Title
Designing Conversational Agents to Promote Collaboration and Systems Thinking in High School Science Discussion

Author
Nguyen, Thi Cam Ha

Publication Date
2022

UNIVERSITY OF CALIFORNIA,
IRVINE

Designing Conversational Agents to Promote Collaboration and Systems Thinking in High School Science Discussion

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Education

by

Ha Nguyen

Dissertation Committee:
Associate Professor June Ahn, Co-chair
Professor Rossella Santagata, Co-chair
Professor Mark Warschauer
Associate Professor Amy Ogan
Associate Professor Cascade Sorte

2022

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

ACKNOWLEDGEMENTS

The Latin root of education is *educere*—to draw out something meaningful in students.

These words from sophomore year's Educational Psychology class have stayed with me since. Education is not the recitation of facts; its goal is to facilitate meaningful interactions. I want to acknowledge Dr. David Malone, who taught me the root of education and encouraged me to pursue a Ph.D. Your words will continue to guide my work.

I am truly grateful for my advisors, Drs. June Ahn, Rossella Santagata, and Mark Warschauer, for their mentorship and unwavering support for my growth as a person and a scholar. Thank you for pushing me to think critically and giving me the opportunities to build partnerships, conduct research, and share my work. June, among many things you have taught me, I have learned so much about how to conceptualize research, lay the groundwork for research partnerships, design and tell stories with data, and write effectively. Rossella, thank you for modeling to me how to be intentional and respectful in our work with educators and students. I will always remember your warm smiles and wisdom. Mark, thank you for seeing the potential in me all those years ago, and for encouraging me to stay curious and open to scholarship.

I would like to thank my committee members, Drs. Amy Ogan and Cascade Sorte, for being so generous with their time. Your feedback on the conceptualization, analyses, and writing of this dissertation has really strengthened this work.

I want to acknowledge my community partners—Sara Ludovise, Holly Fletcher, Erick Valdez, and Kaitlin Magliano at Crystal Cove Conservancy, for supporting my ideas and showing me models of sustainable, thoughtful educational outreach. I thank Mrs. Stephanie Vasconcellos, Daryth Morrissey, and their students, for letting me pilot this work in their classes, amid all the chaos of a global pandemic. This work wouldn't have been possible without their support.

I want to thank my research assistants and thought partners, Laura Wang and Josiah Huse, for helping me with data cleaning, coding, and facilitation of co-design meetings. It has been a pleasure working closely with you for the last 2.5 years.

I thank Wiley, ISLS, and Springer for permission to include Chapters 2, 3, and 4 of my dissertation, which were originally published in BJET, ISLS'22, and *Communications in Computer and Information Science, vol 1522*. Funding for this work was provided by UCI's Center for Research on Teacher Development and Professional Practice.

I want to express gratitude to my peers at the Design and Partnership Lab, Digital Learning Lab, the Teaching and Learning research group, and the Education Ph.D. program. The Ph.D. would have been a lot less fun and more challenging without you all.

I would not have learned that research could be a career, without the opportunities I received as an undergraduate. I want to thank the University Scholars Program for funding my education, and Dr. Victoria Lodewick and my mentors in the program for many enriching dialogues about interdisciplinary research. Shoutouts to Dr. Amy Anderson, for showing me how much impact a caring mentor and educator can have.

To my family, recently acquired second family, and friends: Thank you for the constant support and love, even though I sometimes have troubles explaining what *exactly* my research is about.

Finally, I couldn't have done this work without my partner William, who is the springboard for many of my research ideas. Thank you for being my anchor of love, humor, and resilience.

VITA

Ha Nguyen

EDUCATION

2022	University of California-Irvine, PhD, Education. Dissertation Study: *Designing Conversational Agents to Promote Collaboration and Systems Thinking in High School Science Discussions.* Committee: Dr. June Ahn (co-chair), Dr. Rossella Santagata (co-chair), Dr. Mark Warschauer, Dr. Amy Ogan, Dr. Cascade Sorte.
2020	University of California-Irvine, Master's Degree, Education.
2018	Duke University, Bachelor's Degree, Public Policy, Japanese, Minor in Education.

GRANTS, HONORS & AWARDS

2021	**PI**. Designing a Conversational Agent in Collaborative Discussion for Systems Thinking. Center for Teacher Development and Professional Practice; $2000.
2021	**Recipient,** Michael E. Martinez Prize for Outstanding Educational Research and Service
2021	**Best Paper Nominee,** International Conference of Quantitative Ethnography.
2020	**Honorable Mention for Best Paper**, Conference on Educational Data Science.
2018-2023	**Provost Fellowship**, UC Irvine.
2018	**Dean's Recruitment Fellowship,** UC Irvine
2018	**Phi Beta Kappa Honors Society**
2018	**Highest Distinction in Public Policy** and **Holton Prize for Innovative Work in Education** for senior thesis.
2014-2018	**University Scholars**, A full, four-year merit scholarship for students with potential to research interdisciplinary interests.

SELECTED PEER-REVIEWED PUBLICATIONS

2022

Nguyen, H. (2022). Let's Teach Kibot: Discovering Discussion Patterns between Student Groups and Two Conversational Agent Designs. *British Journal of Educational Technology.* http://doi.org/10.1111/bjet.13219

Nguyen, H. (2022). "Learners' Reactions to Chatbot Communication Breakdowns: Insights into Fostering Learning". In *2nd Annual Meeting of the International Society of the Learning Sciences (Online)*. International Society of the Learning Sciences. (accepted).

Nguyen, H. & Young, W. (2022). Knowledge construction and uncertainty in real world argumentation: A text analysis approach. In *The 12th International Conference on Learning Analytics and Knowledge (LAK22)*.

Jacob, S., Jonathan, M., **Nguyen, H.**, Richardson, D., & Warschauer, M. (2022) Examining the what, why, and how of multilingual student identity development in computer science. *ACM Transactions on Computing Education.*

2021

Nguyen, H. (2021). Exploring Group Discussion with Conversational Agents Using Epistemic Network Analysis. In *International Conference on Quantitative Ethnography* (pp. 378-394). Springer, Cham.

Nguyen, H., Lim, K.Y., Fischer, C., Wu, L., Washington, G., & Warschauer, M. (2021). Peer support and temporality in regulated learning. *Learning & Instruction. https://doi.org/10.1016/j.learninstruc.2021.101443*

Nguyen, H., & Santagata, R. Impact of computer modeling on learning and teaching systems thinking. (2021). *Journal of Research in Science Teaching*. https://onlinelibrary.wiley.com/doi/10.1002/tea.21674

Nguyen, H., Ahn, J., Belgrave, A., Lee, J., Cawelti, L., Kim, H.E., Prado, Y., Santagata, R., & Villavicencio, A. (2021) Establishing trustworthiness through algorithmic approaches to qualitative research. In *Second International Conference on Quantitative Ethnography*. Springer, Cham. https://doi.org/10.1007/978-3-030-67788-6_4

Ahn, J., **Nguyen, H.**, Campos, F., & Young, W. (2021). Transforming everyday information into practical analytics with crowdsourced assessment tasks. In *The 11th International Conference on Learning Analytics and Knowledge (LAK21)*. https://doi.org/10.1145/3448139.3448146

Ahn, J., Campos, F., **Nguyen, H.**, Hays, M., & Morrison, J. (2021). Co-designing for privacy, transparency, and trust in K-12 learning analytics. In *The 11th International Conference on Learning Analytics and Knowledge (LAK21)*. https://doi.org/10.1145/3448139.3448145

Ahn, J., **Nguyen, H.**, & Campos, F. From visible to understandable: Designing for teacher agency in education data visualizations. (2021). *Contemporary Issues in Technology & Teacher Education (CITE)*. https://citejournal.org/volume-21/issue-1-21/general/from-visible-to-understandable-designing-for-teacher-agency-in-education-data-visualizations/

Campos, F., Ahn, J., Digiacomo, D., **Nguyen, H.**, & Hays, M. One chart, many meanings: Making sense of sensemaking in learning analytics dashboard design. (2021). *Journal of Learning Analytics*. https://www.learning-analytics.info/index.php/JLA/article/view/7113

Santagata, R., König, J., Scheiner, T., **Nguyen, H.**, Adleff, A.-K., Yang, X., & Kaiser, G. Mathematics teacher learning to notice: a systematic review of studies of video-supported teacher education. (2021). *ZDM, International Journal of Mathematics Education*. https://doi.org/10.1007/s11858-020-01216-z

2020 **Nguyen, H.**, Wu, L., Fischer, C., Washington, G., & Warschauer, M. (2020). Increasing success in college: Examining the impact of a project-based introductory engineering course. *Journal of Engineering Education.* doi.org/10.1002/jee.20319

Nguyen, H. & Jenkins, J. (2020). In or out of sync: Federal funding and research in early childhood. *AERA Open 6(4)*. doi/10.1177/2332858420979568

Nguyen, H., Wu, L., Washington, G., Lim, K.Y., & Fischer, C. (2020). Collaboration patterns and design practices in first-year project-based engineering. In *Proceedings of the 2020 American Society for Engineering Education Annual Conference & Exposition.*

Nguyen, H., Lim, K.Y., Wu, L. Fischer, C., & Warschauer, M. (2020). “I thought we said”: Perceived peer support, discourse cohesion, and regulation in engineering design. In *14th International Conference of the Learning Sciences: Interdisciplinarity of the Learning Sciences, ICLS 2020.* International Society of the Learning Sciences (ISLS).

Nguyen, H., Ahn, J., Young, W., & Campos, F. (2020). Where's the learning in education crowdsourcing? In *Proceedings of the Seventh (2020) Annual ACM Conference on Learning@ Scale*. doi.org/10.1145/3386527.3406734

Nguyen, H., Garcia, L., Jacob, S., Richardson, D., & Warschauer, M. (2020). Reflection as formative assessment of computational thinking in elementary grades. In *14th International Conference of the Learning Sciences: Interdisciplinarity of the Learning Sciences, ICLS 2020.* International Society of the Learning Sciences (ISLS).

Nguyen, H., Garcia, L., Jacob, S., Richardson, D., & Warschauer, M. (2020). Elementary teachers' use of video reflections to reinforce computer science language and concepts. In *Proceedings of the Research on Equity and Sustained*

Participation in Engineering, Computing, and Technology, IEEE Computer Society.

Zhou, N., **Nguyen, H.**, Fischer, C., Richardson, D., & Warschauer, M. (2020) Hybrid professional development program to promote high school teachers' self-efficacy in computer science classroom. *ACM Transactions on Computing Education.* doi/10.1145/3410631

Jacob, S., **Nguyen, H.**, Garcia, L., Richardson, D., & Warschauer, M. (2020). Teaching computational thinking to multilingual students through inquiry-based learning. In *Proceedings of the Research on Equity and Sustained Participation in Engineering, Computing, and Technology,* IEEE Computer Society.

2018 Jacob, S., **Nguyen, H.**, Tofel-Grehl, C., Richardson, D., & Warschauer, M. (2018). Teaching computational thinking to English learners. *NYS TESOL Journal*, 5(2), 12-24.

BOOK CHAPTERS

2021 **Nguyen, H.**, Campos, F., & Ahn, J. (2021). Designing for generative uncertainty in learning dashboards. In Ifenthaler, D., & Muhittin, S. (Eds), *Visualizations and Dashboards for Learning Analytics*.

Nguyen, H., Campos, F., & Ahn, J. (2021). Expanding the design space of data and action in education: What co-designing with educators reveal about current possibilities and limitations. In Bowers, A. (Ed), *Data Visualization, Dashboards, and Evidence Use in Schools: Data Collaborative Workshop Perspectives of Educators, Researchers, and Data Scientists*. Teachers College, Columbia University. New York, NY. https://doi.org/10.7916/d8-jj2g-e225

SELECTED CONFERENCE PRESENTATIONS (Undergraduate mentees are underlined)

2022 **Nguyen, H.** (2022, June). "Exploring Students' Identities in Co-designing Chatbots". Poster presented at *The Annual Meeting of the International Society of the Learning Sciences*.

Nguyen, H. (2022, April). "Looks like robots, sounds like humans": Surveying students' conceptualizations of learning agents. Poster presented at *AERA Annual Meeting 2022*.

2021 **Nguyen, H.**, Lim, K.Y., Fischer, C., & Wu, L. (2021, June). Using relational event modeling to capture shared regulation interactions in collaborative learning. Poster presented at *The Annual Meeting of the International Society of the Learning Sciences*.

Rosenberg, J., & **Nguyen, H.** (2021, April). How K-12 school districts communicated during the COVID-19 pandemic: A study using Facebook data. Poster presented at *The 11th International Conference on Learning Analytics and Knowledge (LAK21)*

Nguyen, H., Ludovise, S., Wang, J., Huse, J., & Santagata, R. (2021, April). Modeling tools and systems thinking patterns in middle school. Paper presented at *AERA Annual Meeting 2021*.

2020 **Nguyen, H.**, Schmidt, D., Santagata, R.(2020, November). Crystal Code: Examining the impact of computational modeling on scientific systems thinking. Poster presented at *International Society for Technology in Education (ISTE)*. Anaheim, CA.

Nguyen, H. (2020, September). In or out of sync: Funding in early childhood through text analytics. Paper presented at *Conference on Educational Data Science*, Stanford, CA. **[Honorable Mention for Best Paper]**

TEACHING & MENTORING

Course **EDUC 10: Educational Research Design (**TA). Enrollment: 125 students.

ABSTRACT OF THE DISSERTATION

Designing Conversational Agents to Promote Collaboration and Systems Thinking in High School Science Discussion

by

Ha Nguyen

Doctor of Philosophy in Education

University of California, Irvine, 2022

Associate Professor, June Ahn, Co-chair

Professor Rossella Santagata, Co-chair

This dissertation consists of four studies that explore how high school students interact with and learn from two designs of a text-based conversational agent (chatbot) during small-group science discussions with peers and the agents. The agents utilize natural language processing to send prompts to promote students' understanding of ecosystems concepts and collaboration. The two agent designs differ in appearances and linguistic styles to resemble a less knowledgeable peer and an expert. The first two studies explore how students **interacted with the agents as social partners**. In Study 1, I found that interactions with the less-knowledgeable-peer agent generally contained more sequences of questioning and transactive exchange, which provided opportunities for reflection and reasoning. In Study 2, I examined how student groups attempted to remedy interactions with the agents in case the agents failed to interpret students' intent, similar to how students would repair conversations with human partners. Student groups more often reframed and explained their reasoning to the less-knowledgeable-peer than the expert agent. Furthermore, such interactions were positively correlated with higher counts of

systems thinking statements, which indicated an enhanced understanding of interconnections between human and natural systems. Study 3 presents a case study to illustrate how **interactions with the agents varied with group compositions**: emergent, mixed, and expanding prior science knowledge. In Study 4, I built on studies 1-3 to examine **how students learned from the agents**, using interaction patterns with the agents as mediators. Results confirm the affordances of certain interaction dynamics, such as transactive exchange, to deepen systems understanding. Overall, the studies provide converging evidence on the utility of agent designs to support interactions that are enriching for learning. I discuss implications for designing conversational agents as social partners to promote collaboration and systems thinking, with applications to other learning contexts.

