F(1,279) = 7.50$, $p = .006$, $\eta^2 = .026$. This supports hypothesis H2-2, females in mixed-gender teams engaged in CPS interactions with more positive sentiment expressions.

Figure 5

Gender differences in Single-gender and Mixed-gender teams





Male vs. Female in male-majority teams

The results show that females spoke fewer words $F(1,132) = 16.20$, $p = .056$, $\eta^2 = .027$, exhibited fewer proportional turn share $F(1,132) = 5.76$, $p = .018$, $\eta^2 = .041$, than males in male-majority teams, when we control for demographic backgrounds, not including prior physics and STEM majors. However, gender is not significantly associated with word count when we include prior physics and STEM majors as covariates. In this model, we found GPA, $F(1,130) = 7.67$, $p = .007$, $\eta^2 = .056$, Prior Physics, $F(1,130) = 6.05$, $p = .015$, $\eta^2 = .044$, and STEM Major, $F(1,130) = 18.59$, $p < .001$, $\eta^2 = .125$ all show significant association with word count, with STEM Major being the most significant predictor among the covariates included in the model. Other covariates such as Age, English as a first language, school year, and block level do not have a significant association with word count. This indicates that in male-majority teams, higher GPA, a background in physics, or majoring in STEM fields are more important factors to word count. Similarly, Prior Physics, $F(1,130) = 4.29$, $p = .04$, $\eta^2 = .030$, and STEM

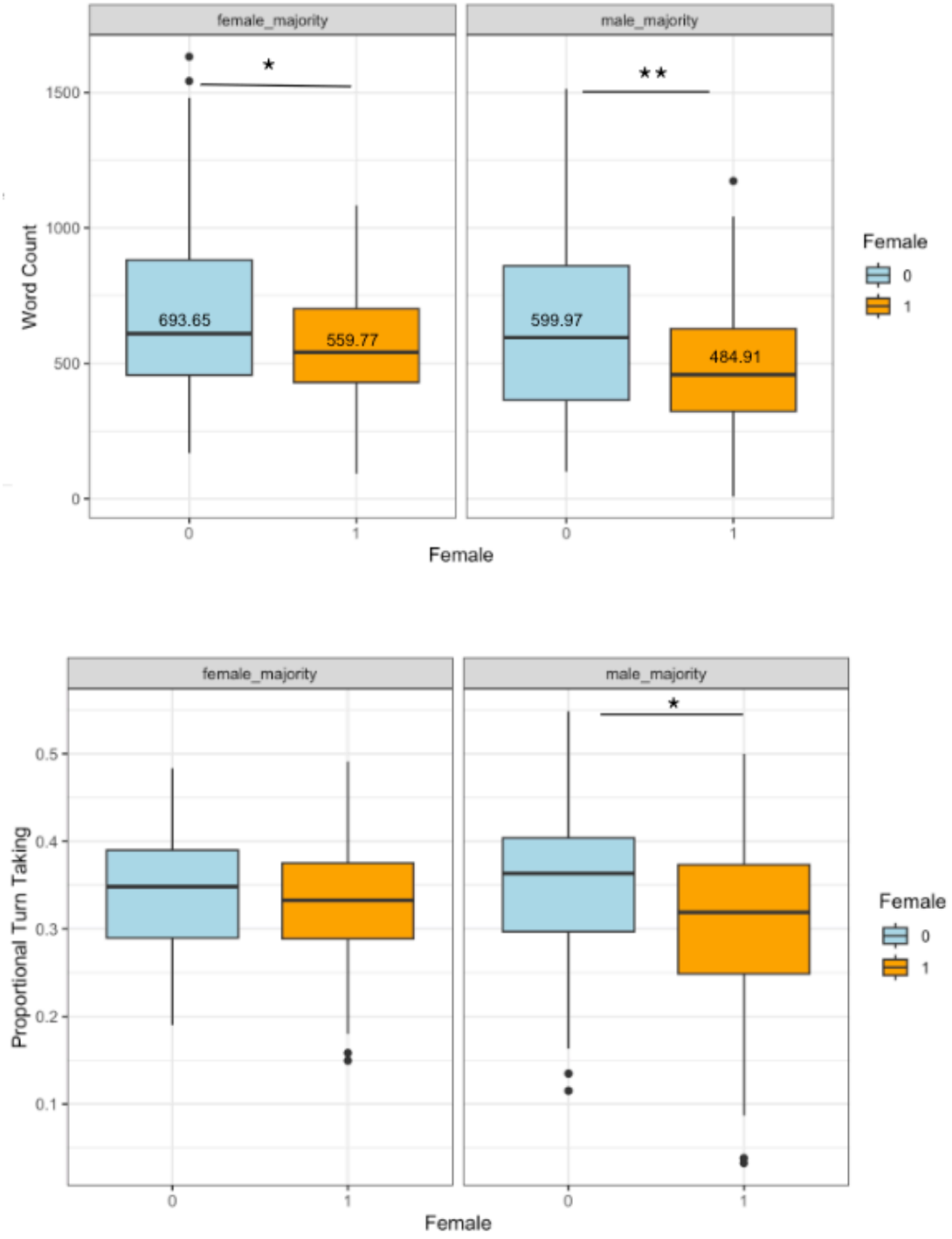
Major, $F(1,130) = 7.20$, $p = .008$, $\eta^2 = .052$ show significant association with turns. In terms of proportional turn-taking, only STEM Major showed a significant association, $F(1,130) = 12.65$, $p = .004$, $\eta^2 = .061$. These findings highlight the effects of academic background rather than gender on CPS communication in male-majority groups. With regards to sentiment, we found that females expressed more positive relative sentiment to their teammates in their CPS interactions $F(1,130) = 3.79$, $p = .054$, $\eta^2 = .028$.

Male vs. Female in female-majority teams

The results show that females spoke fewer words $F(1,140) = 10.43$, $p = .001$, $\eta^2 = .069$, than males in female-majority teams. English as a first language $F(1,140) = 7.36$, $p = .008$, $\eta^2 = .050$, and STEM major $F(1,140) = 6.92$, $p = .009$, $\eta^2 = .047$, also showed significant relation to word count. English as a first language also correlates to turns, $F(1,140) = 4.69$, $p = .032$, $\eta^2 = .032$, although no other factors were associated with this dependent variable. No meaningful differences were found on proportional turn-taking. We did find the females expressed more positive sentiment in their speech compared to their teammates, $F(1,140) = 5.35$, $p = .022$, $\eta^2 = .036$, so does students who are in STEM majors $F(1,140) = 4.40$, $p = .038$, $\eta^2 = .030$.

Figure 6

Gender differences in female-majority and male-majority teams



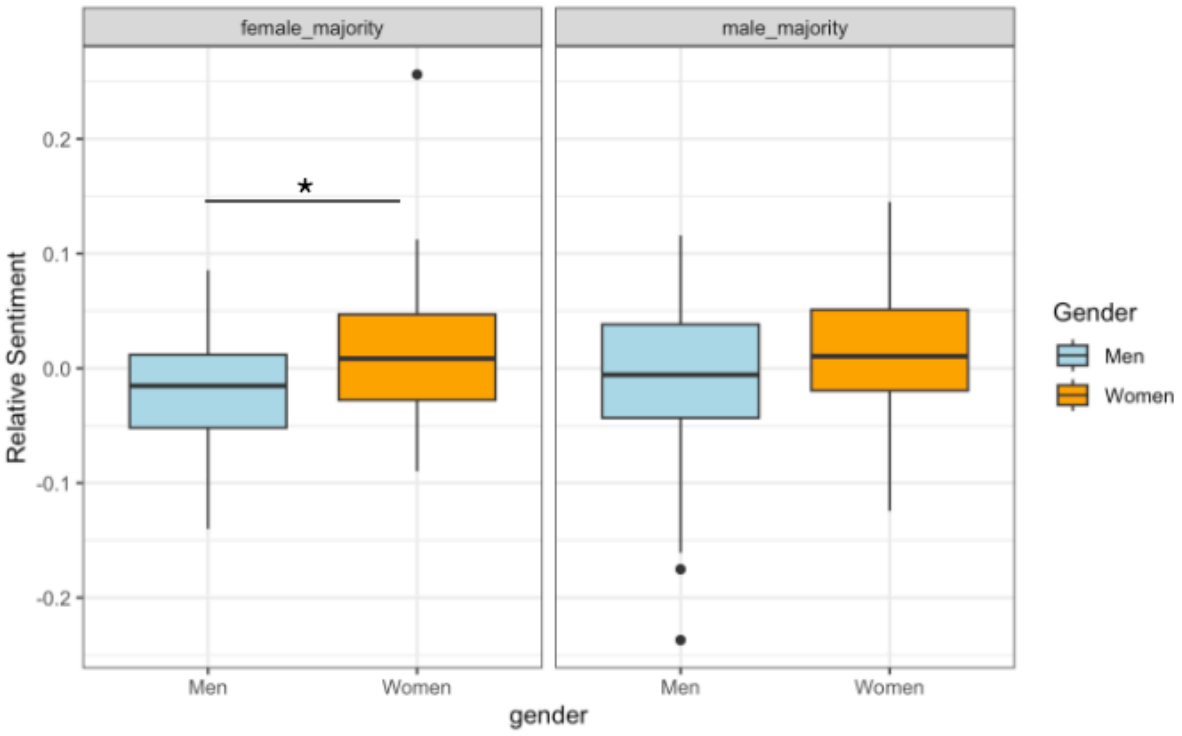


Table 10

Summary of results at the individual level.