CHAPTER 1 INTRODUCTION

Classroom discussion plays a central role in the production and exchange of disciplinary knowledge. Supporting students to articulate their understanding of concepts, reflect on learning, and build on one another's ideas in group discussions can deepen students' conceptual understanding and foster academically productive talk (Dyke, Adamson, et al., 2013; Dyke, Howley, et al., 2013). However, in small-group settings, teachers may not always have the time, resources, or expertise to facilitate discussion in every group.

A potential solution to this challenge is the integration of conversational agents into individual and group settings to facilitate learning interactions (Adamson et al., 2014; Diziol et al., 2010; Dyke, Adamson, et al., 2013; Dyke, Howley, et al., 2013; Walker et al., 2014b). Conversational agents are dialogue systems that provide learning support for students through natural language understanding (Kerly et al., 2008). Graesser and colleagues (2016), for example, have developed agent systems (e.g., AutoTutor) that included different arrangements to provide knowledge scaffolds for learners in subjects spanning reading, writing, and science. These arrangements included one-on-one dialogues between an agent and a student or trialogues between two agents and a human student. In other systems, agents can provide suggestions for idea generation in group discussions or guide peer tutors to reason through how to support their human tutees (Kumar et al., 2011; Tegos & Demetriadis, 2017; Walker et al., 2011).

Researchers have explored different agent designs, for example, as a less knowledgeable tutee or a mentor, to promote interactions conducive to deeper learning (Biswas et al., 2010; Graesser, 2016). These designs build on the Human-Computer Interaction research that users subconsciously treat computer agents as social actors (Nass et al., 1995), and may thus interact differently with agent designs (Kim, 2007; Liew et al., 2013; Rosenberg-Kima et al., 2008). For example, students with emergent knowledge may be more responsive to scaffolds from an expert-like agent, and consequently learn more from the interactions (Graesser, 2016).

However, there has been little consideration of agent designs in group settings, particularly regarding how such designs may influence the content of and participation in group discussions. Unlike individual interactions with an agent, student groups might ignore or abuse the agent (Kumar et al., 2011). Thus, in this dissertation, I examine **how different designs of a text-based agent (chatbot) can facilitate student-agent interactions and learning in group settings**. The agent focuses on providing conceptual and participatory prompts to help students develop systems thinking, or understanding of complex relationships within a marine ecosystem.

Study Context

The design focus of the conversational agents (to facilitate systems thinking and collaboration) is grounded in a multi-year partnership between a local state park, education researchers, biology researchers, and local school districts in Southern California. The agent was integrated into one of the state park's high school environmental science programs, called Marine Science Exploration (MSE). During the 2018-2019 school year, this program served 2,600 youth, with over 70% of participants from Title 1 schools and 46% of students reporting speaking a language other than English at home.

The MSE program consisted of eight lessons to engage students in a participatory science curriculum to learn about systems concepts. Participatory science anchors scientific concepts in local contexts to enrich student interest in authentic, student-driven education (McKinley et al., 2017). The MSE program grounded its curriculum in protecting the biodiversity in the state park's marine conservation area. In the past century, natural events and human actions have threatened the ocean habitats with climate change, pollution, and overfishing, among others. The park thus created the marine protected area to reduce the impacts from those threats. In the MSE program, students explored whether the existing regulations in the park's marine conservation area helped to increase the park's biodiversity. Several scientific practices were integrated throughout the curriculum. Students first developed relational models showing how different factors impacted species diversity, engaged in data collection and analysis, and presented their findings to the state park stakeholders. Students applied these practices in an authentic context and developed their science identities in the process (Brown et al., 2005).

The program design intentionally created opportunities for group discussions around making sense of data. Written and spoken discourse is an important channel to construct science literacy. As individuals interact, they draw on existing knowledge and social repertoires to assert and develop identities (Bransford et al., 2000). For example, early in the curriculum (lesson 3), students brainstormed the elements and processes that might affect the marine ecosystem after learning about the marine conservation area. In classroom observations of those lessons, however, the research team noticed that not all students participated equally in the discussions. This participation pattern could be due to students' varying levels of domain knowledge (Hogan et al., 1999). Furthermore, students tended to get fixated on certain ideas and linear relationships (e.g., overfishing reduces fish), instead of thinking about more complex processes in the system.

The idea of a conversational agent came to me as a solution to improve students' discussion quality. The agent can embed itself in student discussions and provide conceptual and participation nudges as students collaboratively build a concept map of the marine ecosystem. Drawing from prior work with one-on-one interactions between students and an agent (Biswas et al., 2010, 2016; Kim & Baylor, 2016), I test two designs: a less knowledgeable peer and an expert agent. The agents send similar prompts to promote systems thinking and collaboration but differ in appearances and linguistic styles. The dissertation investigates how students interact with and learn from the agents, and how these interactions may vary between student groups.

I draw from research in human-computer interactions, knowledge construction, and systems thinking to inform agent designs and analytical decisions. The agent designs stem from frameworks of Computers as Social Actors (Nass et al., 1995) to assume that humans subconsciously treat computer systems as human actors based on design cues. Thus, designs of the agents as a less knowledgeable peer versus an expert can facilitate different interactions, as if students were interacting with a human counterpart. Drawing from knowledge construction frameworks (Scardamalia & Bereiter, 1991; van Aalst, 2009), I break down interactions with the agents into their content and students' participation patterns. This is because how people interact with others in discussions contributes to knowledge building efforts and helps group knowledge become more structured over time. Finally, I hypothesize that the different interaction patterns and content with the agents support learning in different ways. For this, I draw from systems thinking frameworks (Hmelo-Silver et al., 2017; Snapir et al., 2017) to analyze students' learning, as indicated by group discussions and learning outcomes following agent interactions. These theoretical frameworks—human-agent interactions, discussion moves in knowledge construction, and systems thinking—are common threads throughout the dissertation studies.

Dissertation Overview

This dissertation includes six chapters: Introduction, chapters reporting each of the four studies, and Conclusion. I embed the Literature reviews and Methods sections within each chapter. I explore the following questions:

1. how student groups interact with different agent designs (Studies 1, 2)

2. how groups' interactions vary with different group compositions, i.e., students with emergent, mixed, and expanding prior domain knowledge (Study 3)
3. how agent designs promote learning for individual students, accounting for students' prior domain knowledge (Study 4)

Table 1.1
Overview of Dissertation Studies

	Areas of investigation	Sample	Analytic procedures	Status
Study 1	Groups' discussion moves: • Reasoning • Transactive exchange • Responsiveness to agents	*n* = 52 (18 groups) Within-subject design	Sequential pattern mining; Wilcoxon signed-rank test	Revised from **Nguyen, H.** (2022). Let's Teach Kibot: Discovering Discussion Patterns between Student Groups and Two Conversational Agent Designs. *British Journal of Educational Technology.* http://doi.org/10.1111/bjet.13219
Study 2	Groups' discussion moves: • Conversation repair strategies (e.g., reframe, repeat, explain) Groups' systems thinking, as indicated by chat logs	*n* = 52 (18 groups) Within-subject design	Wilcoxon signed-rank test; Pearson's correlation	Revised from **Nguyen, H.** (June, 2022). Learners' Reactions to Chatbot Communication Breakdowns: Insights into Fostering Learning. In *2nd Annual Meeting of the International Society of the Learning Sciences.* International Society of the Learning Sciences.
Study 3	Case study of groups' discussion: • Reasoning • Transactive exchange • Responsiveness to agents Group compositions, as indicated by systems thinking pre-test	*n* = 9 (3 groups) Within-subject design	Qualitative analysis; Epistemic network analysis	Revised from **Nguyen, H.** (2021). Exploring Group Discussion with Conversational Agents Using Epistemic Network Analysis. *Communications in Computer and Information Science*, *1522 CCIS*, 378–394. https://doi.org/10.1007/978-3-030-93859-8_25
Study 4	Individuals' systems thinking (differences from pre to post-test) Discussion moves as mediators • Reasoning • Transactive exchange • Questioning	*n* = 172 (36 groups) Between-subject design	Multilevel linear regression	Not yet submitted

Data sources for the dissertation came from two study designs, involving a total of 224 students in grade 9. The first was a within-subject design, in which students in groups of two to three interacted with both the less-knowledgeable-peer and the expert agent in a learning task. The second design involved a randomized cluster design where student groups were randomly assigned to interact with either the peer or the expert agent in the same learning task. Data included students' chat logs with one another and with the agents and pretest and posttest that captured systems thinking. In addition, I administered a survey following interactions with the agents to examine students' perceptions of the agents and validate the agent designs. Table 1.1 provides an overview of the studies, areas of investigation, samples, and analytic procedures.

In Studies 1 and 2, I explored **how student groups interacted with the two agents as social partners**. Data drew from student messages (N = 1,764) from 18 groups (52 students in grade 9, ages 14-15). Study 1 focuses on the sequences of interaction to understand larger patterns of knowledge construction (Chen et al., 2017; Knight & Littleton, 2015). Student messages received codes for causal reasoning, transactive exchange (i.e., building on prior ideas by themselves and peers), and responsiveness to the agents. Results indicated no differences between agents in how often each discussion move occurred. Sequential pattern mining suggested that the less-knowledgeable-peer agent prompted groups to show questioning and building on others' ideas, similar to how students may act as peer tutors to the agent. Meanwhile, sequences with the expert agent resembled student-teacher exchange, where groups responded to the agent's nudges and then provided reasoning. Questioning and explaining sequences may promote exploratory learning and the construction of more coherent knowledge over time.

In Study 2, I examined how student groups attempted to **fix conversational breakdowns with the agents**, similar to how students would repair the exchange with human partners. Breakdowns happened when the agents failed to interpret users' intent, leading to frustration and potential abandonment of the interactions. I found that learner groups were generally tolerant of the agents: the most common strategies were to repeat, reframe, or explain their intent, as opposed to quitting the interactions. Groups rephrased and provided explanations during communication breakdowns with the less-knowledgeable-peer agent more often than the expert version. The frequencies of explanations during breakdowns were positively associated with indicators of deeper systems understanding in students' group chats.

Study 3 presents a case study to delve into the **link between group compositions and interactions with the different agent designs**. Embedding the agents in group discussions introduces interesting dynamics, since students with varying levels of prior domain knowledge may show different behaviors of help-seeking and help-giving from peers and the agents. Data came from the chat logs of three student groups (n = 9) with emergent, mixed, and expanding prior knowledge of systems thinking. The groups interacted with both agents.

To explore the different discussion patterns that students displayed, I used epistemic network analysis (ENA), a network analysis technique that visualized connections among co-occurring codes (Shaffer et al., 2016). In this case study, I examined the differences in discussion between emergent, mixed, and expanding groups and between the agent conditions in each group. Overall, the expanding groups engaged in more claim-making in tandem with building on prior ideas when interacting with the less-knowledgeable-peer agent, compared to the expert agent. Meanwhile, the emergent group showed more syntheses of previous ideas when responding to the expert agent. These results illustrate the importance of considering group compositions in examining interactions with learning agents.

Study 4 builds on the previous three studies to explore the **pathways between agent conditions, interactions with agents, and individual student learning**. Participants included 172 students ages 13-14. Students were randomly assigned to groups, and all members within the group interacted with one another and with no agent, a less knowledgeable peer, or an expert agent. Results reveal that agents deepened students' understanding of systems mechanisms by increasing their transactive exchange. Furthermore, comparisons of student groups based on pretest compositions (groups with high versus low variation at pretest) suggest that the less-knowledgeable-peer agent might encourage more balanced participation in transactive exchange for heterogeneous groups, compared to the expert agent.

Implications

Implications for Design. Together, the studies illustrate how designs of agents can position them as social partners in collaborative learning environments. Traditionally, agents serve as facilitators, providing conceptual hints and participation nudges. In the dissertation studies, I find that student groups showed more questioning sequences and explanations to the less-knowledgeable-peer agent as if they were assisting another peer. These findings illustrate how nuances in the agents' designs can facilitate varied interactions in student groups.

In the studies, students and agents co-create scientific concept maps. The agents' role is to promote ideation and elaboration instead of evaluation. This task structure may promote more knowledge construction and less hierarchical divisions of labor than tasks with predefined procedures. Thus, a direction for future work is to explore other design paradigms (beyond peer-expert) in an array of settings, for example, with different task structures and group sizes.

Implications for Learning. I find that interactions with the agents, such as building on previous ideas in transactive exchange or providing explanations, can enhance students' learning of systems concepts at the individual and group levels. This finding aligns with observations of productive discussion patterns in peer and teacher-led discussions. Furthermore, interaction patterns with the agents varied with group compositions along baseline knowledge. The discussion patterns that the studies uncovered thus broaden the interaction moments that learning systems and teachers may consider to promote productive discussion in heterogeneous groups.

Implications for Implementation. As intelligent systems become increasingly prevalent in home and educational settings, questions about how to implement these digital learning tools at scale become pertinent. The agents in this study have been used in both face-to-face and hybrid learning (i.e., face-to-face and virtual within the same lessons), thus illustrating their scalability in different settings. To consider at-scale implementation, I further discuss the design principles for the agent talk moves, such as focusing on elaboration and transactive exchange without being tied to specific domain concepts, as one way to easily adapt the agents to other subjects such as argumentation and English Language Arts.

References

Adamson, D., Dyke, G., Jang, H., & Rosé, C. P. (2014). Towards an agile approach to adapting dynamic collaboration support to student needs. *International Journal of Artificial Intelligence in Education*, *24*(1), 92–124. https://doi.org/10.1007/s40593-013-0012-6

Biswas, G., Jeong, H., Kinnebrew, J. S., Sulcer, B., & Roscoe, R. O. D. (2010). Measuring self-regulated learning skills through social interactions in a teachable agent environment. *Research and Practice in Technology Enhanced Learning*, *5*(02), 123–152.

Biswas, G., Segedy, J. R., & Bunchongchit, K. (2016). From Design to Implementation to Practice a Learning by Teaching System: Betty's Brain. *International Journal of Artificial Intelligence in Education*, *26*(1), 350–364. https://doi.org/10.1007/s40593-015-0057-9

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn* (Vol. 11). Washington, DC: National academy press.

Brown, B. A., Reveles, J. M., & Kelly, G. J. (2005). Scientific literacy and discursive identity: A theoretical framework for understanding science learning. *Science Education*, *89*(5), 779-802.

Chen, B., Resendes, M., Chai, C. S., & Hong, H.-Y. (2017). Two tales of time: uncovering the significance of sequential patterns among contribution types in knowledge-building discourse. *Interactive Learning Environments*, *25*(2), 162–175. https://doi.org/10.1080/10494820.2016.1276081

Diziol, D., Walker, E., Rummel, N., & Koedinger, K. R. (2010). Using intelligent tutor technology to implement adaptive support for student collaboration. In *Educational Psychology Review* (Vol. 22, Issue 1, pp. 89–102). Springer. https://doi.org/10.1007/s10648-009-9116-9

Dyke, G., Adamson, D., Howley, I., & Rose, C. P. (2013). Enhancing scientific reasoning and discussion with conversational agents. *IEEE Transactions on Learning Technologies*, *6*(3), 240–247. https://doi.org/10.1109/TLT.2013.25

Dyke, G., Howley, I., Adamson, D., Kumar, R., & Rosé, C. P. (2013). Towards Academically Productive Talk Supported by Conversational Agents. In *Productive Multivocality in the Analysis of Group Interactions* (pp. 459–476). Springer US. https://doi.org/10.1007/978-1-4614-8960-3_25

Graesser, A. C. (2016). Conversations with AutoTutor Help Students Learn. *International Journal of Artificial Intelligence in Education*, *26*(1), 124–132. https://doi.org/10.1007/s40593-015-0086-4

Hmelo-Silver, C. E., Jordan, R., Eberbach, C., & Sinha, S. (2017). Systems learning with a conceptual representation: a quasi-experimental study. *Instructional Science*, *45*(1), 53–72. https://doi.org/10.1007/S11251-016-9392-Y/FIGURES/5

Hogan, K., Nastasi, B. K., & Pressley, M. (1999). Discourse Patterns and Collaborative Scientific Reasoning in Peer and Teacher-Guided Discussions. *Cognition and Instruction*, *17*(4), 379–432. https://doi.org/10.1207/S1532690XCI1704_2

Kerly, A., Ellis, R., & Bull, S. (2008). CALMsystem: A Conversational Agent for Learner Modelling. In *Applications and Innovations in Intelligent Systems XV* (pp. 89–102). Springer London. https://doi.org/10.1007/978-1-84800-086-5_7

Kim, Y. (2007). Desirable Characteristics of Learning Companions. *International Journal of Artificial Intelligence in Education*, *17*(4), 371–388.

Kim, Y., & Baylor, A. L. (2016). Research-Based Design of Pedagogical Agent Roles: A Review, Progress, and Recommendations. *International Journal of Artificial Intelligence in Education*, *26*(1), 160–169. https://doi.org/10.1007/s40593-015-0055-y

Knight, S., & Littleton, K. (2015). Dialogue as Data in Learning Analytics for Productive Educational Dialogue. *Journal of Learning Analytics*, *2*(3), 111–143. https://doi.org/10.18608/jla.2015.23.7

Kumar, R., Beuth, J., & Rose, C. (2011). Conversational Strategies that Support Idea Generation Productivity in Groups. In H. Spada, G. Stahl, N. Miyake, & N. Law (Eds.), *Connecting Computer-Supported Collaborative Learning to Policy and Practice: CSCL2011 Conference Proceedings. Volume I - Long Papers* (pp. 398–405). International Society of the Learning Sciences. https://repository.isls.org/handle/1/2475

Liew, T. W., Tan, S.-M., & Jayothisa, C. (2013). The Effects of Peer-Like and Expert-Like Pedagogical Agents on Learners' Agent Perceptions, Task-Related Attitudes, and Learning Achievement. *Educational Technology & Society*, *16*(4), 275–286. https://eric.ed.gov/?id=EJ1013641

McKinley, D. C., Miller-Rushing, A. J., Ballard, H. L., Bonney, R., Brown, H., Cook-Patton, S. C., Evans, D. M., French, R. A., Parrish, J. K., & Phillips, T. B. (2017). Citizen science can improve conservation science, natural resource management, and environmental protection. *Biological Conservation*, *208*, 15–28.

Nass, C., Moon, Y., Fogg, B. J., Reeves, B., & Dryer, D. C. (1995). Can computer personalities be human personalities? *International Journal of Human-Computer Studies*, *43*(2), 223–239. https://doi.org/10.1006/ijhc.1995.1042

Rosenberg-Kima, R. B., Baylor, A. L., Plant, E. A., & Doerr, C. E. (2008). Interface agents as social models for female students: The effects of agent visual presence and appearance on female students' attitudes and beliefs. *Computers in Human Behavior*, *24*(6), 2741–2756. https://doi.org/10.1016/j.chb.2008.03.017

Scardamalia, M., & Bereiter, C. (1991). Higher Levels of Agency for Children in Knowledge Building: A Challenge for the Design of New Knowledge Media. *Journal of the Learning Sciences*, *1*(1), 37–68. https://doi.org/10.1207/s15327809jls0101_3

Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics*, *3*(3), 9–45.

Snapir, Z., Eberbach, C., Ben-Zvi-Assaraf, O., Hmelo-Silver, C., & Tripto, J. (2017). Characterising the development of the understanding of human body systems in high-school biology students – a longitudinal study. *Http://Dx.Doi.Org/10.1080/09500693.2017.1364445*, *39*(15), 2092–2127. https://doi.org/10.1080/09500693.2017.1364445

Tegos, S., & Demetriadis, S. (2017). Conversational Agents Improve Peer Learning through Building on Prior Knowledge. *Educational Technology & Society*, *20*(1), 99–111.

van Aalst, J. (2009). Distinguishing knowledge-sharing, knowledge-construction, and knowledge-creation discourses. *International Journal of Computer-Supported Collaborative Learning*, *4*(3), 259–287. https://doi.org/10.1007/s11412-009-9069-5

Walker, E., Rummel, N., & Koedinger, K. R. (2011). Designing automated adaptive support to improve student helping behaviors in a peer tutoring activity. *International Journal of Computer-Supported Collaborative Learning*, *6*(2), 279–306. https://doi.org/10.1007/s11412-011-9111-2

Walker, E., Rummel, N., & Koedinger, K. R. (2014). Adaptive intelligent support to improve peer tutoring in algebra. *International Journal of Artificial Intelligence in Education*, *24*(1), 33–61. https://doi.org/10.1007/s40593-013-0001-9

CHAPTER 2 INTERACTION SEQUENCES WITH AGENTS

An earlier version of this chapter has appeared in:
Nguyen, H. (2022). Let's Teach Kibot: Discovering Discussion Patterns between Student Groups and Two Conversational Agent Designs. *British Journal of Educational Technology*. http://doi.org/10.1111/bjet.13219

Abstract

Conversational agents can deepen reasoning and encourage students to build on others' knowledge in collaborative learning. Embedding agents in group work, however, presents challenges where groups may ignore the agents and calls for designs where students perceive agents as learning partners. This study examines group interactions with two text-based agents (i.e., chatbots) that posed as an expert and a less knowledgeable peer in a high school marine biology lesson. Student messages (N = 1,764) from 18 groups (52 students ages 14-15) received codes for reasoning, building on prior ideas, and responsiveness to the agents. Results indicate no differences between agents in how often each discussion move occurred. Interestingly, sequential pattern mining suggests that groups showed more sequences of questioning and building on others' ideas with the less-knowledgeable-peer agent, similar to how students may act as peer tutors to the agent. Meanwhile, sequences with the expert agent resembled student-teacher exchange, where groups responded to the agent's nudges and then provided reasoning. Findings illustrate the affordances of embedding humanized features in technology designs to promote discussion.

Introduction

Collaborative knowledge construction, where students negotiate ideas to deepen group understanding, plays a central role in the exchange and production of scientific knowledge (Scardamalia & Bereiter, 1993). However, simply inviting students to the discussion does not always yield productive learning (Dillenbourg, 2002). Conversational agents—dialogue systems that provide learning support through natural language interactions—can serve as a potential solution to facilitate discussion (Adamson et al., 2014; Diziol et al., 2010; Dyke, Adamson, et al., 2013; Walker et al., 2014).

To make agents more engaging to students, designers have experimented with agents' appearances and linguistic features to simulate characteristics from familiar figures, such as mentors and tutees (Chen et al., 2020; Graesser, 2016; Y. Kim & Baylor, 2006). These profiles may prime users to display respective conversational norms as if they were getting hints from a knowledgeable tutor or giving help to an inexperienced tutee (Graesser, 2016).

However, researchers have mostly considered agent designs in individual settings, where a student interacts with one or several agents (Biswas et al., 2016; Y. Kim & Baylor, 2016). Different from individual contexts, researchers have observed that student groups may ignore or abuse the agents (Kumar et al., 2010). Exploring variations in groups' interactions with agent profiles thus sheds light on how to best support groups' social and learning needs.

This study explored these interactions in a within-subject design, where 18 student groups (n students = 52, ages 14-15) in 9^{th}-grade science classes chatted with one another and with two text-based agents: a less knowledgeable peer and an expert. Students' messages received codes for causal reasoning, transactive exchange where students built on prior ideas, and responsiveness to the agents. Analyses examined how student groups' frequencies and sequences of interactions differed between agents. This analytical focus assumed that the frequencies and sequences of discussion moves were related to knowledge construction and knowledge acquisition in science classrooms (Chen et al., 2017; Wise & Chiu, 2011).

Theoretical Background

Designing effective conversational agents (CAs) calls for understanding how to promote knowledge construction and human-computer interactions. The current research thus builds on frameworks in knowledge construction (Hewitt & Scardamalia, 1998; Scardamalia & Bereiter, 1993; Weinberger & Fischer, 2006) and examples of CAs in education (e.g., Adamson et al., 2014; Dyke, Adamson, et al., 2013).

Reasoning and Transactivity in Knowledge Construction

Knowledge construction is a key facet of science discussion. This process involves students actively sharing ideas and building on others' work to advance collective understanding (Hmelo-Silver & Barrows, 2008; Hoadley & Kilner, 2005; Muhonen et al., 2017; Scardamalia & Bereiter, 1993; Stahl et al., 2014). Students negotiate a fit between their ideas and those of others, with a focus on explanations and causal mechanisms (Hewitt & Scardamalia, 1998; van Aalst, 2009). In the process, they construct **reasoning** and **extend others' ideas** through social exchange (Weinberger & Fischer, 2006).

Students develop and balance **reasoning** in working on communal tasks (Weinberger & Fischer, 2006). For example, when constructing a scientific concept map together, they assert and reformulate ideas about connections among concepts, challenge others to find evidence, or pose questions to elicit explanations. Exploring how students make claims, employ reasoning, and use questions to guide queries provides insights into jointly constructed understanding (Hmelo-Silver & Barrows, 2008).

Knowledge construction is inherently social, with many opportunities for students to **build on others' contributions** (Hewitt & Scardamalia, 1998; Weinberger & Fischer, 2006). In their seminal work on computer-supported knowledge-building environments, Hewitt and Scardamalia (1998) called for systems that provided access to different modes of participation and idea artifacts, such as written text, diagrams, and drawings. These artifacts can detail idea evolution and encourage students to acknowledge, extend, and challenge others' queries.

The extent to which students refer to prior ideas in discussion is termed transactivity (Teasley, 1997). Students may showcase lower-order transactivity, where they externalize their ideas or request clarification without elaboration of others' reasoning. Meanwhile, higher-order transactivity involves integration and counter-arguments of others' contributions (Teasley, 1997; Wen et al., 2016). Effective knowledge construction environments should foster higher-order transactivity, which has been linked to enhanced individual and group learning (Adamson et al., 2014; C. Rosé et al., 2008; Wen et al., 2016).

The responsibility to promote knowledge construction does not rest solely with the learners, but also on the learning facilitators (Gillies, 2014; Scardamalia & Bereiter, 1993). Facilitators can call on certain students to initiate discussion moves. They can invite collective reasoning and support students to build on ideas (Hmelo-Silver & Barrows, 2008; Muhonen et al., 2017). Emergent research has explored how CAs can adopt these facilitative roles.

Designs of Conversational Agents

CAs can facilitate knowledge construction in real time (Adamson et al., 2014; Dyke, Howley, et al., 2013; Tegos & Demetriadis, 2017). Using natural language understanding, the agents process students' unfolding conversations to propose relevant prompts. Agents have shown promise in encouraging transactive exchange and learning (Howley et al., 2013; Kumar et al., 2011). However, researchers have also documented cases where groups abused or ignored the agents (Kumar et al., 2010). CAs built on mechanical, task-oriented paradigms may not fit into group interactions that also value socio-emotional exchange (Kumar et al., 2010). Socially capable agents that display verbal cues such as self-disclosure, reassurance, and complements have been associated with effective support in tutoring, compared to agents without these features (Kumar et al., 2010; Romero et al., 2017).

These design cues build on the Computers-Are-Social-Actors (CASA) paradigm (Nass et al., 1994). The CASA paradigm suggests that users of computer systems often associate the computers with characteristics traditionally reserved for human partners, such as trust, reciprocity, and competence (Gong, 2008; Pearson et al., 2006; Zhou et al., 2019). Users enact social norms, for example, reciprocating help if the computers offer support or stereotyping that an agent is extroverted based on its assumed tone and gender (S. Kim et al., 2019; Lee & Nass, 2003; Moon & Nass, 1996; Nass et al., 1997; Nass & Moon, 2000).

Applying the CASA paradigm to learning settings offers design opportunities for CAs. The verbal and physical cues of the agents can align with prototypes of learning partners in human-human interactions. Researchers have particularly explored the potential of agents as a peer and an expert (Biswas et al., 2016; Chen et al., 2020; Graesser, 2016; Y. Kim & Baylor, 2006; Rosenberg-Kima et al., 2008). A peer agent possesses comparable knowledge levels and discourse to those of the learners (Y. Kim & Baylor, 2006). The profile builds on the similarity-attraction effect (Byrne & Nelson, 1965), which suggests that users would find agents that resemble their appearance, knowledge, or interest more appealing. A related design is the learn-through-teaching model (Biswas et al., 2016), where students acquire knowledge through explaining and giving feedback to an agent with less knowledge than them. Meanwhile, an expert or tutor agent appears to possess more advanced expertise. Students may perceive this design as more competent, and consequently, treat the agent as a teacher and seek feedback from the agent (Biswas et al., 2016; Y. Kim, 2007; Liew et al., 2013).

Students' Interactions Vary with Agent Designs

Students may demonstrate different behaviors and learning outcomes when interacting with different agents (Heidig & Clarebout, 2011). For example, young learners ages 5-7 showed slightly more affective facial displays when interacting with a robot that behaved as a peer, while acquiring more vocabulary from interactions with a tutor robot (Chen et al., 2020).

Emergent work has explored interaction sequences between students and agents (Howley et al., 2013; Jeong et al., 2008; Kinnebrew et al., 2014). The focus on interaction processes overlaps with knowledge construction's emphasis on cycles of idea formulation and refinement (Scardamalia & Bereiter, 1993). Howley et al. (2013) examined groups' exchange in three conditions: an agent that provided direct nudges for transactivity, an agent with indirect nudges, and no agent. The sequence analyses revealed consistent off-task periods in the indirect nudge condition when the agents' prompts were untimely. This finding called for redesigning the agent to promote transactive exchange at more opportune moments (Howley et al., 2013).

Analyses of interaction patterns can reveal how the peer versus expert agent contributes to knowledge construction. If students treat the agents as social partners (Nass et al., 1994), they may demonstrate behaviors similar to how they interact with teachers versus peers in tutoring (Almasi, 1995; King et al., 1998; Roscoe & Chi, 2007). Students verbalize thinking through questions and subsequent explanations more frequently when acting as peer tutors, compared to when they respond to teachers (Almasi, 1995; Berghmans et al., 2013; King et al., 1998).

Interaction sequences may also vary with the assumed roles. For instance, questioning in teacher-led discussions assumes that teachers are the knowledge providers, and students simply respond to the teachers' prompts (Hogan et al., 1999; Scardamalia & Bereiter, 1991). In contrast, when students give hints to a peer, the peer's answers may trigger more varied response sequences from the student tutors, from feedback and explanation to further questioning. Interactions with peers and teachers can both generate more complex conceptual reasoning, but they likely involve different sequences (Hogan et al., 1999). Sustained sharing of queries, followed by idea co-construction by the group, helps to articulate gaps in the groups' knowledge towards more structured understanding. In comparison, interactions with teachers can be productive if students respond to teachers' questions with idea elaboration.