Corresponding RQ/Hypothesis	Conditions	Results	Results
RQ1(H1-1)	Single-gender teams;	Partially supported	No gender difference in individual outcomes, except on Word Count (F<M).
RQ1(H1-2)	Mixed-gender teams	Partially Supported	Word count (F<M) Turns (F<M, not significant) Proportional turn share (F<M, marginally significant)
RQ1(H1-3)	Male-majority teams;	Partially supported	when controlling for basic demographic background: Word count (F<M); Proportional turnshare (F<M) Academic background in STEM has a stronger effect (STEM major and prior physics experience)

RQ1(H1-4)	Female-majority teams	Partially supported	Word count (F<M)
RQ2 (H2-1)	Single-gender teams	Supported	No evident gender differences in Sentiment.
RQ2 (H2-2)	Mixed-gender teams	Supported	Speaker Sentiment (F>M) Relative sentiment (F>M)
RQ2 (H2-3)	Male-majority teams;	Partially supported	Speaker Sentiment (Not significant) Relative sentiment (F>M)
RQ2 (H2-4)	Female-majority teams	Partially supported	Speaker Sentiment (Not significant) Relative sentiment (F>M)

Team level outcomes

We examined the relationship between gender composition on CPS teams' communication. At the group level, we focused on three outcomes: Participation Density, Participation Balance, and group sentiment.

Single-gender vs. mixed-gender teams

Table 11 demonstrates the means and standard deviations of dependent variables at team level between single and mixed-gender compositions. A Welch t-test to compare the mean differences between single-gender and mixed-gender groups. We found no difference between single-gender and mixed-gender groups in participation density. However, the results indicate a significant difference in means between the groups on participation balance $t(98) = 2.2$, $p = .03$. We did not observe differences between single-gender and mixed-gender groups on group sentiment.

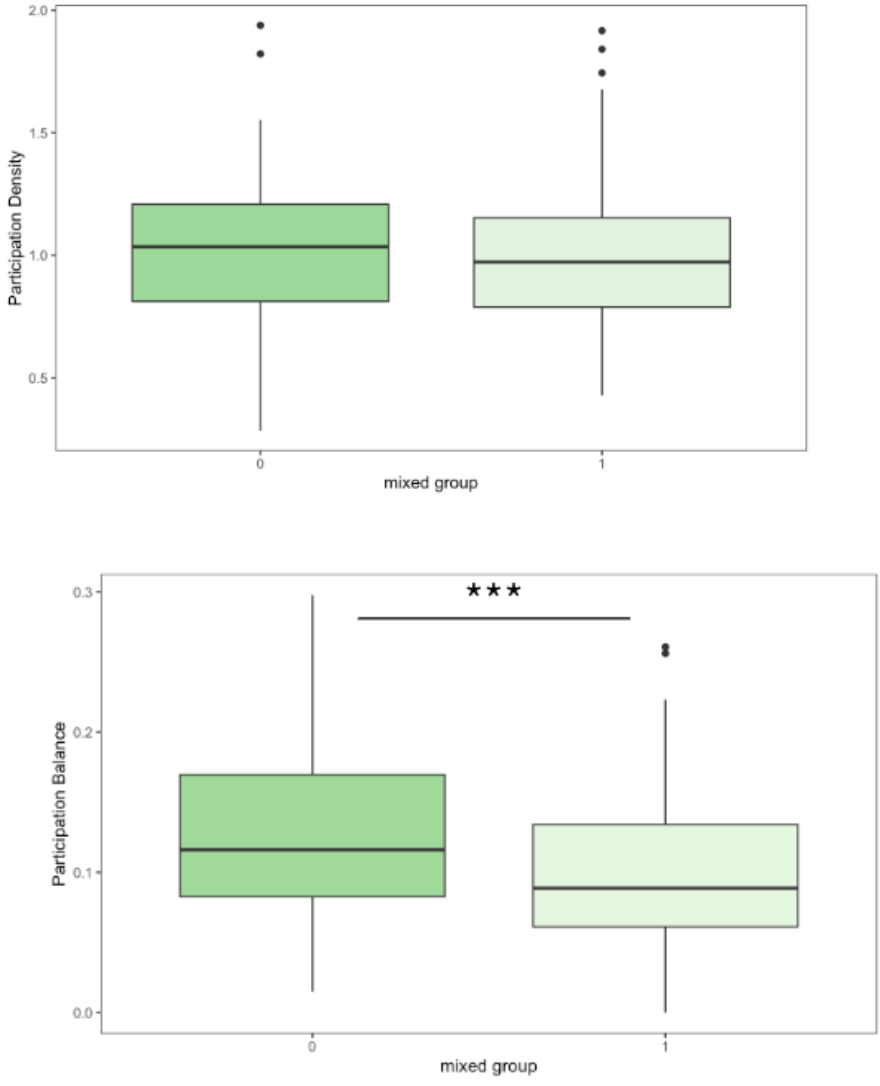
Table 11

Means and standard deviations of dependent variables at the team level with single and mixed gender compositions.

Gender composition	*N*	Participation density *M*(SD)	Participation balance *M*(SD)	Group sentiment *M*(SD)
Single gender teams	54	1.011(.321)	.119(.057)	.219(.046)
Mixed gender teams	119	1.004(.318)	.099(.055)	.210(.036)

Figure 7

Participation Balance in Single-gender and Mixed-gender team



All team compositions

Table 12 demonstrates the means and standard deviations of dependent variables across all team compositions. We examined whether these differences between groups are meaningful.

Participation Density. The ANOVA test revealed a statistically significant difference in participation density among the gender composition groups, $F(3, 169) = 2.64$, $p = .05$, $\eta^2 = .044$. This means that at least one group mean is significantly different from the others. Post-hoc comparisons using the Tukey HSD test further indicated that the mean participation density for the male-majority group was significantly lower than the female-majority group (mean difference = -0.148, $p = .05$).

Participation Balance. The ANOVA test demonstrated a statistically significant difference in participation density among the gender composition groups, $F(3,169) = 4.09$, $p < .01$, $\eta^2 = .068$. Based on the subsequent post-hoc Tukey HSD test, we found a significant difference in participation balance between female-majority teams and all-female teams ($p < .01$), where all-female teams ($M = .13$) exhibited higher participation balance than female-majority teams ($M = .09$). This rejects the hypothesis that female-majority groups will have higher participation balance than male-majority groups. In fact, female-majority displayed slightly lower balance than male-majority, although the result is not significant. This also suggests that the proportion of female participants is not necessarily associated with participation balance.

Group Sentiment. We did not observe significant differences across groups on group sentiment.

Figure 8

Participation Balance across all team compositions

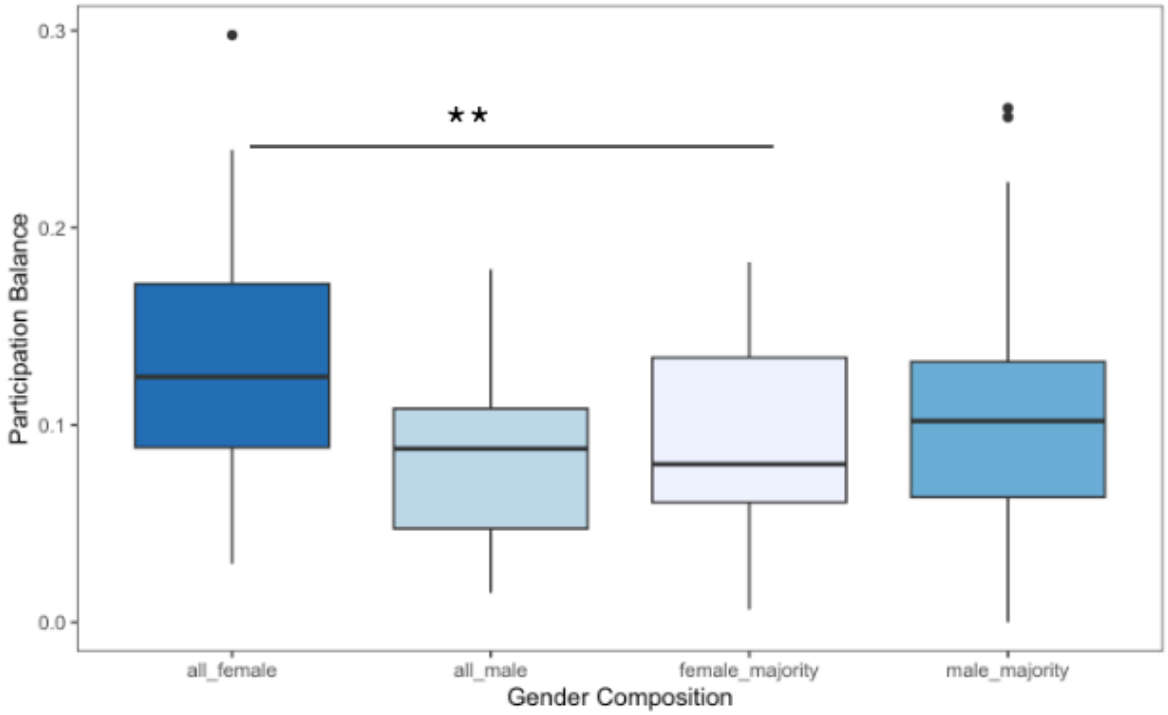


Table 12

Means and standard deviations of dependent variables at team level of all compositions.