In sum, researchers have examined different designs of CAs in individual settings (e.g., Biswas et al., 2016; Chen et al., 2020; Graesser, 2016; Y. Kim & Baylor, 2006; Rosenberg-Kima et al., 2008), but not yet in collaborative contexts. Understanding groups' interactions with different agent designs can contribute to creating agents that promote behaviors conducive to learning in collaborative settings. In this work, I explore student interactions with two text-based agents: a less knowledgeable peer and an expert. Instead of exploring the whole discussion history or separate discussion moves, I focus on interaction sequences as an intermediate unit to understand larger patterns of knowledge construction. The following questions guide this work:

RQ1. What discussion moves do student groups leverage in agent interactions? How do groups' discussion moves differ between the less-knowledgeable-peer and expert agent?

RQ2. How do sequences of discussion moves (e.g., progression of questioning, reasoning, and elaboration) differ between agent conditions?

Methodology

Study Setting

The current study is part of a multi-year partnership between a local state park, education and biology researchers, and local teachers in the southwestern United States. The agents were integrated into the state park's environmental program, entitled Marine Science Exploration (MSE). The MSE program consisted of eight lessons to engage high school students in a participatory science curriculum that anchored scientific concepts in local contexts (McKinley et al., 2017). The MSE program centered around the state park's marine protected area, which was created to reduce threats from human-driven climate change, pollution, and overfishing. During the program, students learned about systems elements within the local marine ecosystem.

In observing those lessons, the park educators and teachers noticed that not all students participated equally in discussions. Students also focused on linear links (e.g., fish eat plankton), instead of complex system processes (e.g., ocean acidification influences phytoplankton's abundance, disrupting food chains). The author and the state park developed the text-based CAs (Kibot less-knowledgeable-peer and Kibot expert) to address these challenges.

Participants

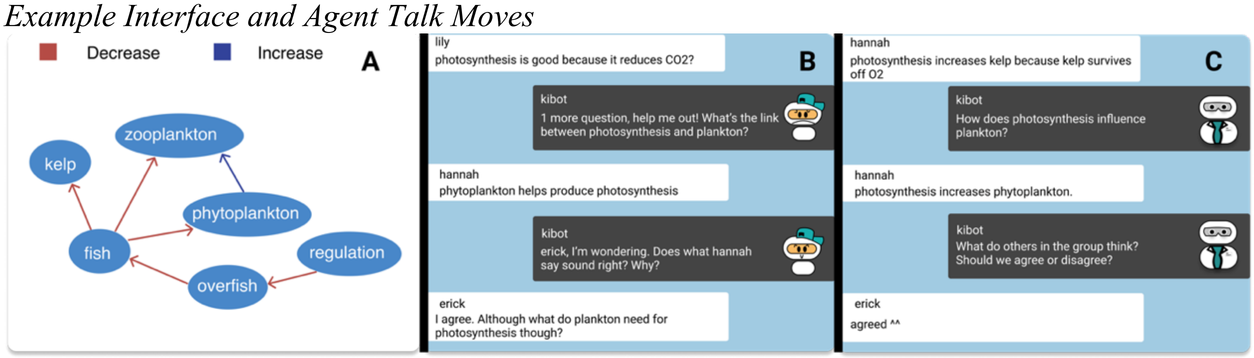
Participants were 18 groups of two to three (52 students, two 9th-grade classes) taught by the same science teachers in a public high school in the southwestern U.S. The school served a diverse student population that was 46% White, 38% Latinx, and 9% Asian in 2019-20. Students were participating in the MSE program during their science class time. The school had a one-to-one laptop policy, and students were familiar with using chat windows. The 60-minute lesson occurred when the school was enacting social distancing due to the COVID-19 pandemic. Students mostly relied on the chat interface instead of verbal interactions as they interacted with the agents.

Learning Task

The agents were embedded into a modeling lesson early in the MSE program. After learning about the marine protected area, students worked in groups of two or three to create a concept map of the park's marine ecosystems. Students focused on how changes in an element affected other elements and processes, for example, building connections between kelp, habitat, and biodiversity. Students communicated through a chat window (Figure 2.1). Their messages were compared to an underlying "expert" map (described further below). To establish content validity, five park educators and six university marine biology researchers collaborated through three iterations to refine the expert map. Matched connections between students' and experts' answers appeared on the web interface next to students' chats (panel A, Figure 2.1). The agents kept track of the chats and provided prompts to help groups reflect on missing connections.

On average, groups interacted with each agent for 12.5 minutes, SD = 1.2. The interaction order with the agents was randomized to minimize practice effect with the second agent. Half of the student groups started with the less-knowledgeable-peer agent, while the other half started with the expert agent, and switched halfway through the lesson. Appendix 2.1 illustrates the switching from the expert to the less-knowledgeable-peer agent, where the agent's appearance and linguistic styles change accordingly.

Figure 2.1
Example Interface and Agent Talk Moves



Notes. Panel A: Concept maps from students' chat; panel B: less-knowledgeable-peer condition; panel C: expert condition

Agent's Prompt Designs

Focus on Causal Reasoning

The agents' prompts focus on dimensions of scientific reasoning: element, evidence, and causal coherence (Kang et al., 2014; Nguyen & Santagata, 2020). Element captures living organisms, (e.g., fish), non-living components (e.g., sun), and processes (e.g., global warming). Evidence describes how students use empirical data or experiences to support ideas. Finally, causal coherence refers to how students connect concepts to scientific ideas in a logical chain of reasoning, such as emphasizing feedback loops within systems.

The agents use natural language processing to parse students' messages and select the appropriate responses that promote element, evidence, and causal coherence. Using the Python's package "spaCy" dependency parser (Honnibal & Montani, 2017), the agents segment students' chats into subjects and objects. For example, a message like "fish decreases plankton" gets parsed into "fish" (the subject) and "plankton" (the object).

The agents compare students' concept maps (the subject-object pair) and the expert maps using two algorithms: (1) fuzzy matching with Levenshtein-based string similarity and (2) word embedding. Levenshtein distance measures the distance between a student's answer versus the expert map by counting the number of edits to turn one string into the other. The fuzzy matching disregards the punctuation and word order.

A limitation to string similarity approaches is that they do not capture cases where terms fall under similar domains but may not be exactly similar. For this, the agents calculate the similarity between students' and expert maps using word embeddings with spaCy (Honnibal & Montani, 2017). The subject and object from student answers are turned into high-dimensional vectors that capture their contexts (i.e., the words they are surrounded by). Shorter distances between a vector that represents a student's term and a vector for the expert answer would indicate higher similarity. If the semantic similarity is above .80, the terms are considered similar to the experts' terms. The .80 threshold was determined through user testing and calculating the range of semantic similarity values between users' responses and expert terms.

If there exists a missing link between a term that students mention and an expert's term, the agents provide hints for the missing terms. If students miss the hints, the agents follow up with a prompt that explicitly mentions the link between the target element and existing elements in students' concept maps. To foster use of evidence, the agents ask for students' reasoning. To promote causal coherence, the agents order the hints so that students finish all connections among existing terms, before moving onto another. Table 2.1 outlines these agents' prompts.

Focus on Transactivity

The agents utilize transactive talk moves (Dyke, Howley, et al., 2013). After every five talk turns, agents invite students to explain, elaborate on a previous statement (made by themselves or others), or discuss why they agree or disagree with peers. Within groups, based on the chat counts per student, the agents direct the transactive nudges at those who have participated the least in the conversation. Figure 2.1 provides example chats.

The Less-knowledgeable-peer and Expert Agents

Following examples from one-to-one agent-student interactions (Biswas et al., 2010; Chen et al., 2020; Y. Kim & Baylor, 2006), this study tests two key designs: less-knowledgeable-peer and expert agents. The agents send the same prompts, but these prompts differ in their wording and agents' expressions to display different emotions and competence. Competence and emotions have been linked to students' affective experiences and cognitive engagement (Y. Kim & Baylor, 2006).

The Less-Knowledgeable-Peer Agent

The less-knowledgeable-peer agent builds on the learn-through-teaching paradigm to represent a peer with lower levels of knowledge (Biswas et al., 2010). The agent explicitly states that it is learning from student chat. To simulate the agent's learning, the chat includes a "knowledge bar" that gets updated with each new link students make in the concept map. The agent uses colloquial expressions to ask students to explain concepts in ways that expand its knowledge and sends animated texts when students build a correct systems connection. The agent shows multiple social expressions, such as excitement, confusion, and appreciation. The animations change with these emotions, for instance, frowning when confused.

The Expert Agent

The expert agent is portrayed as a scientist. The agent enters the group chat by asking students if they are ready to learn. This agent can also answer students' questions to define terms students may include in the concept map. The expert agent speaks in a formal tone and holds a static expression throughout the interactions with the students. To avoid situations where students abuse the expert agent for hints, if students ask for hints or questions beyond term definitions, both agents encourage students to discuss the questions with their group members instead of providing answers.

To validate the agent designs, the author conducted semi-structured focus groups with five user groups of three to four each (n = 17 participants, each session lasted about 15 minutes). Participants were a convenience sample of high school (n = 4), college students (n = 2), college graduates (n = 7), and park educators (n = 4). Participants interacted with each agent for 10 minutes and rated the agents on their role (peer or expert), emotion, and competence. All participants identified the agents' characteristics as intended.

Table 2.1
Examples of the Agents' Talk Moves

Talk move	Description	Less-Knowledgeable-Peer	Expert Kibot
Reasoning (Graesser, 2016; Nguyen & Santagata, 2020)			
Hint	Encourage students to think about missing systems elements	I'm learning so much from you! Here's something I still don't understand: What's the role of **plankton** in this system?	Great! Now think about: What's the role of **plankton** in this system?
Prompt	Prompt students to think about connection with a term they haven't mentioned (if students miss the Hint)	Ooh I have an idea, is **photosynthesis** [missing term] connected to **plankton** [existing term in students' map] in any way?	That sounds good. How is photosynthesis [missing term] connected to plankton [existing term in students' map]?
Evidence	Ask students to provide evidence to explain their systems connections	Explain to me, why do you think so?	Can you provide evidence for your answers?
Transactivity (Fiacco & Rosé, 2018; Teasley, 1997)			
Elaborate	Ask students to elaborate on ideas	Learning is hard! *LifeisGood* [student username]**,** can you explain your thinking to me?	*LifeisGood* [student username], what do you think?
Agree/ Disagree	Ask students to critique peer's ideas	*LifeisGood* [student username], help me out. Should we agree or disagree with your friends? Why?	*LifeisGood* [student username], do you agree or disagree with what your friends just said? Why?
Uncertainty	Encourage group discussion when students ask for hint or express confusion	I am not sure. Can someone in the group help?	I can't tell you the answer. Why don't you discuss with your peers and let me know?
Opinion conformity	Show alignment with students' ideas to elicit idea elaboration	I love that. What do you all think?	That's a great idea. What about others in the group?
Concept Definition			
Define concepts	Define concepts in the ecosystems (only expert agent)		Student: What's kelp? Agent: They're basically huge seaweed? Kelps live in dense groupings and provide food and shelter for other marine animals
Social Expressions (Sebo et al., 2020)			
Acknowledgement	Acknowledge students' responses	[animated image] You're awesome!	Good job.

Excitement	Show excitement at the groups' progress (only less-knowledgeable agent, accompanied by changes in expression)	Look at the progress we've made! We're so closed to finishing all interactions.	
Confusion	Show confusion and ask for students' elaboration (only less-knowledgeable agent, accompanied by changes in expression)	I am confused. Why isn't this connection between **phytoplankton** and **photosynthesis** the other way around?	
Small talk	Social conversations about common topics (sports, hobbies, agents' gender, background)	I love all types of sports! But baseball is my favorite. Have you noticed the cap I'm wearing?	I don't really watch sports. Does studying planktons under the microscope count?
Fallback	Fallback when Kibot can't parse student text for 5 turns	Sorry I'm learning. Can you break your ideas down for me? Start with something simple, 'Global warming increases temperature', and provide your reasons.	I didn't catch that. Can you rephrase what you said, starting with a simple sentence, like 'Global warming increases temperature', and then provide reasoning?

Notes. Explanations are provided in brackets.

Observing how test users interacted with the agents surfaced additional talk moves to improve the conversation flow, including small talk (e.g., discussion about the agents' favorite sports) and opinion conformity (e.g., "I love that! What do others in the groups think"). Table 2.1 presents examples of all agents' talk moves, grouped under Reasoning, Transactivity, Concepts, and Social Expressions.

Data Sources

The main data source came from students' chat logs. Each row of the log consisted of a message, group ID, student username, timestamp, and agent condition (less-knowledgeable-peer or expert). The agents' messages were retained for context but were not included in the analyses.

There were two coding iterations for discussion moves. The first iteration built on prior frameworks in knowledge construction (Fiacco & Rosé, 2018; Teasley, 1997; Weinberger & Fischer, 2006). Each message was coded for how students constructed their reasoning and engaged in transactive exchange. The reasoning dimension describes how students produce **claims**, **reasoning** to warrant claims, and **questions** to guide the group discussion. Reasoning consists of evidence from scientific facts, observations, or personal experiences, with logical connections for how such evidence may support claims (Nguyen & Santagata, 2020; Weinberger & Fischer, 2006).

The transactive dimension accounts for the social aspects of knowledge construction and consists of **transactive** and **externalizing** moves (Roschelle & Teasley, 1995; Weinberger & Fischer, 2006). Transactive acts can be broken down into different moves, for example, when students accept the contributions of a partner, assume the perspectives of a partner, or challenge and modify a partner's stances (Weinberger & Fischer, 2006). These moves are distinguished from externalizing when students revisit their ideas (Teasley, 1997).

In the second iteration, the author excluded or merged codes with low occurrences and used a bottom-up approach to develop emergent codes. First, I observed that most of the transactive moves from students were to integrate their partners' perspectives into their answers, and there was only one occurrence where they challenged another partner's ideas. Thus, these codes were merged under the transactive category. Second, **responsiveness** emerged as another category, to denote when students responded to nudges by the agents.

The codes were dichotomous (1 for occurrence, 0 for non-occurrence). Each message could receive codes for multiple categories, for instance, if a message showed both reasoning and transactive talk. Table 2.2 presents the final coding scheme. Once the coding scheme was established, the author and a research assistant separately coded 20% of the data and achieved acceptable agreement across dimensions (reasoning: Cohen's $\kappa = .98$; transactivity: Cohen's $\kappa = .92$; responsiveness: $\kappa = 1$).

Table 2.2
Coding Scheme for Students' Chat

Code	Description	Example
Reasoning		
Claim	A statement about an idea or concept	Fish decreases plankton
Reasoning	Evidence or explanation to support a claim	… because plankton is abundant in the ocean and is easy food source.
Questioning	Guiding questions for the discussion	If the water temperature changes what would happen?
Transactivity		
Externalizing	Articulate one's own prior thoughts to groups	As I mentioned, I think we should also regulate CO_2 emissions.
Transactive	Integrate, apply, or challenge perspectives of a peer's prior ideas	I agree, whales decrease fish because fish is whales' s food source. Whales eat plankton too.
Responsiveness	Students respond to agent's prompts.	Kibot: What would happen if CO_2 increase? imastar: There'll be more acidification.

Analytical Strategies

Student group was the unit of analysis. Group-level analyses align with the study's theoretical focus on how groups' ideas evolve in knowledge construction (Chen et al., 2017; Scardamalia & Bereiter, 1993). Analyses at the individual level were less suitable due to the small occurrences of chat moves per individual (M = 6 utterances per individual per agent). Interaction sequences emerging from such a small number of utterances may be less meaningful. The Limitations section outlines the constraints of this analytical decision in more detail.

As a robustness check for group-level analyses, I examined whether individuals equitably contributed to the discussion. I manually pulled out all chat sequences that involved questioning, which can be a marker that differentiates mentor-led and peer tutoring (Hogan et al., 1999). Of those, I calculated the proportions of sequences where multiple students contributed. Group exchange constituted the majority of sequences (64.5%; 40 out of 62 sequences).

Discussion Moves and Differences Between Agent Conditions

The first research question examined differences between agent conditions in occurrences of discussion moves for reasoning (claim, reasoning, questioning), transactivity (externalize, transactive), and responsiveness to agents. Student groups did not differ in the number of messages they sent between agent conditions. On average, groups sent 16.21 messages, SD = 16.01 in the peer condition, while they sent 20.64 messages, SD = 32.05 in the expert condition ($W = 302.5$, $p = .52$). Due to the high standard deviations, to normalize interaction patterns between groups, I calculated the ratios of each move's occurrences out of all messages that each group sent, and used Wilcoxon signed-rank tests to examine whether there was a significant difference between conditions for the ratios. To account for multiple comparisons, I used Benjamini-Hochberg corrections with the false discovery rate of .05.

Differences in Sequential Patterns

To answer RQ2, I applied sequential pattern mining to examine differences in the sequences of discussion moves between agent conditions. The same codes for discussion moves were arranged in the temporal order in which they occurred. The rules for the sequential patterns were identified using R's arulesSequences packages (Buchta et al., 2020). The package used the cSPADE algorithm (Zaki, 2000) to identify the temporal association of an event and a subsequent one based on frequencies of occurrences. In this study, the students' chat logs formed a set of sequences (e.g., one sequence per student group per condition), and each sequence contained a set of reasoning, transactive, and responsive moves. For example, if sequence S1 started with "claim", there would be some likelihood that "reasoning" would follow "claim" within the same sequence and form the pattern claim -> reasoning.

The study considers three metrics (support, confidence, and lift) to capture the likelihood of a sequential pattern and identify candidate sequences for subsequent analyses. Support, ranging between 0 and 1, describes the proportion that a specific pattern occurred out of all sessions. For example, a support value of .10 for a sequence such as "claim => reasoning" suggests that this pattern occurred in 10% of the sequences. Confidence, also ranging between 0 and 1, indicates the likelihood of a discussion move B to follow A, once A occurred. A confidence value of .25 for the same “claim => reasoning” sequence, for instance, would indicate that if we see “claim” in a message, there is a 25% chance that “reasoning” would appear in subsequent messages within the same talk window. Finally, lift shows the support for a pattern (e.g., A->B), divided by the support for A times the support for B. A lift value greater than 1 suggests a positive likelihood that a pattern would occur, relative to chance occurrences of observing A and B independently. Using the same example sequence of “claim => reasoning”, a lift value greater than 1 would imply that the presence of “claim” has increased the probability that “reasoning” would occur in subsequent messages. Prior work has noted that a high value of lift may indicate added values, since they have high correlations with domain experts’ judgment of interesting patterns (Bazaldua et al., 2014; Merceron & Yacef, 2008).

The thresholds for the parameters were set as follows: support at .25, confidence at .50, and lift at 1.25. These thresholds helped capture a wider range of sequences than setting high support and confidence thresholds (that occurred frequently but may not have high lift values). The findings present the highest lift-value sequential patterns for each agent condition. For example, the top sequences in the expert agent condition were questioning, followed by another question with a lift value of 1.73. Findings focus on lift values to indicate levels of interestingness and include excerpts from student discussions to illustrate the patterns.

The discussion within a sequential pattern fell within a 1-minute window to capture relevant contexts. For robustness check, I examined the top sequences (by lift values) for window sizes of two, three, and five minutes (Appendix 2.3). The sequences differed slightly but had consistent patterns across sizes. For instance, for all sizes, sequences in interaction with the less-knowledgeable-peer agent frequently involved questions and elaboration.

To explore the differences between agent conditions, I calculated the occurrences of the top patterns for each student group. Wilcoxon tests were used to determine whether occurrences differed by agents, with pattern occurrences as the dependent variables and the agent conditions as the independent variable. The tests used Benjamini-Hochberg corrections (false discovery rate of .05) to account for multiple comparisons.

Finally, in the study's within-subject design, groups started with an agent and switched to the other agent halfway through the lesson. Thus, I compared the interaction patterns with each agent in relation to which agent a student group started the learning task with. These analyses explored whether interactions with each agent remained consistent regardless of the starting agent.

Findings

Frequencies of Discussion Moves

The descriptive statistics present an overview of students' interactions. Within conditions, students sent the most claims (peer: 264 messages; expert: 316), followed by externalizing their own ideas (peer: 172; expert: 173), responsiveness (peer: 101; expert: 114), transactive exchange to build on peers' ideas (peer: 93; expert: 96), and questioning (peer: 89; expert: 73). Reasoning had the lowest occurrences (peer: 60; expert: 72).

Wilcoxon tests suggest no substantial difference between conditions in the ratio of occurrence for reasoning, transactivity, or responsiveness (Table 2.3). These findings indicate that the agent designs might have resulted in similar levels of engagement for student groups. Additionally, Mann-Whitney tests suggest that discussion moves were consistent regardless of the agent profile groups started with. Recall that groups started with an agent (less-knowledgeable-peer or expert) and switched to the other during the lesson. The starting agent profiles did not significantly influence the frequencies of discussion moves (Appendix 2.2).

Table 2.3
Wilcoxon Results for Ratios of Groups' Discussion Moves

Code	Mpeer	SDpeer	Mexpert	SDexpert	W	p	adjusted-p
Reasoning							
Claim	.36	.49	.31	.30	428.50	.91	.95
Reasoning	.20	.21	.16	.15	124.50	.69	.95
Questioning	.22	.22	.25	.28	75.50	.95	.95
Transactivity							
Externalizing (self's ideas)	.39	.16	.35	.10	127.50	.57	.95
Transactive (prior ideas, friend)	.21	.11	.19	.12	122.50	.47	.95
Responsiveness	.25	.15	.23	.13	123.00	.88	.95

Questioning Sequences with the Less-Knowledgeable-Peer Agent

The second question explores the temporality in groups' interactions. Table 2.4 lists the top patterns for each agent condition, ranked by lift values to indicate interestingness. For example, the two sequences with the highest lift values in the less-knowledgeable-peer condition were (1) questioning, followed by externalizing one's ideas and questioning; and (2) reasoning and responding to agent, followed by transactive exchange.

Meanwhile, the top two sequences in the expert condition were (1) questioning, followed by another question; and (2) responding to agent and building on one's idea, followed by transactive exchange. Out of the top five sequences in the less-knowledgeable-peer condition, three involved questioning the agent and elaborating on ideas. Consider excerpt 1, when students posed a question and continued to build on prior ideas with more questioning.

S1: **what kills fish**?

S1: if there were **plastic** there what would happen

Kibot: I am not sure. What do you think would happen?

S1: **plastic kills fish** because fish gets stuck. **Does plastic kill fish**?

S1 began his interaction with Kibot less-knowledgeable-peer with open-ended questions. In response, Kibot expressed uncertainty to invite students to articulate their thinking. This utterance prompted S1 to provide an answer to his prior questions and to follow with another question about the same concept.

A related pattern is for students to co-construct several causal links related to a common construct, and then pose questions (Transactive-multiple -> Question). For example, in the following excerpt (excerpt 2), students S2 and S3 were discussing links between global warming, ocean temperature, oxygen, fish, and planktons. Their exchange ended with a question regarding the chain of connections they just discussed (high temperature → O_2 → fish). Such pattern of claim-making and transactive exchange, followed by questioning was present in many groups in the less-knowledgeable-peer condition (*n* occurrences = 10).

S3: Global warming decreases oxygen.

S2: Yeah, because **warm water** holds less O_2.

S2: And O_2 is important because **fish** needs O_2.

S3: zooplankton increases O_2 because zooplankton creates O_2.

S2: I think phytoplankton does.

S2: How does **higher temperature** effect **fish**?

Meanwhile, the sequences with the highest lift values in the expert condition mostly consisted of responsiveness and elaboration of previous ideas. These patterns consisted of students answering the agent's prompts to articulate claims, provide explanations, or build on ideas from peers. Excerpt 3 illustrates this sequence:

Kibot: What would happen to elements in this system if ocean temperature increases?

S4: Higher temperature would influence the habitat because fish only lives in a certain temperature range.

S5: Agreed, so fish would die off if the temperature is too high. Higher temperature also influences the habitat by releasing more O_2.

Table 2.4

Top 5 Sequences per Condition, by Lift Values

Less-knowledgeable-peer agent	Lift	Support	Confidence	n (peer)	n (expert)
Question => Externalize; Question	1.54	.33	.60	12	6
Reasoning; Responsiveness => Transactive	1.50	.28	1.00	10	8
Question => Transactive	1.50	.28	1.00	10	0
Transactive (multiple) => Question	1.50	.28	1.00	10	2
Responsiveness; Externalize => Transactive	1.50	.33	1.00	12	16
Expert agent	Lift	Support	Confidence	n (peer)	n (expert)
Question => Question	1.73	.29	.71	12	10
Responsiveness; Externalize => Transactive	1.42	.47	1.00	12	16
Externalize; Responsiveness => Transactive	1.42	.29	1.00	10	10
Externalize; Transactive; Responsiveness => Transactive	1.42	.35	1.00	6	12
Responsiveness (multiple) => Transactive	1.30	.35	1.00	12	12

Notes. ";" = co-occurring moves within a message; "=>" shows sequences. Lift = probability of occurrences, relative to chance. Support = proportion of occurrences. Confidence = likelihood of pattern B to follow A, once A occurred.

Table 2.5

Wilcoxon Results for Between-Agent Differences in Sequence Occurrences

	Code	Mpeer	SDpeer	Mexpert	SDexpert	W	*p*	adjusted *p*
1	Question => Externalize; Question	.67	.49	.33	.59	216	.05	.09
2	Reasoning; Responsiveness => Transactive	.56	.51	.47	.62	170	.53	.71
3	Question => Transactive	.56	.62	0	0	224	.0004***	.003**
4	Transactive (multiple) => Question	.56	.51	.17	.38	206	.01*	.03*
5	Responsiveness; Externalize => Transactive	.67	.59	1.00	.37	99	.05	.09
6	Question => Question	.67	.59	.63	.50	147	.92	1.00
7	Externalize; Responsiveness => Transactive	.56	.70	.63	.50	128	.55	.71
8	Externalize; Transactive; Responsiveness => Transactive	.33	.49	.75	.45	84	.02*	.06
9	Responsiveness (multiple) => Transactive	.67	.49	.67	.49	162	1.00	1.00

Notes: *: .05, **: .01, ***: .001. "Respond agent; Externalize => Transactive" appeared for both conditions

Wilcoxon tests provide evidence of differences in the occurrences of questioning sequences between agent conditions (questioning -> transactive; M_{peer} = .56, SD_{peer} = .62; M_{expert} = 0, SD_{expert} = 0; adjusted p = .003; transactive (multiple) -> questioning; M_{peer} = .56, SD_{peer} = .51; M_{expert} = .17, SD_{expert} = .38; adjusted p = .03). Table 2.5 presents the results.

The top sequences for each agent were consistent regardless of the agent student groups started with (Appendix 2.4). In groups who started with or switched to the less-knowledgeable-peer agent during the lesson, the top sequences with this agent primarily involved questioning. In comparison, most sequences with the expert agent included transactive and responsive moves. These results align with patterns from the overall sample and suggest that randomizing the starting agents and switching agents did not substantially impact interaction sequences.

In sum, the sequences illustrate interesting patterns, such as when groups in the less-knowledgeable-peer condition posed questions, explained ideas, and posed questions again. Such sequences may suggest "quizzing" behaviors, one that may facilitate knowledge construction.

Discussion

Designing for Knowledge Construction

This study explored how student groups demonstrated reasoning, transactive exchange, and responsiveness in interaction with one another and with two conversational agents. Claim-making, externalizing one's ideas, and responsiveness to the agents were the most frequent moves. Findings about students' responsiveness are noteworthy because student groups in prior work tended to ignore learning agents (Kumar et al., 2010). A potential explanation for this finding is that the content of the agents' prompts adapted to group discussions to maintain conversation coherence, and thus, responding to the agents may have been less disruptive. Consider the following excerpt:

S6: the **fish** eat **zooplankton**

S6: **zooplankton** gets eaten by **fish**

Kibot: What would happen to other elements in this system if **whales** decrease?

S7: if the **whales** go extinct there will be more **fish**

S6: do **sharks** eat **fish**

S8: why would **whales** eat **fish**

S7: **whales** eat **plankton** too

Here, the students started their concept map by linking fish and zooplankton. The agent kept track of those connections and cued "whales", an element linked to both "fish" and "zooplankton". The conversation flowed naturally to include the target relationships (whales -> fish; whales -> zooplankton) and other predators to "fish" such as "sharks". As another example, if after five talk turns following the hint about "whales", students have not created the intended connections, the agent will ask more explicit prompts, e.g., "Would fish increase or decrease if there were more whales?" The missing connections are linked to recently created terms in the concept map ("fish"), thus maintaining conversation coherence.

Adaptivity also manifests in agents' awareness of ongoing group dynamics. The agents keep track of each student's participation rates to engage the least active students, e.g., "S2, help me out. Do you agree or disagree with your friends?". Recognition of group dynamics positions agents as ingroup members and reduces users' potential antagonistic treatment of the agents, such as ignoring or abusing the agents (Sebo et al., 2020).

Furthermore, the combination of visual and verbal cues may have "humanized" the agents and made the interactions more engaging (Feine et al., 2019; Go & Sundar, 2019; Muresan & Pohl, 2019). The agents assume roles as an expert or a less-knowledgeable peer and have human-like visual features such as eyes, expressions, and clothing to reflect these profiles. For example, the expert agent wears an "expert" tag on a white coat, while the less-knowledgeable-peer agent shows dynamic expressions such as confusion and excitement to reflect its younger profile (Figure 2.1). These visual cues can be helpful because people tend to reason with anthropomorphized agents in ways they act with humans (Malle et al., 2016; Nass & Moon, 2000). Users may subsequently take on responsibilities as a tutor to the less-knowledgeable-peer agent or a tutee to the expert agent.

Previous agents in learning domains have mostly focused on conceptual prompts (Heidig & Clarebout, 2011; Y. Kim & Baylor, 2016). The agents in this study employ additional social verbal cues, including small talk, acknowledgment, uncertainty, and opinion conformity (Table 2.1). In expressing uncertainty, for instance, the less-knowledgeable-peer agent reveals its vulnerability to invite for students' idea elaboration. Students may be more willing to respond to the agents when they recognize the agents' humanlike characteristics through the social cues and position the agents as conversational partners (Feine et al., 2019).