Gender composition	N	Participation density *M*(SD)	Participation balance *M*(SD)	Group sentiment *M*(SD)
All male	14	1.095(.207)	.091(.048)	.226(.044)
All female	40	.981(.349)	.129(.057)	.217(.047)
Male majority group (2M1F)	57	1.075(.252)	.106(.060)	.206(.033)
Female majority group (2F1M)	62	.927(.355)	.092(.049)	.214(.038)

Table 13
Summary of results at group level.

Corresponding RQ	Variable of interest	Assumptions	Hypothesis supported?	Results
RQ3(H3-1)	Participation Density, Participation Balance	Mixed-gender teams will be associated with higher participation density and lower participation balance.	Partially supported	Mixed-gender teams exhibited lower participation balance, though no difference in participation density.
RQ3(H3-2)	Participation Density	Male-majority teams will have higher Participation Density than female-majority teams.	Not supported	No evident differences on participation density across gender composition.
RQ3(H3-3)	Participation Balance	Female-majority > than male-majority teams.	Not supported	All-female > female-majority.
RQ4(H4-1)	Group sentiment	Single-gender vs. mixed teams	Not supported	No evident differences across gender composition.
RQ4(H4-2)	Group sentiment	all compositions	Not supported	No evident differences across gender composition.

Discussion

This study investigates gender differences in collaborative problem-solving interactions and examines the effects of gender composition on STEM teams. The findings reveal gender discrepancies in CPS participation, particularly within mixed-gender teams, while interactions in single-gender teams show minimal gender differences. This suggests that when working with same-gender peers, students' CPS participation is more comparable, with the exception of word count. Women spoke fewer words in all-female teams compared to men in all-male teams. This

may imply that women were either more succinct in their communication or provided less elaboration compared to men. We did not observe evident gender differences on other individual level outcomes, suggesting that gender homogeneity may benefit cohesion and balance.

Women exhibited significantly lower word count and proportional turn-shares compared to men in mixed-gender teams. Combined with the group-level results indicating that mixed-gender teams exhibit significantly lower participation balance, we further conclude that mixed-gender teams are more susceptible to imbalanced participation. As a result, mixed-gender teams may require additional effort to manage communication disparities effectively. Interestingly, female-majority groups exhibited the lowest participation balance, which was significantly different from all-female teams that showed the highest participation balance. One interpretation is that the presence of a single opposite-gender member in female teams disrupts group dynamics more significantly than in male teams. Psychological literature suggests that the presence of a dominant gender can elicit a perceived threat in the underrepresented group in STEM fields (Shapiro and Williams, 2012; Deemer et al., 2014). Combined with women's greater propensity toward social sensitivity (Woolley et al., 2010), this may lead to perceptiveness and reactivity among women's communication when male peers are present. Meanwhile, men are considered more naturally gifted in STEM and more likely to assume leadership roles in team settings, may be less sensitive to females' communication styles and less inclined to adapt. This may explain our results wherein no notable differences between all-male teams and male-majority teams were observed. Further empirical evidence will be needed to support this reasoning.

Amongst mixed-gender teams, we found that men tend to dominate the conversations characterized by higher proportional turn-share than females. This male dominance in turn-share is most pronounced in male-majority groups, as indicated by higher F-statistics reflecting larger

between-group variance and larger partial eta square indicating larger effect sizes. Notably, in male-majority groups, proportional turn-share is even more significantly influenced by STEM major and prior physics coursework. However, when we account for students' academic backgrounds, the significant effect of gender diminishes, revealing a more nuanced picture of gender differences in STEM and highlighting the importance of considering academic background in such analyses.

Regarding affective communications in CPS teams, we found that women communicate more positively in mixed-gender teams, even after controlling for demographic and academic backgrounds. Females consistently exhibited more positive CPS interactions compared to their male counterparts, both in male-majority and female-majority groups. This finding aligns with previous literature suggesting that women tend to be more supportive and encouraging in collaborative communications. This is significant as Sun et al. (2022) previously found that complimenting and encouraging fellow team members helps foster a positive collaborative environment, which is crucial for successful CPS task performance. Furthermore, it highlights the consistency of females' affective CPS interaction pattern, even when they are in the minority position within the group. At the group level, we did not find evidence to suggest group sentiment vary significantly across different gender compositions.

This study has several limitations. First, we did not account for the potential impact of existing connections between participants. Forty-six students from 25 teams (26%) indicated they knew at least one teammate prior to the study. Familiarity among participants can affect collaboration dynamics, as previous studies have shown (Janssen et al., 2009). Familiarity may introduce different group norms and dynamics that influence team effectiveness and outcomes (Muskat et al., 2022). While this factor might be more crucial for long-term collaborative teams,

it could still have an impact on short-term, spontaneously formed teams like those in our study. Future analyses should consider the "friendship factor" to strengthen our conclusions. Second, we considered communication as our outcome in this study. Future studies that establish a link between team composition, communication, and performance outcomes will provide a more robust argument and elucidate whether certain compositions are linked to more productive CPS processes in teams. Third, our exploration of CPS communication used a binary gender perspective. The academic community increasingly recognizes the need for more inclusive gender definitions. Kube et al. (2024) pointed out that investigating collaboration patterns solely through the lens of male and female differences risks reproducing gender stereotypes. Future research should aim to incorporate a more inclusive and nuanced approach to gender. Lastly, conducting longitudinal studies to monitor the long-term impact of different team compositions on gender dynamics in STEM across various courses in authentic classroom settings would be beneficial. This would help understand the generalizability of our findings.

Overall, our findings indicate that gender representation within teams can influence equitable participation in collaborative problem-solving (CPS) activities. However, merely optimizing numerical gender representation is insufficient; it must be integrated with other targeted interventions to be effective. The persistent participation divides shown in our study suggest communication imbalances in teams may still hinder women from truly feeling welcome and integrated in STEM. As Joshi and Roh's (2009) pointed out, fields classified as more “masculine,” such as engineering and physics, higher rates of gender diversity do little to improve performance outcomes. Furthermore, we must be careful to avoid generating the perception of being a “token” candidate for women, as it can backfire and negatively impact women’s sense of belonging in STEM (Cain and Leahy, 2014). Therefore, intentional facilitation

to improve team dynamics is important to ensure that all members feel valued and heard. Additionally, targeted training intervention that builds interpersonal team dynamics may also help cultivate a sense of psychological safety for team members (Dusenberry and Robinson, 2020).

Contributions to the Field

This dissertation makes a significant contribution to the field by providing empirical evidence on the current state of gender equity and identifying potential areas for future intervention. Studies 1 and 3 challenge some prior trends regarding gender gaps in STEM and highlight the importance of re-examining these issues in novel contexts. Specifically, Study 1 and Study 2 underscore the critical role of a sense of belonging in promoting STEM equity and explore the extent to which belonging can be influenced by less salient factors, such as group experiences. Study 3 further elucidates the intricacies of gender equity in team-based learning environments. It reveals the significant impact of group composition on online STEM interactions and emphasizes the importance of interpersonal interactions in advancing gender equity within STEM teams. Collectively, these three studies provide valuable insights into the issue of gender equity in STEM. They offer actionable recommendations for designing interventions that promote gender equity in online STEM interactions and provide concrete strategies for creating equitable learning environments that drive positive learning experiences for all students.

Study 1 contributes valuable empirical evidence to the growing body of literature on gender inequalities in STEM fields. By examining previously underexplored actors and contexts, the study offers fresh insights into this critical issue within a novel online learning

environment—an intensive, short-term, interdisciplinary STEM program with globally distributed learners. Notably, our research transcends the traditional confines of accredited higher education institutions, diverging from the predominantly location-bound contexts of prior studies. The findings both support and challenge existing beliefs about gender disparities in STEM, highlighting the limited generalizability of previous trends. Consequently, this underscores the need to thoroughly understand the specific context before designing and implementing any interventions.

Study 2 underscores the potential of teamwork in promoting equity within STEM fields. Utilizing empirical data from a large online course with globally diverse learners, we demonstrate the critical role of group experiences in fostering a sense of belonging, particularly for women. We found that women's experiences in group settings in an online course are more strongly correlated with their sense of belonging to the field of study than their male peers. This emphasizes the need for moderated group interactions to foster positive learning experiences and cultivate an inclusive environment, and prompted future studies to explore the links between effectiveness of intervention on group experiences and sense of belonging. While the shift in the entire cultural narrative around gender stereotypes might be slow, interventions targeting the context of teams can be effective.

Study 3 contributes insights for the development of more inclusive and effective collaborative environments in STEM education and beyond. The gender composition of teams significantly impacts gender gaps in participation, with mixed-gender teams being more prone to developing participation disparities compared to single-gender teams. In particular, female-minority teams require more attention and moderation to ensure gender-balanced interactions. While gender diversity can be beneficial to collaborative problem-solving (CPS), it also

introduces potential challenges in communication. We suggest that while team composition is often considered a key strategy, it is equally important to implement facilitation techniques that promote equitable participation in STEM. Institutions and policymakers should prioritize training on collaboration and communication skills to enable more effective and inclusive interpersonal interactions. Additionally, inclusiveness must be explicitly addressed as part of the collaborative classroom culture (Prinsen, 2007). It is important to create a learning environment where all individuals feel empowered to contribute actively and meaningfully.