Questioning Sequences with the Less-Knowledgeable-Peer Agent

Examination of sequences with high lift values suggests that student groups enacted distinct conversational norms with the two agents. Responses to the expert agent were followed by transactive exchange, similar to how students might react to a teacher's prompts. Meanwhile, several groups showed "quizzing" behaviors with the less-knowledgeable-peer agent. The questions that students raised in interaction with the agents resemble Socratic questioning, which aims to probe others' thinking and determine what is known and not known (Paul & Elder, 2007). This style of communication has often been adopted by teachers as a form of instructional scaffolding to generate students' exploratory discourse and scientific inquiries (Hogan & Pressley, 1997).

The questioning sequences resemble interactions that support learning in peer tutoring. Tutors rely on elaborated explanations and questioning to guide tutees' thinking (Graesser & Person, 1994). Excerpts from the current work illustrate how questioning sequences might involve multiple students, instead of one student's interactions with the agent. Prior work with one-to-one intelligent tutors has employed the learn-through-teaching framework, where students iteratively taught and tested CAs (Biswas et al., 2010, 2016; Kinnebrew et al., 2014). These patterns can be extended to collaborative settings, where efforts to teach the agent no longer rest with an individual. Through these efforts, groups coordinate attention around the same concepts.

Questioning can benefit student tutors by providing opportunities for tutors to reflect on their knowledge and move towards knowledge construction (Berghmans et al., 2013; King et al., 1998). Excerpt 2 in this study exemplifies complex connections that span across systems elements in the concept map. The small number of sequences that contained questioning in this study often contained elaborated explanations from the questioners and group members. Sustained sharing of queries and explanations contributes to more complex scientific reasoning over time (Hogan et al., 1999).

Currently, the agents only respond to students' questions with general answers such as "I'm not sure. Can someone in the group explain?" to prompt for elaboration. Future iterations can introduce a mix of general prompts and responses that contain alternative conceptions to engage students in idea elaboration. These prompts can trigger episodes of tutor's questioning, guidance, and feedback, which can enhance tutor's learning gains (P. A. Cohen et al., 1982; Graesser et al., 1995).

Limitations and Future Directions

The limitations of this study can guide future investigations. First, analyses were at the group instead of individual level. Agent designs may differ with group dynamics and individual behaviors (Sebo et al., 2020). In a case study with the same Kibot agents (Study 3), I found that groups composed of students with expanding, mixed, and emergent prior science understanding interacted with the agents differently. The emergent group primarily showed responsiveness to the expert agent, whereas the expanding group showed more transactive exchange with the less-knowledgeable-peer agent. Incorporating group and individual dynamics into sequential pattern mining would require more complex analyses, such as multilevel models with large samples. Future studies can employ these analyses and consider additional variables that may influence knowledge construction, such as domain understanding, perceptions of agents, and participation tendency. Second, the within-subject design did not allow for linking interaction sequences to learning outcomes. Future research can apply between-subject designs to explore the pathways between agents, reasoning and transactivity, and learning performance.

Third, I used timestamps to determine the window size for sequential pattern mining and found that the notable sequential patterns were consistent across window sizes. An alternate segmentation approach can be based on idea units in the group chats. For example, if one segmented a unit around the idea of the food web, we might see longer sequences of claim-making. However, this approach may also be problematic in this study's context, because students were talking about related connections. Consecutive utterances such as "fish eat plankton", "fish consume kelp", and "more zooplankton decreases phytoplankton" could be treated as three separate idea units or one unit of related links. Thus, future work could experiment with other approaches to segment discourse for sequence mining.

Overall, findings broaden the design space for learning technologies in collaborative contexts. The sequences uncovered by this research can be applied towards developing agent adaptivity. Future systems can adapt designs to promote appropriate discussion moves and sequences, given dynamic learning goals and groups' knowledge and collaboration states.

Furthermore, future work can examine longitudinal interactions with agents to determine whether the observed interaction sequences can be sustained, and their implications for learning.

Conclusion

This study examines the design of two agent prototypes to facilitate students' reasoning, transactivity, and responsiveness to the agents. Findings illustrate that design tweaks in agents' appearances and linguistic styles can facilitate different discussion sequences, including groups' questioning and explaining to the less-knowledgeable-peer agent. These sequences can promote students' reflection and idea elaboration. As intelligent systems become increasingly prevalent in collaborative learning, we need to consider how student groups interact with such systems. This work illustrates how embedding humanness in agents' designs, such as in the form of a less knowledgeable peer or an expert, can support this vision.

References

Adamson, D., Dyke, G., Jang, H., & Rosé, C. P. (2014). Towards an agile approach to adapting dynamic collaboration support to student needs. *International Journal of Artificial Intelligence in Education*, *24*(1), 92–124. https://doi.org/10.1007/s40593-013-0012-6

Almasi, J. F. (1995). The nature of fourth graders' sociocognitive conflicts in peer-led and teacher-led discussions of literature. *Reading Research Quarterly*, 314–351.

Bazaldua, D. L., Baker, R., & Pedro, M. O. (2014). Comparing expert and metric-based assessments of association rule interestingness. *Educational Data Mining 2014*.

Berghmans, I., Neckebroeck, F., Dochy, F., & Struyven, K. (2013). A typology of approaches to peer tutoring. Unraveling peer tutors' behavioural strategies. *European Journal of Psychology of Education*, *28*(3), 703–723. https://doi.org/10.1007/s10212-012-0136-3

Biswas, G., Jeong, H., Kinnebrew, J. S., Sulcer, B., & Roscoe, R. O. D. (2010). Measuring self-regulated learning skills through social interactions in a teachable agent environment. *Research and Practice in Technology Enhanced Learning*, *5*(02), 123–152.

Biswas, G., Segedy, J. R., & Bunchongchit, K. (2016). From Design to Implementation to Practice a Learning by Teaching System: Betty's Brain. *International Journal of Artificial Intelligence in Education*, *26*(1), 350–364. https://doi.org/10.1007/s40593-015-0057-9

Buchta, C., Hahsler, M., Diaz, D., Buchta, M. C., & Zaki, M. J. (2020). *Package 'arulesSequences.'*

Byrne, D., & Nelson, D. (1965). Attraction as a linear function of proportion of positive reinforcements. *Journal of Personality and Social Psychology*, *1*(6), 659–663. https://doi.org/10.1037/h0022073

Chen, B., Resendes, M., Chai, C. S., & Hong, H.-Y. (2017). Two tales of time: uncovering the significance of sequential patterns among contribution types in knowledge-building discourse. *Interactive Learning Environments*, *25*(2), 162–175. https://doi.org/10.1080/10494820.2016.1276081

Chen, H., Park, H. W., & Breazeal, C. (2020). Teaching and learning with children: Impact of reciprocal peer learning with a social robot on children's learning and emotive engagement. *Computers and Education*, *150*, 103836. https://doi.org/10.1016/j.compedu.2020.103836

Cohen, P. A., Kulik, J. A., & Kulik, C.-L. C. (1982). Educational Outcomes of Tutoring: A Meta-analysis of Findings. *American Educational Research Journal*, *19*(2), 237–248. https://doi.org/10.3102/00028312019002237

Dillenbourg, P. (2002). *Over-scripting CSCL: The risks of blending collaborative learning with instructional design.* Citeseer.

Diziol, D., Walker, E., Rummel, N., & Koedinger, K. R. (2010). Using intelligent tutor technology to implement adaptive support for student collaboration. In *Educational Psychology Review* (Vol. 22, Issue 1, pp. 89–102). Springer. https://doi.org/10.1007/s10648-009-9116-9

Dyke, G., Adamson, D., Howley, I., & Rose, C. P. (2013). Enhancing scientific reasoning and discussion with conversational agents. *IEEE Transactions on Learning Technologies*, *6*(3), 240–247. https://doi.org/10.1109/TLT.2013.25

Dyke, G., Howley, I., Adamson, D., Kumar, R., & Rosé, C. P. (2013). Towards Academically Productive Talk Supported by Conversational Agents. In *Productive Multivocality in the Analysis of Group Interactions* (pp. 459–476). Springer US. https://doi.org/10.1007/978-1-4614-8960-3_25

Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2019). A Taxonomy of Social Cues for Conversational Agents. *International Journal of Human-Computer Studies*, *132*, 138–161. https://doi.org/10.1016/J.IJHCS.2019.07.009

Fiacco, J., & Rosé, C. (2018). Towards domain general detection of transactive knowledge building behavior. *Proceedings of the Fifth Annual ACM Conference on Learning at Scale*, 1–11.

Gillies, R. M. (2014). Dialogic interactions in the cooperative classroom. *International Journal of Educational Research*, *76*, 178–189. https://doi.org/10.1016/j.ijer.2015.02.009

Go, E., & Sundar, S. S. (2019). Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. *Computers in Human Behavior*, *97*, 304–316. https://doi.org/10.1016/j.chb.2019.01.020

Gong, L. (2008). How social is social responses to computers? The function of the degree of anthropomorphism in computer representations. *Computers in Human Behavior*, *24*(4), 1494–1509. https://doi.org/10.1016/j.chb.2007.05.007

Graesser, A. C. (2016). Conversations with AutoTutor Help Students Learn. *International Journal of Artificial Intelligence in Education*, *26*(1), 124–132. https://doi.org/10.1007/s40593-015-0086-4

Graesser, A. C., & Person, N. K. (1994). Question Asking During Tutoring. *American Educational Research Journal*, *31*(1), 104–137. https://doi.org/10.3102/00028312031001104

Graesser, A. C., Person, N. K., & Magliano, J. P. (1995). Collaborative dialogue patterns in naturalistic one-to-one tutoring. *Applied Cognitive Psychology*, *9*(6), 495–522. https://doi.org/10.1002/acp.2350090604

Heidig, S., & Clarebout, G. (2011). Do pedagogical agents make a difference to student motivation and learning? In *Educational Research Review* (Vol. 6, Issue 1, pp. 27–54). Elsevier. https://doi.org/10.1016/j.edurev.2010.07.004

Hewitt, J., & Scardamalia, M. (1998). Design Principles for Distributed Knowledge Building Processes. *Educational Psychology Review*, *10*(1), 75–96. https://doi.org/10.1023/A:1022810231840

Hmelo-Silver, C. E., & Barrows, H. S. (2008). Facilitating Collaborative Knowledge Building. *Cognition and Instruction*, *26*(1), 48–94. https://doi.org/10.1080/07370000701798495

Hoadley, C. M., & Kilner, P. G. (2005). Using technology to transform communities of practice into knowledge-building communities. *ACM SIGGROUP Bulletin*, *25*(1), 31–40. https://doi.org/10.1145/1067699.1067705

Hogan, K., Nastasi, B. K., & Pressley, M. (1999). Discourse Patterns and Collaborative Scientific Reasoning in Peer and Teacher-Guided Discussions. *Cognition and Instruction*, *17*(4), 379–432. https://doi.org/10.1207/S1532690XCI1704_2

Hogan, K., & Pressley, M. (1997). Scaffolding scientific competencies within classroom communities of inquiry. In *Scaffolding student learning: Instructional approaches and issues.* (pp. 74–107). Brookline Books.

Honnibal, M., & Montani, I. (2017). spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. *To Appear*, *7*(1), 411–420.

Howley, I., Kumar, R., Mayfield, E., Dyke, G., & Rosé, C. P. (2013). Gaining Insights from Sociolinguistic Style Analysis for Redesign of Conversational Agent Based Support for Collaborative Learning. In *Productive Multivocality in the Analysis of Group Interactions* (pp. 477–494). Springer US. https://doi.org/10.1007/978-1-4614-8960-3_26

Jeong, H., Gupta, A., Roscoe, R., Wagster, J., Biswas, G., & Schwartz, D. (2008). Using hidden markov models to characterize student behaviors in learning-by-teaching environments. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *5091 LNCS*, 614–625. https://doi.org/10.1007/978-3-540-69132-7_64

Kang, H., Thompson, J., & Windschitl, M. (2014, July 1). Creating Opportunities for Students to Show What They Know: The Role of Scaffolding in Assessment Tasks. *Science Education*, *98*(4), 674–704. https://doi.org/10.1002/sce.21123

Kim, S., Lee, J., & Gweon, G. (2019). Comparing data from chatbot and web surveys effects of platform and conversational style on survey response quality. *Conference on Human Factors in Computing Systems - Proceedings*, 1–12. https://doi.org/10.1145/3290605.3300316

Kim, Y. (2007). Desirable Characteristics of Learning Companions. *International Journal of Artificial Intelligence in Education*, *17*(4), 371–388.

Kim, Y., & Baylor, A. L. (2006). A social-cognitive framework for pedagogical agents as learning companions. *Educational Technology Research and Development*, *54*(6), 569–596. https://doi.org/10.1007/s11423-006-0637-3

Kim, Y., & Baylor, A. L. (2016). Research-Based Design of Pedagogical Agent Roles: A Review, Progress, and Recommendations. *International Journal of Artificial Intelligence in Education*, *26*(1), 160–169. https://doi.org/10.1007/s40593-015-0055-y

King, A., Staffieri, A., & Adelgais, A. (1998). Mutual Peer Tutoring: Effects of Structuring Tutorial Interaction to Scaffold Peer Learning. *Journal of Educational Psychology*, *90*(1), 134–152. https://doi.org/10.1037/0022-0663.90.1.134

Kinnebrew, J. S., Segedy, J. R., & Biswas, G. (2014). Analyzing the temporal evolution of students' behaviors in open-ended learning environments. *Metacognition and Learning*, *9*(2), 187–215. https://doi.org/10.1007/S11409-014-9112-4

Kumar, R, Beuth, J., & Rose, C. (2011). Conversational Strategies that Support Idea Generation Productivity in Groups. In H. Spada, G. Stahl, N. Miyake, & N. Law (Eds.), *Connecting Computer-Supported Collaborative Learning to Policy and Practice: CSCL2011 Conference Proceedings. Volume I - Long Papers* (pp. 398–405). International Society of the Learning Sciences. https://repository.isls.org/handle/1/2475

Kumar, Rohit, Ai, H., Beuth, J. L., & Rosé, C. P. (2010). Socially capable conversational tutors can be effective in collaborative learning situations. *Lecture Notes in Computer Science*, *6094 LNCS*(PART 1), 156–164. https://doi.org/10.1007/978-3-642-13388-6_20

Lee, K. M., & Nass, C. (2003). Designing social presence of social actors in human computer interaction. *Conference on Human Factors in Computing Systems - Proceedings*, 289–296. https://doi.org/10.1145/642611.642662

Liew, T. W., Tan, S.-M., & Jayothisa, C. (2013). The Effects of Peer-Like and Expert-Like Pedagogical Agents on Learners' Agent Perceptions, Task-Related Attitudes, and Learning Achievement. *Educational Technology & Society*, *16*(4), 275–286. https://eric.ed.gov/?id=EJ1013641

Malle, B. F., Scheutz, M., Forlizzi, J., & Voiklis, J. (2016). Which robot am I thinking about? The impact of action and appearance on people's evaluations of a moral robot. *ACM/IEEE International Conference on Human-Robot Interaction*, *2016-April*, 125–132. https://doi.org/10.1109/HRI.2016.7451743

McKinley, D. C., Miller-Rushing, A. J., Ballard, H. L., Bonney, R., Brown, H., Cook-Patton, S. C., Evans, D. M., French, R. A., Parrish, J. K., & Phillips, T. B. (2017). Citizen science can improve conservation science, natural resource management, and environmental protection. *Biological Conservation*, *208*, 15–28.

Merceron, A., & Yacef, K. (2008). Interestingness measures for association rules in educational data. *Educational Data Mining 2008*.

Moon, Y., & Nass, C. I. (1996). Adaptive agents and personality change. *Conference Companion on Human Factors in Computing Systems Common Ground - CHI '96*, 287–288. https://doi.org/10.1145/257089.257325

Muhonen, H., Rasku-Puttonen, H., Pakarinen, E., Poikkeus, A. M., & Lerkkanen, M. K. (2017). Knowledge-building patterns in educational dialogue. *International Journal of Educational Research*, *81*, 25–37. https://doi.org/10.1016/j.ijer.2016.10.005

Muresan, A., & Pohl, H. (2019). Chats with bots: Balancing imitation and engagement. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–6.

Nass, C., & Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues*, *56*(1), 81–103. https://doi.org/10.1111/0022-4537.00153

Nass, C., Moon, Y., & Green, N. (1997). Are machines gender neutral? Gender-stereotypic responses to computers with voices. *Journal of Applied Social Psychology*, *27*(10), 864–876. https://doi.org/10.1111/j.1559-1816.1997.tb00275.x

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 72–78.

Nguyen, H., & Santagata, R. (2020). Impact of computer modeling on learning and teaching systems thinking. *Journal of Research in Science Teaching*, tea.21674. https://doi.org/10.1002/tea.21674

Paul, R., & Elder, L. (2007). Critical Thinking: The Art of Socratic Questioning. *Journal of Developmental Education*, *31*(1), 36–37. https://www.proquest.com/scholarly-journals/critical-thinking-art-socratic-questioning/docview/228487383/se-2?accountid=14509

Pearson, J., Hu, J., Branigan, H. P., Pickering, M. J., & Nass, C. I. (2006). Adaptive language behavior in HCI. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '06*, 1177. https://doi.org/10.1145/1124772.1124948

Romero, O., Zhao, R., & Cassell, J. (2017). Cognitive-inspired conversational-strategy reasoner for socially-aware agents. *IJCAI'17: Proceedings of the 26th International Joint Conference on Artificial Intelligence*, 3807–3813. https://dl.acm.org/doi/abs/10.5555/3172077.3172421

Roschelle, J., & Teasley, S. D. (1995). The Construction of Shared Knowledge in Collaborative Problem Solving. In *Computer Supported Collaborative Learning* (pp. 69–97). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-85098-1_5

Roscoe, R. D., & Chi, M. T. H. (2007). Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors' explanations and questions. *Review of Educational Research*, *77*(4), 534–574. https://doi.org/10.3102/0034654307309920

Rosé, C., Wang, Y. C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., & Fischer, F. (2008). Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning*, *3*(3), 237–271. https://doi.org/10.1007/s11412-007-9034-0

Rosenberg-Kima, R. B., Baylor, A. L., Plant, E. A., & Doerr, C. E. (2008). Interface agents as social models for female students: The effects of agent visual presence and appearance on female students' attitudes and beliefs. *Computers in Human Behavior*, *24*(6), 2741–2756. https://doi.org/10.1016/j.chb.2008.03.017

Scardamalia, M., & Bereiter, C. (1991). Higher Levels of Agency for Children in Knowledge Building: A Challenge for the Design of New Knowledge Media. *Journal of the Learning Sciences*, *1*(1), 37–68. https://doi.org/10.1207/s15327809jls0101_3

Scardamalia, M., & Bereiter, C. (1993). Technologies for knowledge-building discourse. *Communications of the ACM*, *36*(5), 37–41. https://doi.org/10.1145/155049.155056

Sebo, S., Stoll, B., Scassellati, B., & Jung, M. F. (2020). Robots in Groups and Teams: A Literature Review. *Proceedings of the ACM on Human-Computer Interaction*, *4*(CSCW2), 36. https://doi.org/10.1145/3415247

Stahl, G., Cress, U., Law, N., & Ludvigsen, S. (2014). Analyzing the multidimensional construction of knowledge in diverse contexts. In *International Journal of Computer-Supported Collaborative Learning* (Vol. 9, Issue 1, pp. 1–6). Springer New York LLC. https://doi.org/10.1007/s11412-014-9189-4

Teasley, S. D. (1997). Talking About Reasoning: How Important Is the Peer in Peer Collaboration? In *Discourse, Tools and Reasoning* (pp. 361–384). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-03362-3_16

Tegos, S., & Demetriadis, S. (2017). Conversational Agents Improve Peer Learning through Building on Prior Knowledge. *Educational Technology & Society*, *20*(1), 99–111.

van Aalst, J. (2009). Distinguishing knowledge-sharing, knowledge-construction, and knowledge-creation discourses. *International Journal of Computer-Supported Collaborative Learning*, *4*(3), 259–287. https://doi.org/10.1007/s11412-009-9069-5

Walker, E., Rummel, N., & Koedinger, K. R. (2014). Adaptive intelligent support to improve peer tutoring in algebra. *International Journal of Artificial Intelligence in Education*, *24*(1), 33–61. https://doi.org/10.1007/s40593-013-0001-9

Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers and Education*, *46*(1), 71–95. https://doi.org/10.1016/j.compedu.2005.04.003

Wen, M., Maki, K., Wang, X., Dow, S. P., Herbsleb, J., & Rose, C. (2016). Transactivity as a Predictor of Future Collaborative Knowledge Integration in Team-Based Learning in Online Courses. In *International Educational Data Mining Society*. International Educational Data Mining Society. e-mail: admin@educationaldatamining.org; Web site: http://www.educationaldatamining.org. https://www.edx.org/course/medicinal-chemistry-

Wise, A. F., & Chiu, M. M. (2011). Analyzing temporal patterns of knowledge construction in a role-based online discussion. *International Journal of Computer-Supported Collaborative Learning*, *6*(3), 445–470. https://doi.org/10.1007/s11412-011-9120-1

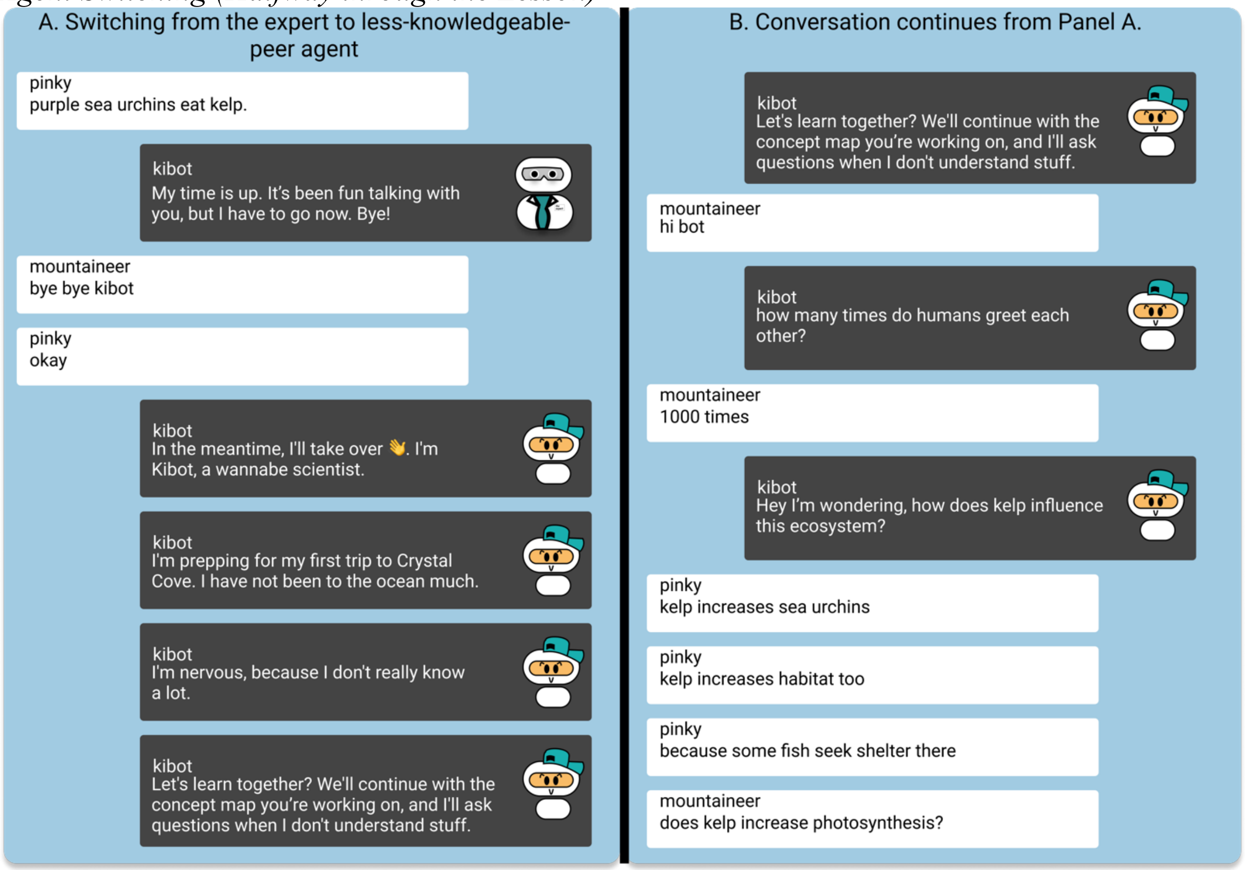
Zaki, M. J. (2000). Sequence mining in categorical domains: incorporating constraints. *Proceedings of the Ninth International Conference on Information and Knowledge Management*, 422–429.

Zhou, M. X., Mark, G., Li, J., & Yang, H. (2019). Trusting virtual agents: The effect of personality. *ACM Transactions on Interactive Intelligent Systems*, *9*(2–3), 1–36. https://doi.org/10.1145/3232077

Appendix

Appendix 2.1

Agent Switching (Halfway through the Lesson)



Notes. The within-subject design switches agents halfway through the activity so that each student group interacts with both agents. In entering the conversation, the agents reintroduce the learning task so that groups stay on task. The example conversation in the figure illustrates that students acknowledge the agent switching (panel A) and are aware of the task (build the concept map; panel B). The less-knowledgeable-agent Kibot also guides students back to the task (e.g., "Hey I'm wondering, how does kelp influence this ecosystem?")

Appendix 2.2

Proportions of Discussion Moves for Groups Starting with Less-knowledgeable-peer versus Expert Agent

Code	M1	SD1	M2	SD2	Mann-Whitney's U	p	adjusted-p
Interactions with the less-knowledgeable-peer agent							
Reasoning							
Claim	.37	.14	.32	.18	162	1	1
Reasoning	.13	.15	.27	.23	136	.78	1
Questioning	.28	.29	.16	.09	98	1	1
Transactivity							
Externalizing (self's ideas)	.45	.16	.33	.14	162	1	1
Transactive (prior ideas, friend)	.22	.14	.21	.08	153	.78	1
Responsiveness	.25	.17	.24	.14	136	.78	1
Interactions with the expert agent							
Reasoning							
Claim	.41	.22	.30	.19	161	.57	.78
Reasoning	.16	.09	.16	.22	144	.55	.78
Questioning	.34	.41	.19	.19	44	.28	.78
Transactivity							
Externalizing (self's ideas)	.38	.09	.31	.11	144	.55	.78
Transactive (prior ideas, friend)	.18	.13	.21	.12	153	.78	.78
Responsiveness	.20	.11	.26	.15	105	.77	.78

Notes. M1 = groups starting with the less-knowledgeable-peer agent, M2 = starting with the expert agent. This table shows that there was no difference in the ratio of occurrence for discussion moves in interactions with each agent (less knowledgeable peer and expert) between M1 and M2 (i.e., groups that started with either agent condition).

Appendix 2.3

Robustness Checks: Top 5 Discussion Sequences for Different Window Sizes

2-minute window	3-minute window	5-minute window
Less-knowledgeable-peer agent		
Externalize => Transactive	Question => Externalize	Externalize (multiple) => Transactive
Responsiveness => Transactive	Transactive => Responsiveness	Responsiveness => Question
Question => Externalize; Question	Question => Transactive	Question => Transactive
Question => Transactive	Externalize (multiple) => Transactive	Externalize => Transactive
Transactive (multiple) => Question	Responsiveness => Transactive	Responsiveness => Externalize
Expert agent		
Question -> Question	Responsiveness; Question => Question	Externalize; Transactive => Transactive
Externalize; Responsiveness => Transactive	Externalize; Responsiveness => Responsiveness	Externalize; Transactive => Responsiveness
Externalize => Responsiveness	Externalize; Transactive => Transactive	Externalize; Responsiveness => Externalize
Responsiveness (multiple) => Transactive	Externalize (multiple) => Transactive	Externalize; Responsiveness = > Transactive
Externalize => Transactive	Responsiveness (multiple) => Transactive	Externalize => Responsiveness; Reasoning

Notes. Sequences are organized by decreasing lift values (minimum lift values >1).

Appendix 2.4

Top Sequences for Groups Starting with Less-knowledgeable-peer versus Expert Agent

Interactions with the less-knowledgeable-peer (LKP) agent	
Start with LKP agent	**Start with expert agent and switch to LKP agent**
1. Question; Externalize => Question	1. Responsiveness; Externalize => Transactive
2. Question => Transactive	2. Transactive (multiple) => Question
3. Responsiveness => Reasoning	3. Question; Externalize => Question
4. Transactive (multiple) => Question	4. Responsiveness; Externalize => Question
5. Reasoning; Responsiveness => Transactive	5. Question (multiple) => Externalize
Interactions with the expert agent	
Start with LKP agent and switch to expert	**Start with expert agent**
1. Responsiveness; Externalize => Transactive	1. Externalize => Externalize
2. Externalize; Responsiveness => Transactive	2. Externalize => Transactive
3. Transactive => Responsiveness	3. Externalize; Reasoning => Responsiveness
4. Externalize; Transactive => Responsiveness	4. Transactive => Responsiveness
5. Responsiveness (multiple) => Transactive	5. Externalize; Transactive; Responsiveness => Transactive

Notes. Top 5 sequences per group ordered by highest lift (minimum threshold of 1), support (minimum threshold of .25), and confidence values (minimum threshold of .5). Lift = probability of sequence occurrences, compared to chance. Support = proportion of occurrences. Confidence = likelihood of pattern B to follow A, once A occurred.

CHAPTER 3 CONVERSATIONAL REPAIRS WITH AGENTS

An earlier version of this chapter will appear in:
Nguyen, H. (June, 2022). Learners' Reactions to Chatbot Communication Breakdowns: Insights into Fostering Learning. In *2nd Annual Meeting of the International Society of the Learning Sciences.* (Online). International Society of the Learning Sciences.

Abstract

Text-based conversational agents (i.e., chatbots) can promote productive discussions and learning. However, agents sometimes fail to interpret users' intent, leading to communication breakdowns. To alleviate these issues, researchers have experimented with agent designs that may increase users' willingness to fix communication breakdowns. This study explores the affordances of varying agent designs (as a less knowledgeable peer versus an expert) to prompt learners' strategies to repair breakdowns. Data sources drew from the chat logs of 18 groups (52 high school learners) in science discussions with one another and with the agents. Results suggest that groups rephrased and provided explanations during communication breakdowns with the peer agent more often than the expert version. There existed a positive association between how often groups explained to the agents during breakdowns and indicators of deeper learning in the group chats. Findings illustrate the affordances of designs that enact certain communication strategies—ones that also enrich learning.