Implications

Study 1 indicates that although women expressed lower belonging to the field, they were equally driven to pursue a STEM career compared to men. We illuminate areas of opportunity for future interventions. For instance, interventions that focus on both field belonging and general self-efficacy might be more effective in increasing female learners' persistence desire. While male learners are often assumed to be more advantageous in STEM, we should not ignore their learning needs and adequate support, as we illustrated a negative association between their self-efficacy and academic readiness. Overall, this study calls for contextualized support that prioritizes improvement in both learning outcomes and experience for learners of diverse backgrounds. We invite future studies to further investigate novel online learning communities, as broadening STEM participation in a global context involves greater consideration than gender.

Study 2 provides valuable insights and practical implications for the development of future programs or interventions, highlighting the importance of structured group activities in enhancing learning experience and a sense of belonging. We found that interventions such as frequent check-ins with the group and collecting feedback from students about team dynamics

may be beneficial to reducing disengagement issues and promote positive experience. For instance, Providing additional warm-ups for learners who lack solid foundations in certain topics in advance may reduce the friction in group work due to knowledge gaps. Our findings also demonstrated the crucial role of characteristics of teammates, which could either buffer or exacerbate the challenges knowledge discrepancies add to the groups. As a result, providing collaboration guidelines to enhance group cohesion and collaborative behavior could be beneficial. Specifically, for the nature of computational neuroscience courses and other computation heavy courses alike, providing scaffolding and guidelines on how to code collaboratively and distribute work in collaborative teams could foster more organized teamwork. Lastly, while it presents unique challenges in managing a large number of synchronous groups, mechanisms that allow students to opt out of inactive groups or automatically switch non-responsive students as observers might increase learners' agency.

Study 3 indicates that mixed-gender teams may face more communication challenges, the pronounced gender differences can also introduce productive dynamics, such as positive interactions. The aim is not to determine the optimal composition for CPS interactions but rather to highlight potential opportunities and challenges associated with gender compositions in order to inform strategies to enhance future STEM CPS teams. Literature has documented that the inclusion of historically underrepresented identities can alter group attitudes and behaviors (Soler-Gallart, 2017). Specifically, when team members engage in dialogue and interactions rather than leveraging prestige for power, they are more open to reconsidering concepts upon encountering new information. Therefore, facilitating mixed-gender teams with an eye on participation balance provides an opportunity for students to improve their collaborative skills in order to engage more meaningfully in CPS tasks.

Together, these findings have several implications for creating a more inclusive and equitable STEM learning environment. First, we demonstrated that female-majority teams show more positive sentiment and relatively balanced participation, educators and team leaders should consider forming teams with a higher proportion of women to foster more positive and equitable interactions. Given that females in the minority are more likely to participate less and may have their contributions overshadowed by men, it is crucial to pay attention to scenarios where females are the minority in teams. For mixed-gender teams, facilitation techniques such as structured turn-taking or designated roles, may help promote equal participation. Second, providing information on team communication that makes all team members aware of the dynamics of gendered communication may help students see the communication gaps and encourage more balanced participation, as oftentimes participants are unaware of their own communication patterns. Additionally, doing reflection exercises and establishing feedback mechanisms where team members can reflect on their own and their teammate's participation and communication can promote self-awareness and team cohesion over time (Hesse et al., 2015). Third, specific guidance on the importance of equal contribution, active listening, and respectful dialogue can help break down stereotypical communication patterns.

Future Studies

An underexplored area is the intersection of culture and gender in online learning. Research has stated the importance of considering how stereotypes vary across different cultural contexts and how they subtly and perniciously shape women's ability to feel a sense of fit and belonging in STEM, experiences, and motivation for career pursuits (Schmader, 2023). The global online learning communities offer a fertilized ground for future research on this topic.

Future research should consider tracking students' sense of belonging and experiences over time. For instance, weekly surveys can be distributed to examine whether the relationship between group experience and belonging fluctuates throughout the duration of a course. In Studies 1 and 2, I examined the link between group experience and a sense of belonging within the field of study. Future studies should explore whether students' sense of belonging to the group, a more immediate environment that they interact with on a more frequent basis, mediates this relationship. Additionally, conducting follow-up surveys after one month and six months could enhance the validity of the field belonging measure by capturing any changes in students' responses over time. I also recommend revising the sampling of field belonging with more specific disciplinary anchors, especially in interdisciplinary courses, where a lack of distinction can complicate determining to which field students feel more or less belonging. In my future work, I plan to test this in other courses offered by the Neuromatch Academy, such as deep learning, to test for generalizability. This will provide an interesting comparison as the teaching team remains largely the same while the curriculum focuses more on computing aspects.

Another thing we learned from study 2 and a concern voiced by many students was the knowledge gap. Future research should provide further insights into how disciplinary backgrounds and the divide between experts and novices impact learning experiences and outcomes. Such findings could help determine the extent to which knowledge gaps should be managed in group-based learning environments. Additionally, future work should examine the role of teaching assistants in facilitating group learning and its effect on learners' experiences. Theoretical and empirical evidence strongly suggest that teacher characteristics can influence learner experiences. In the community we studied, this might be especially true, as the TAs are crowdsourced volunteers who are either competent in the subject area or have previously

completed the course. Given their similar age to the learners, this context presents an intriguing opportunity to investigate the modeling effect, particularly for women and marginalized groups.

Building on the findings of Study 3, future research should investigate how communication dynamics impact CPS outcomes. Beyond merely examining the link between the numerical representation of female and male students within groups and their outcomes, future research should aim to understand why certain group compositions yield positive or negative effects. Researchers should explore in depth how group compositions influence various group and individual processes, such as communication dynamics and trust building, and how these, in turn, impact learners' experiences and learning gains. Both instructors and researchers will benefit from a deeper understanding of the underlying mechanisms that affect achievement and experience, particularly in underserved groups. While the quantity of participation is a significant predictor of collaborative problem-solving performance, examining the socio-cognitive aspects of multiparty interactions could provide a more nuanced understanding of the quality of collaboration and meaningful contributions (Hesse et al., 2015; Dowell et al., 2019).

Advances in online conferencing and automatic transcription have made data collection for studying collaborative teams and online learning more convenient than ever. The improvement of text mining tools and large language models makes it more feasible to characterize collaborative processes at scale. The rapid improvement of text mining tools and large language models has made it feasible to characterize collaborative processes at scale. These developments present researchers with unparalleled opportunities to extract fine-grained insights from conversations and to uncover various processes associated with a range of student characteristics, such as gender, ethnicity, cultural background, motivation, prior experience and so on. It is crucial for the field to identify significant features and indicators that are both

theoretically and pedagogically meaningful. As novel learning contexts and new forms of analytics emerge in this digital age, existing theories will need to be updated. New frameworks may be required to guide future analyses effectively, with empirical evidence continuously enriching and broadening these frameworks.

The rapid integration of artificial intelligence (AI) into education has sparked essential discussions about designing AI chatbots for various contexts. Researchers are investigating the immense potential of Generative AI to revolutionize education by enabling personalized learning experiences and providing timely support. Within this framework, there are numerous potential applications aimed at advancing equitable participation in online STEM learning. For instance, AI-driven analytic tools can be developed to detect imbalanced participation and intervene in hostile interactions by alerting teaching teams. Social agents might also be designed with rule-based prompts to facilitate online learning groups. Such innovations are already emerging in intelligent tutoring systems, and the future holds promise for AI agents that assist in fostering equal participation, constructive conflict, and inclusive collaborative environments (Seeber et al., 2020). However, the development of these tools must involve a critical examination of algorithmic bias, given the ingrained biases against women and minoritized groups present in training data. Institutions and organizations must also ensure transparency and accountability in their AI use. Finally, it is crucial to involve stakeholders such as educators, students, and decision-makers in designing these learner-centered tools.

Diversity and equity issues in STEM represent entrenched systemic challenges that will require a long-term commitment to address (Berry et al., 2022). Addressing these challenges requires persistent efforts across institutional, interpersonal, and individual levels. Although changes may not happen overnight, we can start by focusing on one group, one course, or one

community at a time, fostering small shifts in culture and narratives. I hope this dissertation serves as a small contribution to this collective endeavor toward a more equitable future.

References

Alvarez, D., Gonzalez, A., & Ubani, C. (2021). The portrayal of men and women in digital communication: Content analysis of gender roles and gender display in reaction GIFs. *International Journal of Communication (19328036)*, *15*.

Aguillon, S. M., Siegmund, G. F., Petipas, R. H., Drake, A. G., Cotner, S., & Ballen, C. J. (2020). Gender differences in student participation in an active-learning classroom. CBE—*Life Sciences Education*, *19*(2), ar12.

Aries, E. (1987). *Gender and communication*. Sage Publications, Inc.

Archer, W., & Davison, J. (2008). *Graduate employability: What do employers think and want* (pp. 1-20). London: The council for industry and Higher Education.

Baxter-Magolda, M.B. (1992).*Knowing and reasoning in college: Gender-related patterns in students' intellectual development*. San Francisco: Jossey-Bass.

Belenky, M., Clinchy, B., Goldberger, N., & Tarule, J. (1986).*Women's ways of knowing*. New York: Basic Books.

Bear, J. B., & Woolley, A. W. (2011). The role of gender in team collaboration and performance. *Interdisciplinary science reviews*, *36*(2), 146-153.

Beede, D. N., Julian, T. A., Langdon, D., McKittrick, G., Khan, B., & Doms, M. E. (2011). Women in STEM: A gender gap to innovation. *Economics and Statistics Administration Issue Brief*, (04-11).

Berry, A., McKeever, S., Murphy, B., & Delany, S. J. (2022). Addressing the" leaky pipeline": a review and categorisation of actions to recruit and retain women in computing education. *arXiv preprint arXiv:2206.06113*.

Blum, K.D. (1999). Gender differences in asynchronous learning in higher education: Learning styles, participation barriers and communication patterns. *Journal of Asynchronous Learning Networks*, *1*(3). Retrieved April 18, 2001, from the World Wide Web:http://http://www.aln.org/alnweb/journal/Vol3_issue1/blum.htm

Blair-Loy, M., Rogers, L. E., Glaser, D., Wong, Y. A., Abraham, D., & Cosman, P. C. (2017). Gender in engineering departments: Are there gender differences in interruptions of academic job talks?. *Social Sciences*, *6*(1), 29.

Burke, A., Okrent, A., Hale, K., & Gough, N. (2022). The State of US Science & Engineering 2022. National Science Board Science & Engineering Indicators. NSB-2022-1. *National Science Foundation*.

Bosch, N., D'Mello, S., Baker, R., Ocumpaugh, J., Shute, V., Ventura, M., Wang, L. and Zhao, W. 2015. Automatic detection of learning-centered affective states in the wild. *Proceedings of the 20th international conference on intelligent user interfaces* (2015), 379–388.

Bonta, V., Kumaresh, N., & Janardhan, N. (2019). A comprehensive study on lexicon based approaches for sentiment analysis. *Asian Journal of Computer Science and Technology, 8*(S2), 1-6.

Botella, C., Rueda, S., López-Iñesta, E., & Marzal, P. (2019). Gender diversity in STEM disciplines: A multiple factor problem. *Entropy*, *21*(1), 30.

Butler, J. (1988). Performative Acts and Gender Constitution: An Essay in Phenomenology and Feminist Theory. *Theatre Journal*, *40*(4), 519–531. https://doi.org/10.2307/3207893

Cain, C. L., & Leahey, E. (2014). Cultural correlates of gender integration in science. Gender, *Work & Organization*, *21*(6), 516-530.

Casad, B. J., Petzel, Z. W., & Ingalls, E. A. (2019). A model of threatening academic environments predicts women STEM majors' self-esteem and engagement in STEM. *Sex Roles*, *80*, 469-488.

Chang, M. J., Cerna, O., Han, J., & Saenz, V. (2008). The contradictory roles of institutional status in retaining underrepresented minorities in biomedical and behavioral science majors. *The Review of Higher Education*, *31*(4), 433-464.

Cheryan, S., Ziegler, S. A., Montoya, A. K., & Jiang, L. (2017). Why are some STEM fields more gender balanced than others?. *Psychological bulletin*, *143*(1), 1.

Cheryan, S., Plaut, V. C., Davies, P. G., & Steele, C. M. (2009). Ambient belonging: how stereotypical cues impact gender participation in computer science. *Journal of personality and social psychology*, *97*(6), 1045.

Chirikov, I., Semenova, T., Maloshonok, N., Bettinger, E., & Kizilcec, R. F. (2020). Online education platforms scale college STEM instruction with equivalent learning outcomes at lower cost. *Science advances*, *6*(15), eaay5324.

Dasgupta, N., & Stout, J. G. (2014). Girls and women in science, technology, engineering, and mathematics: STEMing the tide and broadening participation in STEM careers. *Policy Insights from the Behavioral and Brain Sciences*, *1*(1), 21-29.

Dasgupta, N., Scircle, M. M., & Hunsinger, M. (2015). Female peers in small work groups enhance women's motivation, verbal participation, and career aspirations in engineering. *Proceedings of the National Academy of Sciences*, *112*(16), 4988-4993.

Deemer, E. D., Thoman, D. B., Chase, J. P., & Smith, J. L. (2014). Feeling the threat: Stereotype threat as a contextual barrier to women's science career choice intentions. *Journal of Career Development*, *41*(2), 141-158.

Dennehy, T. C., & Dasgupta, N. (2017). Female peer mentors early in college increase women's positive academic experiences and retention in engineering. *Proceedings of the National Academy of Sciences*, *114*(23), 5964-5969. https://doi.org/10.1073/pnas.1613117114

Dehbozorgi, N., Maher, M. L., & Dorodchi, M. (2020, October). Sentiment analysis on conversations in collaborative active learning as an early predictor of performance. In *2020 IEEE frontiers in education conference (FIE)* (pp. 1-9). IEEE.

Dowell, N. M., Nixon, T. M., & Graesser, A. C. (2019). Group communication analysis: A computational linguistics approach for detecting sociocognitive roles in multiparty interactions. *Behavior research methods*, *51*, 1007-1041.

Dowell, N., Lin, Y., Godfrey, A., & Brooks, C. (2019). Promoting inclusivity through time-dynamic discourse analysis in digitally-mediated collaborative learning. In *Artificial Intelligence in Education: 20th International Conference, AIED 2019, Chicago, IL, USA, June 25-29, 2019, Proceedings, Part I 20* (pp. 207-219). Springer International Publishing.

Dupas, P., Modestino, A. S., Niederle, M., & Wolfers, J. (2021). *Gender and the dynamics of economics seminars* (No. w28494). National Bureau of Economic Research.

Duran, N. D., Paige, A., & D'Mello, S. K. (2024). Multi-Level Linguistic Alignment in a Dynamic Collaborative Problem-Solving Task. *Cognitive Science*, *48*(1), e13398.

Dusenberry, L., & Robinson, J. (2020). Building psychological safety through training interventions: Manage the team, not just the project. *IEEE Transactions on Professional Communication*, *63*(3), 207-226.

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative science quarterly*, *44*(2), 350-383.

Estrada, M., Woodcock, A., Hernandez, P. R., & Schultz, P. W. (2011). Toward a Model of Social Influence That Explains Minority Student Integration into the Scientific Community. *Journal of Educational Psychology*, *103*(1), 206–222.

Eloy, L., EB Stewart, A., Jean Amon, M., Reinhardt, C., Michaels, A., Sun, C., D'Mello, S., et al. (2019, October). Modeling team-level multimodal dynamics during multiparty collaboration. *In 2019 International Conference on Multimodal Interaction* (pp. 244–258).

Farr, B., Rotermund, S., Ho, P., Radwin, D., & Robles, J. (2011). Evaluation of the Academy for College Excellence: Year 1 interim report.

Feng, Q., Luo, H., Li, W., Chen, T., & Song, N. (2023). Effects of gender diversity on college students' collaborative learning: From individual gender to gender pairing. *Heliyon*, *9*(6).

Ferreira-Mello, R., André, M., Pinheiro, A., Costa, E., & Romero, C. (2019). Text mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *9*(6), e1332.

Fischer, S. (2017). The downside of good peers: How classroom composition differentially affects men's and women's STEM persistence. *Labour Economics*, *46*, 211-226.

Fines-Neuschild, M., Hlavacek-Larrondo, J., & Arguin, J. F. (2024). Empowering educators: the key to achieving gender parity in STEM fields. *Communications Physics*, *7*(1), 78.

Fry, R., Kennedy, B., & Funk, C. (2021). STEM jobs see uneven progress in increasing gender, racial and ethnic diversity. *Pew Research Center*, *1*

Gallus, J., & Bhatia, S. (2020). Gender, power and emotions in the collaborative production of knowledge: A large-scale analysis of Wikipedia editor conversations. *Organizational Behavior and Human Decision Processes*, *160*, 115-130.

Griffin, P., Care, E., & McGaw, B. (2012). The Changing Role of Education and Schools. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment and Teaching of 21st Century Skills* (pp. 1–15). Springer Netherlands. https://doi.org/10.1007/978-94-007-2324-5_1

Goff, E. E., Mulvey, K. L., Irvin, M. J., & Hartstone-Rose, A. (2019). The effects of prior informal science and math experiences on undergraduate STEM identity. *Research in Science & Technological Education*, *38*(3), 272-288. https://doi.org/10.1080/02635143.2019.1627307

Gosling, S. D., Rentfrow, P. J., & Swann Jr, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in personality*, *37*(6), 504-528.

Good, C., Rattan, A., & Dweck, C. S. (2012). Why do women opt out? Sense of belonging and women's representation in mathematics. *Journal of personality and social psychology*, *102*(4), 700.

Hall, J. A., & Matsumoto, D. (2004). Gender differences in judgments of multiple emotions from facial expressions. *Emotion*, *4*(2), 201.

Hammond, A., Rubiano Matulevich, E., Beegle, K., & Kumaraswamy, S. K. (2020). *The equality equation: Advancing the participation of women and girls in STEM*. World Bank.

Hardt, D., Mayer, L., & Rincke, J. (2023). Who does the talking here? The impact of gender composition on team interactions.

He, W., Xu, G., & Kruck, S. E. (2014). Online IS education for the 21st century. *Journal of Information Systems Education*, *25*(2), 101-106.

Henricks, G., Perry, M., & Bhat, S. (2024, March). The Relation Among Gender, Language, and Posting Type in Online Chemistry Course Discussion Forums. In *Proceedings of the 14th Learning Analytics and Knowledge Conference* (pp. 189-199).

Herring, S. (1996). Posting in a different voice: Gender and ethics in computer-mediated communication. *Philosophical perspectives on computer-mediated communication*, *115*, 45

Hesse, F., Care, E., Buder, J., Sassenberg, K., & Griffin, P. (2015). A framework for teachable collaborative problem solving skills. *Assessment and teaching of 21st century skills: Methods and approach*, 37-56.

Hughes, R., & Pace, C. R. (2003). Using NSSE to study student retention and withdrawal. *Assessment Update*, *15*(4), 1–2.

Hutto, C., & Gilbert, E. (2014, May). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media* (Vol. 8, No. 1, pp. 216–225).

Hilmy, S., De Silva, T., Pathirana, S., Kodagoda, N., & Suriyawansa, K. (2019, December). Moocs recommender based on user preference, learning styles and forum activity. In *2019 International Conference on Advancements in Computing* (ICAC) (pp. 180-185). IEEE.

Janssen, J., Erkens, G., Kirschner, P. A., & Kanselaar, G. (2009). Influence of group member familiarity on online collaborative learning. *Computers in Human Behavior*, *25*(1), 161-170.

Jacobi, T. and D. Schweers (2017): Justice, Interrupted: The Effect of Gender, Ideology, and Seniority at Supreme Court Oral Arguments, *Virginia Law Review*, *103*, 1379–1496.

Jiang, M., Li, Y., Zheng, J., & Han, X. (2017). Gender group differences on behavior patterns in collaborative problem solving through LEGO. *Journal of Computers in Education*, *4*, 127-145

Johnson, D.W., Johnson, R.T.: *Learning Together and Alone. Cooperative, Competitive, and Individualistic Learning*, 4th edn. Allyn and Bacon, Needham Heights (1994)

Johnson, D. W., Johnson, R. T., & Smith, K. (2007). The state of cooperative learning in postsecondary and professional settings. *Educational psychology review*, *19*, 15-29.

Johnson, G. M. (2015). On-campus and fully-online university students: Comparing demographics, digital technology use and learning characteristics. *Journal of University Teaching & Learning Practice*, *12*(1), 4.

Juavinett, A. L. (2022). The next generation of neuroscientists needs to learn how to code, and we need new ways to teach them. *Neuron*, *110*(4), 576-578.

Kim, J. Y., & Meister, A. (2023). Microaggressions, interrupted: The experience and effects of gender microaggressions for women in STEM. *Journal of Business Ethics*, *185*(3), 513-531

Kline, T. J. (2005). *Psychological testing: A practical approach to design and evaluation*. Sage publications.

Knöös, J., & Rääf, S. A. (2021). Sentiment Analysis of MOOC learner reviews: What motivates learners to complete a course?.

Kozlowski, D., Larivière, V., Sugimoto, C. R., & Monroe-White, T. (2022). Intersectional inequalities in science. *Proceedings of the National Academy of Sciences*, *119*(2), e2113067119.

Koh, E., & Lim, J. (2012). Using online collaboration applications for group assignments: The interplay between design and human characteristics. *Computers & Education*, *59*(2), 481–496. https://doi.org/10.1016/j.compedu.2012.02.002

Kube, D., Weidlich, J., Kreijns, K., & Drachsler, H. (2024). Addressing gender in STEM classrooms: The impact of gender bias on women scientists' experiences in higher education careers in Germany. *Education and Information Technologies*, 1-28.

Kumi-Yeboah, A., Kim, Y., Sallar, A. M., & Kiramba, L. K. (2020). Exploring the use of digital technologies from the perspective of diverse learners in online learning environments. *Online Learning*, *24*(4), 42-63.

Kucuktunc, O., Cambazoglu, B. B., Weber, I., & Ferhatosmanoglu, H. (2012, February). A large-scale sentiment analysis for Yahoo! answers. In *Proceedings of the fifth ACM international conference on Web search and data mining* (pp. 633-642).

Kurup, P. M., Yang, Y., Li, X., & Dong, Y. (2021). Interdisciplinary and integrated STEM. *Encyclopedia*, *1*(4), 1192-1199.

LaCosse, J., Canning, E. A., Bowman, N. A., Murphy, M. C., & Logel, C. (2020). A social-belonging intervention improves STEM outcomes for students who speak English as a second language. *Science advances*, *6*(40), eabb6543.

Lavy, V., & Sand, E. (2018). On the origins of gender gaps in human capital: Short-and long-term consequences of teachers' biases. *Journal of Public Economics*, *167*, 263-279.

Lewis, N. A., Sekaquaptewa, D., & Meadows, L. A. (2019). Modeling gender counter-stereotypic group behavior: A brief video intervention reduces participation gender gaps on STEM teams. *Social Psychology of Education*, *22*, 557-577.

Li, L. (2022). Reskilling and upskilling the future-ready workforce for industry 4.0 and beyond. *Information Systems Frontiers*, 1-16.

Light, A. E., Benson-Greenwald, T. M., & Diekman, A. B. (2022). Gender representation cues labels of hard and soft sciences. *Journal of Experimental Social Psychology*, *98*, 104234.

Lin, Y., Dowell, N., Godfrey, A., Choi, H., & Brooks, C. (2019, March). Modeling gender dynamics in intra and interpersonal interactions during online collaborative learning. In

Proceedings of the 9th international conference on learning analytics & knowledge (pp. 431-435).

Lin, Y. T., Wu, C. C., Chen, Z. H., & Ku, P. Y. (2020). How gender pairings affect collaborative problem solving in social-learning context. *Educational Technology & Society*, *23*(4), 30-44.

Lin, Y., & Nixon, N. (2024). Transitioning to Online Instructions and COVID-19 Response: A View from Mining Emergent College Students Discourse in Online Discussion Forum. *International Journal of Artificial Intelligence in Education*, 1-26.

Lin, Y., Yu, R., & Dowell, N. (2020). LIWCs the same, not the same: Gendered linguistic signals of performance and experience in online STEM courses. In *Artificial Intelligence in Education: 21st International Conference*, AIED 2020, Ifrane, Morocco, July 6–10, 2020, Proceedings, Part I 21 (pp. 333-345). Springer International Publishing.

Lin, Y., & Nixon, N. (2024, July). STEM Pathways in a Global Online Course: Are Male and Female Learners Motivated the Same?. In *Proceedings of the Eleventh ACM Conference on Learning@ Scale* (pp. 243-249).

Lowden, K., Hall, S., Elliot, D., & Lewin, J. (2011). Employers' perceptions of the employability skills of new graduates. *London: Edge Foundation*, *201126.*

Love, H. B., Stephens, A., Fosdick, B. K., Tofany, E., & Fisher, E. R. (2022). The impact of gender diversity on scientific research teams: a need to broaden and accelerate future research. *Humanities and Social Sciences Communications*, *9*(1), 1-12.

MacLaren, N. G., Yammarino, F. J., Dionne, S. D., Sayama, H., Mumford, M. D., Connelly, S., ... & Ruark, G. A. (2020). Testing the babble hypothesis: Speaking time predicts leader emergence in small groups. *The Leadership Quarterly*, *31*(5), 101409.

Maltese, A. V., & Tai, R. H. (2011). Pipeline persistence: Examining the association of educational experiences with earned degrees in STEM among US students. *Science Education*, *95*(5), 877-907. https://doi.org/10.1002/sce.20441

Martin, F., & Bolliger, D. U. (2018). Engagement matters: Student perceptions on the importance of engagement strategies in the online learning environment. *Online learning*, *22*(1), 205-222.

Mau, W. C., Chen, S. J., Li, J., & Johnson, E. (2020). Gender difference in STEM career aspiration and social-cognitive factors in collectivist and individualist cultures. *Administrative Issues Journal*, *10*(1), 4.

Means, B., Bakia, M., & Murphy, R. (2014). *Learning online: What research tells us about whether, when and how*. Routledge.

Miller, M. G., & Sutherland, J. L. (2023). The effect of gender on interruptions at congressional hearings. *American Political Science Review*, *117*(1), 103-121

Miguel, J., Andres, L., Mercedes, M., Rodrigo, T. and Sugay, J.O. 2014. An exploratory analysis of confusion among students using Newton's Playground. *Proceedings of the 22nd International Conference on Computers in Education*. (2014).

Murphy, M. C., Steele, C. M., & Gross, J. J. (2007). Signaling threat: How situational cues affect women in math, science, and engineering settings. *Psychological science*, *18*(10), 879-885.

Muskat, B., Anand, A., Contessotto, C., Tan, A. H. T., & Park, G. (2022). Team familiarity—Boon for routines, bane for innovation? A review and future research agenda. *Human Resource Management Review*, *32*(4), 100892.

National Science Foundation. (2023). *Diversity and STEM: Women, Minorities, and Persons with Disabilities*. https://ncses.nsf.gov/pubs/nsf23315/report

Nielsen, W., Chan, E. K., & Jahng, N. (2010). Collaborative learning in an online course: A comparison of communication patterns in small and whole group activities.

Nixon, N., Lin, Y., & Snow, L. (2024). Catalyzing Equity in STEM Teams: Harnessing Generative AI for Inclusion and Diversity. *Policy Insights from the Behavioral and Brain Sciences*, *11*(1), 85-92.

Ortiz-Martínez, G., Vázquez-Villegas, P., Ruiz-Cantisani, M. I., Delgado-Fabián, M., Conejo-Márquez, D. A., & Membrillo-Hernández, J. (2023). Analysis of the retention of women in higher education STEM programs. *Humanities and Social Sciences Communications*, *10*(1), 1-14.

Palomino, M. A., & Aider, F. (2022). Evaluating the effectiveness of text pre-processing in sentiment analysis. *Applied Sciences*, *12*(17), 8765.

Praharaj, S., Scheffel, M., Schmitz, M., Specht, M., & Drachsler, H. (2022, March). Towards collaborative convergence: quantifying collaboration quality with automated co-located collaboration analytics. In Lak22: *12th international learning analytics and knowledge conference* (pp. 358-369).

Prinsen, F. R., Volman, M. L., & Terwel, J. (2007). Gender-related differences in computer-mediated communication and computer-supported collaborative learning. *Journal of Computer Assisted Learning*, *23*(5), 393-409.

Rainey, K., Dancy, M., Mickelson, R., Stearns, E., & Moller, S. (2018). Race and gender differences in how sense of belonging influences decisions to major in STEM. *International Journal of STEM Education*, 5(1), 1-14.

Ramos, R. L., Alviña, K., & Martinez, L. R. (2017). Diversity of graduates from bachelor's, master's and doctoral degree neuroscience programs in the United States. *Journal of Undergraduate Neuroscience Education*, 16(1), A6.

Roberts, R., Denney, V., & Marion, J. W. (2024). Online academic group sentiment analyses informing workplace online group sentiment. *Development and Learning in Organizations: An International Journal*

Roscoe, R. D., Salehi, S., Nixon, N., Worsley, M., Piech, C., & Luckin, R. (2022). Inclusion and Equity as a Paradigm Shift for Artificial Intelligence in Education. In *Artificial Intelligence in STEM Education*. CRC Press.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of statistical software,* 48, 1-36.

Sachmpazidi, D., Olmstead, A., Thompson, A. N., Henderson, C., & Beach, A. (2021). Team-based instructional change in undergraduate STEM: characterizing effective faculty collaboration. *International Journal of STEM Education*, *8*, 1-23.

Sanger, C. S. (2020). Inclusive pedagogy and universal design approaches for diverse learning environments. *Diversity and inclusion in global higher education: Lessons from across Asia*, 31-71.

Sax, L. J., Kanny, M. A., Riggers-Piehl, T. A., Whang, H., & Paulson, L. N. (2015). "But I'm not good at math": The changing salience of mathematical self-concept in shaping women's and men's STEM aspirations. *Research in Higher Education*, *56*, 813-842.

Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of experimental social psychology*, *38*(2), 194-201.

Schmader, T., Johns, M., & Barquissau, M. (2004). The costs of accepting gender differences: The role of stereotype endorsement in women's experience in the math domain. *Sex roles*, *50*, 835-850.

Schmader, T. (2023). Gender inclusion and fit in STEM. *Annual Review of Psychology*, *74*(1), 219-243.

Schweingruber, H. A., Nielsen, N. R., & Singer, S. R. (Eds.). (2012). *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering*. National Academies Press.

Seeber, I., Bittner, E., Briggs, R. O., De Vreede, T., De Vreede, G. J., Elkins, A., ... & Söllner, M. (2020). Machines as teammates: A research agenda on AI in team collaboration. *Information & management*, *57*(2), 103174.

Shapiro, J. R., & Williams, A. M. (2012). The role of stereotype threats in undermining girls' and women's performance and interest in STEM fields. *Sex roles*, *66*, 175-183.

Singer, A., Montgomery, G., & Schmoll, S. (2020). How to foster the formation of STEM identity: studying diversity in an authentic learning environment. *International Journal of STEM Education*, *7*, 1-12.

Smith-Doerr, L., Alegria, S. N., & Sacco, T. (2017). How diversity matters in the US science and engineering workforce: A critical review considering integration in teams, fields, and organizational contexts. *Engaging Science, Technology, and Society*, *3*, 139-153.

Smith, E. (2011). Women into science and engineering? Gendered participation in higher education STEM subjects. *British Educational Research Journal*, *37*(6), 993-1014.

Soler-Gallart, M. (2017). *Achieving social impact: Sociology in the public sphere* (p. 112). Switzerland: Springer.

Song, L., Singleton, E. S., Hill, J. R., & Koh, M. H. (2004). Improving online learning: Student perceptions of useful and challenging characteristics. *The internet and higher education*, *7*(1), 59-70.

Sonnert, G., & Holton, G. J. (1995). *Who succeeds in science?: the gender dimension*. Rutgers University Press.

Stewart, A. E., Amon, M. J., Duran, N. D., & D'Mello, S. K. (2020, April). Beyond team makeup: Diversity in teams predicts valued outcomes in computer-mediated collaborations. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1-13).

Stout, J. G., Dasgupta, N., Hunsinger, M., & McManus, M. A. (2011). STEMing the tide: using ingroup experts to inoculate women's self-concept in science, technology, engineering, and mathematics (STEM). *Journal of personality and social psychology*, *100*(2), 255.

Stout, J., & Tamer, B. (2016, March). Collaborative learning eliminates the negative impact of gender stereotypes on women's self-concept. In *SIGCSE* (p. 496).

Strayhorn, T. L. (2011). *College Students' Sense of Belonging: A Key to Educational Success for All Students*. New York: Routledge.

Sun, C., Shute, V. J., Stewart, A. E., Beck-White, Q., Reinhardt, C. R., Zhou, G., ... & D'Mello, S. K. (2022). The relationship between collaborative problem solving behaviors and solution outcomes in a game-based learning environment. *Computers in Human Behavior*, *128*, 107120.

Sun, A., & Chen, X. (2016). Online education and its effective practice: A research review. *Journal of Information Technology Education: Research*, *15*, 157-190.

Tannen, D. (1990). Gender differences in topical coherence: Creating involvement in best friends' talk. *Discourse processes*, *13*(1), 73-90.

Takeda, S., & Homberg, F. (2014). The effects of gender on group work process and achievement: an analysis through self-and peer-assessment. *British Educational Research Journal*, *40*(2), 373-396.

Tomai, M., Mebane, M. E., Rosa, V., & Benedetti, M. (2014). Can computer supported collaborative learning (CSCL) promote counter-stereotypical gender communication styles in male and female university students?. *Procedia-Social and Behavioral Sciences*, *116*, 4384-4392.

Van Veelen, R., Derks, B., & Endedijk, M. D. (2019). Double trouble: How being outnumbered and negatively stereotyped threatens career outcomes of women in STEM. *Frontiers in psychology*, *10*, 150.

van Viegen, T., Akrami, A., Bonnen, K., DeWitt, E., Hyafil, A., Ledmyr, H., ... & Peters, M. A. (2021). Neuromatch Academy: Teaching computational neuroscience with global accessibility. *Trends in cognitive sciences*, *25*(7), 535-538.

von Davier, A. A., & Halpin, P. F. (2013). Collaborative Problem Solving and the Assessment of Cognitive Skills: Psychometric Considerations. *ETS Research Report Series*, *2013*(2), i–36. https://doi.org/10.1002/j.2333-8504.2013.tb02348.x

West, C., & Zimmerman, D. H. (1987). Doing Gender. *Gender and Society*, *1*(2), 125–151. http://www.jstor.org/stable/189945

Wang, M. T., & Degol, J. L. (2017). Gender gap in science, technology, engineering, and mathematics (STEM): Current knowledge, implications for practice, policy, and future directions. Educa*tional psychology review*, *29*, 119-140.

Whitcomb, K. M., Maries, A., & Singh, C. (2023). Progression in self-efficacy, interest, identity, sense of belonging, perceived recognition and effectiveness of peer interaction of physics majors and comparison with non-majors and Ph. D. students. *Research in Science Education*, *53*(3), 525-539.

Wilson, Z. S., Holmes, L., Sylvain, M. R., Batiste, L., Johnson, M., McGuire, S. Y., Pang, S. S., & Warner, I. M. (2015). Hierarchical Mentoring: A Transformative Strategy for Improving Diversity and Retention in Undergraduate STEM Disciplines. *Journal of Science Education and Technology*, *21*(1), 148-156. https://doi.org/10.1007/s10956-011-9292-5

Williams, J. C., Phillips, K. W., & Hall, E. V. (2016). Tools for change: Boosting the retention of women in the STEM pipeline. *Journal of Research in Gender Studies*, *6*(1), 11.

Williams, J., & Roth, W. M. (2019). Theoretical perspectives on interdisciplinary mathematics education. *Interdisciplinary mathematics education: The state of the art and beyond*, 13-34.

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *science*, *330*(6004), 686-688.

Yuan, C. W., Liu, Y. H., Wang, H. C., & Tseng, Y. C. (2019, May). Gender effects on collaborative online brainstorming teamwork. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-6).

Zhou, J., & Ye, J. M. (2023). Sentiment analysis in education research: a review of journal publications. *Interactive learning environments*, *31*(3), 1252-1264.

Appendix A: Inventory of Psychological Measures - Sense of Belonging, STEM Identity, and Self-Efficacy

Sense of Belonging - Field Focused
Scale: Strongly Disagree = 1, Strongly Agree = 7

Item 1	I feel like I really belong in my field of study.
Item 2	I feel accepted by other students in my field of study.
Item 3	I feel accepted by the instructors in my field of study.
Item 4	I regret choosing my field of study (reverse scoring).
Item 5	I feel like I have a lot in common with other students in my field of study.

Cronbach alpha = .85
Source: Cheryan and colleagues (2009), Lewis et al. (2019)

STEM Identity
Scale: Strongly Disagree = 1, Strongly Agree = 7

Item 1	Being a student in my field of study is an important part of my self-image.
Item 2	Being a student in my field of study is unimportant to my sense of what kind of person I am.
Item 3	Being a student in my field of study has very little to do with how I feel about myself.
Item 4	Being a student in my field of study is an important reflection of who I am.

Cronbach alpha = .83
Source: Schmader (2002), Lewis et al. (2019)

Self-Efficacy

Scale: a slider from 1 to 100 was used. In the analysis, this is rescaled to 1-10.

Item 1	finding strategies to succeed in class?
Item 2	finding a place to study without distractions?
Item 3	thinking of many different ways to earn a good grade?
Item 4	motivating yourself to do schoolwork?
Item 5	doing research?
Item 6	studying even when there are other things to do?
Item 7	taking notes?
Item 8	finishing your homework assignments by deadlines?
Item 9	scheduling your time to accomplish your tasks?
Item 10	thinking of specific ways to do well in your class?

Cronbach alpha = .87

Source: College Student Self Efficacy Survey—CSSAS in Farr et al., (2011)

STEM Persistence Intent

Scale: Strongly Disagree = 1, Strongly Agree = 7

Item 1	How likely is it that you will pursue graduate study related to your major?
Item 2	How likely is it that your eventual career after graduation will directly pertain to mathematics or science?
Item 3	How often do you think about changing your major?
Item 4	How likely is it that you will change your major?

Source: Schmader, T., Johns, M., & Barquissau, M. (2004).

Sense of Belonging - Group Focused
Scale: Not at all true = 1, Completely true = 5

Item 1	I feel like a real part of my pod.
Item 2	People in my pod notice when I'm good at something.
Item 3	It is hard for people like me to be accepted here (in my pod or at NMA - Deep Learning in general).
Item 4	Other students in my pod take my opinions seriously.

Cronbach alpha = .91
Source: College Student Self Efficacy Survey—CSSAS in Farr et al., (2011)

Self-Efficacy - Field Focused
Scale: Strongly Disagree = 1, Strongly Agree = 7

Item 1	When I get new material in my field of study, I'm usually sure I will be able to learn it.
Item 2	I sometimes doubt my ability to perform well in my field of study.
Item 3	I am good at my field of study compared to other people in my major.
Item 4	My field of study has always come pretty easy for me.
Item 5	My field of study has never been easy for me.
Item 6	I feel like I have to work harder than other people in my field of study to do well.

Cronbach alpha = .76
Source: Lewis et al. (2019)

Appendix B. Knowledge Test Items

Item 1	Linear Algebra: What is the importance of eigenvalues and eigenvectors of a matrix?
Item 2	Calculus: What is the solution to this equation: dx/dt = -a*x +b
Item 3	Statistics: What are the defining characteristics of the Gaussian distribution?
Item 4	Model types: What can models do?
Item 5	Modeling practice: How would you start a modeling project?
Item 6	Model fitting: What's the least-biased method of evaluating the performance of a data-generated model?
Item 7	GLM: Generalized linear models allow you to:
Item 8	Dimensionality reduction: Dimensionality reduction is
Item 9	Reinforcement learning: In updating the value function (V := V + a*d), if "d" equals the error, then the parameter “a” represents the:
Item 10	Deep Learning: In an auto-encoder architecture, the relationship between the number of inputs and outputs is:
Item 11	Bayes: Bayes Rule states that P(y|x) = P(x|y)*P(y), the term P(x|y) in this case is often called:
Item 12	Real Neuron Models: In a Leaky Integrate and Fire model, a high membrane time constant would indicate that for a constant input current the neuron has:

Appendix C. Full list of country of origin in Neuromatch

	Country	Student		Country	Student
1	China	231	45	South Africa	3
2	India	151	46	Switzerland	3
3	United States	94	47	Tunisia	3
4	Iran	55	48	Algeria	2
5	Germany	50	49	Bolivia	2
6	Mexico	40	50	Cameroon	2
7	Turkey	40	51	Croatia	2
8	Egypt	35	52	Cyprus	2
9	Canada	30	53	Indonesia	2
10	Russia	26	54	Serbia	2
11	Italy	25	55	Sri Lanka	2
12	Brazil	23	56	Sweden	2
13	United Kingdom	23	57	Thailand	2
14	Nigeria	18	58	Uganda	2
15	South Korea	17	59	Ukraine	2
16	France	15	60	Venezuela	2
17	Argentina	12	61	Afghanistan	1
18	Spain	12	62	Bangladesh	1
19	Australia	11	63	Belarus	1
20	Chile	9	64	Benin	1
21	Peru	9	65	Curacao	1
22	Poland	9	66	Democratic Republic of the Congo	1
23	Ireland	8	67	Denmark	1
24	Colombia	7	68	Dominican Republic	1
25	Japan	7	69	Ecuador	1
26	Greece	6	70	Estonia	1
27	Hungary	6	71	Guatemala	1
28	Israel	6	72	Honduras	1
29	Netherlands	6	73	Iraq	1
30	Taiwan	6	74	Jordan	1
31	Ghana	5	75	Kyrgyzstan	1
32	Lebanon	5	76	Latvia	1
33	Nepal	5	77	Madagascar	1
34	Pakistan	5	78	Malaysia	1
35	Portugal	5	79	New Zealand	1
36	Hong Kong	4	80	Palestine	1
37	Sudan	4	81	Puerto Rico	1
38	Vietnam	4	82	Saudi Arabia	1
39	Austria	3	83	Sierra Leone	1
40	Azerbaijan	3	84	Singapore	1
41	Bulgaria	3	85	Slovenia	1
42	Czechia	3	86	Uruguay	1
43	Philippines	3	87	Uzbekistan	1
44	Romania	3	88	Yemen	1

Appendix D. Codebook for student feedback

Main code	Subcodes
Skill and Knowledge Disparities (pertinent to group formation)	Positive feelings about student's own knowledge and skills
	Negative feelings about student's own knowledge and skills
	Positive feelings about other students knowledge and skills
	Negative feelings about other students knowledge and skills
	Composition of group relative to prior knowledge levels
Engagement and Participation	Engagement with own group
	Engagement with other groups
	Teaching Strategies / Engagement with Course Content and Materials
Other barriers	Workload
	Language
	Shared Interest
	TA Expertise

Appendix E. Sample CPS Transcript

teamname	block	speaker	timestamp	content
ASU-T102	Block1	PD	05:21:00	It's still moving.
ASU-T102	Block1	PB	05:22:00	I think you need to be heavier and you need to like, put it like really far above it. 'Cause it needs to like, do a, like almost a full flip.
ASU-T102	Block1	PD	05:29:00	Mm-hmm (affirmative). It's trying hard, yeah. Here, I'm gonna do a little.
ASU-T102	Block1	PC	05:37:00	Oh, drop a heavy weight on top. Drop like a really heavy weight, just on top of that .Just shove it into the end. Oh. Wait, okay. Okay. Heavier. Go heavier, just keep dropping weights.
ASU-T102	Block1	PB	05:50:00	(laughs).
ASU-T102	Block1	PD	05:50:00	Oh, it's right on the edge.
ASU-T102	Block1	PC	05:56:00	It's, it's stuck on something, yeah.
ASU-T102	Block1	PD	06:00:00	Here, watch this. Nope.
ASU-T102	Block1	PC	06:03:00	It just needs to be heavier, like it, we just need to lift up the whole other side, I think. Delete the, delete the, all that stuff, just get it out of the way so it's not, it doesn't hinder it. There you go. And then just drop, yeah, just drop like a really heavy weight.
ASU-T102	Block1	PC	06:21:00	Hm.
ASU-T102	Block1	PB	06:21:00	Hm.
ASU-T102	Block1	PD	06:32:00	That doesn't do anything.
ASU-T102	Block1	PC	06:34:00	Could you-
ASU-T102	Block1	PB	06:35:00	Oh.
ASU-T102	Block1	PC	06:35:00	... make a spring that would push it back, just, just like, a centimeter-
ASU-T102	Block1	PD	06:39:00	Uh, let me try.
ASU-T102	Block1	PC	06:40:00	... somehow? Yeah, like put one.
ASU-T102	Block1	PD	06:43:00	Oh, I got an idea.
ASU-T102	Block1	PC	06:43:00	Mm-hmm (affirmative). Pushed it. Getting there.
ASU-T102	Block1	PB	06:55:00	I mean, if you like, did a pin on like, the little place where the ball is and put a really heavy weight, like on the right side of the pit, if you put [crosstalk 00:07:08] a pin and put heavy weights under it, then it might, like, go down.
ASU-T102	Block1	PC	07:11:00	I think that might have just screwed us up pretty big time, there.
ASU-T102	Block1	PD	07:13:00	Yeah (laughs).
ASU-T102	Block1	PB	07:14:00	Yeah.
ASU-T102	Block1	PC	07:16:00	Yeah, go, yeah, do the initial idea, and then go way heavier. Actually, no. Wait, could you, could, I'm wondering if you could do the same idea that you were talking about, so instead of dropping a weight, you pull it down, so like, draw a pin on the very edge, even further edge [crosstalk 00:07:34].
ASU-T102	Block1	PD	07:35:00	Or this one?

Appendix F. Exclusion criteria for text processing

Variable	Annotation exclusion	Additional excluded words
Word count	Annotation on crosstalk and laughter	Filler words (e.g. "hm", "oof", "um", "uh", "uhm", "ah", "oh", "er", "hmm", “huh”)
Sentiment	Annotation on crosstalk	

Annotation example:

- Crosstalk: “You guys ready? [crosstalk 00:00:03]. Let's do…”
- Laughter: “(laughs) Um... Uh, try moving it forward so…”