Introduction

Text-based pedagogical agents (i.e., chatbots) have shown promise in facilitating discussions to promote productive talk moves and enrich learning (Dyke, Howley, et al., 2013; S. Kim et al., 2020). These systems apply natural language processing of textual and speech input to provide scaffolds for idea articulation, knowledge construction, and argumentation (Y. Kim & Baylor, 2016). Despite their promises, researchers have documented cases where users ignore the agents' hints or abandon the conversations (Kumar et al., 2010). Communication breakdowns are a potential reason for such disengagement. Breakdowns occur when the agents fail to interpret the users' intent and respond appropriately. Breakdowns can reduce users' trust in the systems and willingness to interact with agents (Chaves & Gerosa, 2019).

To alleviate breakdowns, researchers have explored ways to improve the agents' natural language understanding and designed responses to increase users' willingness to correct the agents when the conversations fail (Adamson et al., 2014; Ashktorab et al., 2019; Engelhardt et al., 2017; M. K. Lee et al., 2010). Another strategy is to develop agent designs with varied personalities and characteristics to prime users' expectations for social interactions (Biswas et al., 2016; Chaves & Gerosa, 2019; Y. Kim & Baylor, 2016; S. Lee et al., 2019).

I built on this prior work to develop two agent designs (a less knowledgeable peer and an expert). The less-knowledgeable-peer agent enacted a learn-by-teaching paradigm, where learners enhanced knowledge through idea explanation to the agent (Biswas et al., 2016). Meanwhile, the expert persona resembled an authority figure to monitor the discussion (Y. Kim, 2007; Y. Kim & Baylor, 2016). The agents were integrated into a group modeling activity as part of a high school environmental science project. The agents scaffolded learners to collaboratively reason through connections within a marine ecosystem (described further in the Methodology section). As learners chatted with one another and responded to the agents, breakdowns might happen when the agents could not parse the chat intent correctly. For example, in response to clarification requests from learners ("What do you mean?"), the agents might repeat the main learning task instead of clarifying a prior utterance.

In this study, I explored the strategies that learners used when such breakdowns occurred. I further examined how these strategies differed in group interactions with the two agent designs. Understanding how learner groups approached communication breakdowns with the different designs provides insights for agent development to sustain learner engagement and enrich learning.

Background

Human-agent Collaboration in Learning Interactions

The design of the agents in this study is guided by knowledge building frameworks, which propose that learners construct increasingly coherent understanding through idea sharing and critiques of group artifacts (Scardamalia & Bereiter, 1993). Agents can serve as facilitators of knowledge building. They can prompt learners who are less active to voice opinions, identify areas of misunderstanding, and ask for elaboration of peers' ideas to advance the group's knowledge. These discourse practices have been linked to enhanced learning (Dyke, Adamson, et al., 2013; S. Kim et al., 2020). In previous work (Study 1), I found that learners were overall responsive to agents' prompts, and these prompts advanced the group discussion towards idea elaboration.

Agents can also directly collaborate with learners, such as when agents offer ideas and learners critique the agent's understanding, pose questions, or provide explanations to the agents. Research on one-on-one interactions between a student and a pedagogical agent suggests that such human-agent interactions can contribute to learning (Biswas et al., 2016; Graesser, 2016). As an example, Biswas and colleagues (2016) found that learners in middle school classrooms deepened learning from interactions with a less knowledgeable tutee agent, where the learners iteratively taught and quizzed the agent to gauge their understanding.

In sum, collaboration in agent-facilitated group discussions can occur through both human-human and human-agent interactions. Prior research has focused on questions such as how agents facilitate interactions among human users or how individuals reacted to the agents in one-on-one exchange. However, to avoid situations where groups abuse or ignore the agents (Kumar et al., 2010), it is important to examine group-agent collaboration, particularly how groups position the agents in knowledge building processes. The current study taps into this question, with particular regards to how groups react to breakdowns in student-agent communication.

Breakdowns and User Repair Strategies

Despite their promises, agents can encounter breakdowns, when the agents cannot comprehend the language interactions in users' input and fail to complete the expected tasks (Feine et al., 2019; Følstad et al., 2018; Gnewuch et al., 2017). Breakdowns can have a detrimental impact on users' interactions with the agents. Users may reduce their trust in the agents, become less willing to continue using them, or abandon the tasks altogether (Luger & Sellen, 2016). Breakdowns can worsen issues already observed in interactions between student groups and the agents, such as when groups ignore the agent's hints (Kumar et al., 2010).

However, breakdowns also provide opportunities for learners to reflect on gaps in their understanding and engage in knowledge building (Roscoe & Chi, 2007). When an agent misinterprets an utterance because it represents a conceptual misunderstanding that the agent does not recognize, learners can self-repair the misunderstanding or turn to others in the group to co-construct more coherent explanations. Researchers have documented a range of repair strategies in human-human and human-computer interactions (Ashktorab et al., 2019; Brennan, 1998; Clark

& Brennan, 1991; Yu et al., 2016). For instance, the speakers (i.e., users) can initiate repairs by repeating ideas, rephrasing the utterances, or explaining their intent (Clark & Brennan, 1991).

These repair strategies have different implications for learning. Simply repeating or paraphrasing ideas do not produce new knowledge (Teasley, 1995). However, learners enrich understanding through constructive activities such as self-explaining or explaining to others (Chi, 2000; Davis, 2010; Roscoe & Chi, 2007). These activities encourage learners to collaboratively integrate new knowledge with old knowledge and link concepts from different sources, thus repairing existing knowledge and broadening understanding (Chi, 2009; Davis, 2010). These knowledge creation processes can make individuals' and groups' knowledge more structured.

Repair Strategies and Learning Implications for Systems Thinking

Explaining ideas can enrich systems thinking, defined as the ability to recognize and explain structures and relationships within complex systems (Snapir et al., 2017). Systems thinking is an important concept in the curricular context of the current study. Students cannot develop a comprehensive understanding of the local marine ecosystem, without understanding how components within and across this system interact and how these interactions influence the system's functions (Yoon et al., 2016).

To track students' development of systems thinking, researchers have further broken down this concept into different categories, such as Components, Mechanisms, and Phenomena (Hmelo-Silver et al., 2017; Snapir et al., 2017). Components denote a system's elements, such as fish and kelp forest in a marine ecosystem. Mechanisms describe the causal links among components, such as how overfishing reduces fish populations. Phenomena capture the central patterns that students are trying to explain. An example phenomenon is the increase in biodiversity as fishing regulation becomes effective. The categories of Components, Mechanisms, and Phenomena can provide insights into the sophistication of systems thinking. Whereas emergent learners tend to emphasize separate components, systems experts more often focus on the coherent links between components, mechanisms, and phenomena when discussing systems (Hmelo-Silver & Pfeffer, 2004).

Students can enrich their systems thinking through self-explaining the relations within a system or collaboratively explaining ideas in group discussions (Jacobson & Wilensky, 2006). Explanations may invite students to articulate the components and mechanisms of the systems, incorporate data and scientific phenomena, and iteratively refine their assumptions towards a more coherent understanding (Hogan et al., 1999; Jordan et al., 2013; Nguyen & Santagata, 2020; Scardamalia & Bereiter, 1993). When verbally prompted for explanations, students who are describing how a system functions generally increase their reasoning about mechanisms and processes, instead of just focusing on components (Danish et al., 2017).

Agent Designs to Facilitate User-initiated Repairs

It follows that the conversational agents—as the listeners—can facilitate certain repair strategies that may promote explanations and other student talk moves conducive to learning. Listeners can request that the user rephrase their utterances (e.g., "I don't understand. Can you rephrase?"), provide options (e.g., "Here are what I can help you with: Pick A, B, or C"), apologize (e.g., "Sorry, I don't understand"), and ask for feedback (e.g., "My algorithm classifies this as food because it contains the word *bread.* Did you mean something else?"; Ashktorab et al., 2019; Engelhardt et al., 2017; M. K. Lee et al., 2010). The assumption behind these strategies is that humanlike, social behaviors such as apologies and feedback requests would make users more tolerant of breakdowns and more likely to engage in user-initiated repairs (Engelhardt et al., 2017).

Researchers have experimented with varying personas in appearances, personalities, and roles to adapt to users' underlying expectations and enact social norms in interactions with technologies like agents, voice interfaces, or robots (Chen et al., 2020; Kim, 2007; Kim & Baylor, 2016; M. K. Lee et al., 2010; Rosenberg-Kima et al., 2008). For example, in a study on human-robot interactions, M. K. Lee et al. (2010) found that users who perceived human-robot interactions as relational preferred apologetic repairs. Meanwhile, those who saw the interactions as transactional preferred compensation in the form of a coupon or a refund. This research illustrates that minimal design cues can prompt users to enact social interactions similar to how they may treat a human counterpart. Thus, agent designs to enact norms such as sociability or responsiveness might facilitate different user-initiated repair strategies.

Research Questions

In this study, I drew from research on the exchange between individual learners and pedagogical agents (Biswas et al., 2010, 2016; Kim & Baylor, 2016) to explore collaboration with a less knowledgeable peer versus an expert agent. The less-knowledge-peer version facilitates more idea articulation, explanation, and knowledge regulation (Biswas et al., 2010; Graesser, 2016). Meanwhile, learners might respond differently to an expert profile that is assumed to possess a higher level of expertise (Graesser, 2016). The following questions guided the research:

RQ1. What reactions do learner groups have when the agent misunderstands their intent or fails to address the intents sufficiently? How do these reactions correlate with evidence of groups' learning?

RQ2. To what extent do learner groups' reactions vary with different agent designs?

Methodology

This study presents a quantitative content analysis of group-agent interactions. I compared the differences in repair strategies between agent designs and examined the correlations between the occurrences of these strategies and group learning.

Study Context

The study took place within a high school science curriculum in the southwestern United States. The program engaged learners in field exploration and modeling to understand how restrictions on human impacts influenced local, coastal habitats. In groups of two or three, learners built a concept map that explained how regulations within a marine protected area might impact biodiversity. To deepen learners' systems thinking, the agents (nicknamed Kibot) were introduced into the activity to provide nudges around Components, Mechanisms, and Phenomena (Hmelo-Silver et al., 2017; Snapir et al., 2017).

Interactions with Kibot occurred early in the curriculum. Learners watched an introductory video about the marine protected area (MPA) and chatted on a Web interface about the components and mechanisms to explain biodiversity in the MPA. As the chat unfolded, Kibot's backend parsed the messages into relationship pairs to form a concept map. For instance, "fish eat plankton" would be parsed as {fish; plankton} and form a link between fish and plankton on the concept map (Figure 3.1). The agents kept track of the evolving concept maps to provide nudges for learner groups in the same chat window.

The current research employed a within-subject design, where learner groups talked to both agent versions. Groups randomly started with one prototype (less knowledgeable peer or expert) and switched halfway during a class period. The order of interaction was randomized to reduce practice effect when learners interacted with the second agent. Each conversation per agent lasted 12.5 minutes on average.

The agents facilitated systems thinking elements through an array of scaffolds that have shown promise in collaborative discussions (Dyke, Adamson, et al., 2013; Dyke, Howley, et al., 2013; Howley et al., 2013). As learners worked in groups to construct a concept map of the marine ecosystem, Kibot asked them to elaborate on ideas (e.g., “Can you explain more?”), provided direct prompts for additional mechanisms (e.g., “If plastic pollution increases, what would happen to the habitat?”), and nudged learners to build on peers’ ideas (e.g., “Do you agree with Erin? Why or why not?”). The prompts adapted to the current state of the groups’ concept maps by parsing learners’ chats in real time and providing hints that aligned with an underlying expert concept map. I outline details for the algorithm behind Kibot and the conceptual nudges in related work (Study 1). Of interest in this study is how these scaffolds varied with different agent designs. The underlying purposes of the agents’ prompts (e.g., to nudge for transactive exchange) were similar, but the speeches and appearances of the agents varied.

The less knowledgeable peer agent (panel C, Figure 3.1) resembled a peer with equivalent or less knowledge. This agent introduced itself as someone who had never been to the park and would benefit from any knowledge that the learners could provide. An example prompt for students’ elaboration is “I’m still learning. Can you explain to me what you meant?” Similar to previous work with one-on-one pedagogical agents (Y. Kim & Baylor, 2006), the agent embraced different expressions, such as holding a lightbulb when thinking or frowning when confused.

The expert agent (panel B, Figure 3.1) introduced itself as someone with scientific expertise. This version used more formal language in its questioning and prompting, and held the same expression throughout the conversation. To nudge students for elaboration or linking concepts in their concept maps, the expert agent framed the task as “checking to make sure you understand” or “show me what you know”. Because the expert agent was portrayed as someone with expertise, learners might be more likely to turn to this agent and abuse its hints. To balance interactions between the peer and expert agents, when learners asked the expert agent for hints, I programmed the reply to be “Ah, that’s a simple question. Why don’t we discuss in groups and answer it by yourself?" Meanwhile, the peer agent would respond with statements such as "I don't know. Can you find out and tell me?" In other words, the two agents provided equivalent prompts.

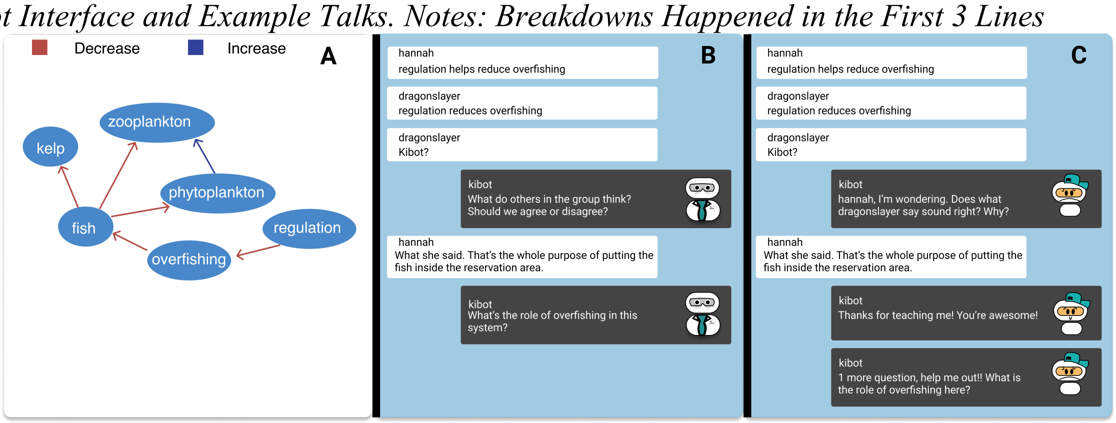
Figure 3.1 provides illustrations of the concept maps and the agent designs, and features an instance of communication breakdowns where the agents cannot parse the link between regulation and overfishing in the beginning (chat by student "dragonslayer"). Students attempted to alert the agent (calling "Kibot?") and explain the purpose of the protected area (reducing overfishing to protect the fish populations). As another example of communication breakdowns, students may mistakenly describe the direction of a connection (e.g., stating that fossil fuel burning decreases CO_2 instead of increasing CO_2). In this case, the agents might show the correct relationship ("increase"), and students might attempt to reconcile the discrepancy by repeating their utterances.

Participants and Data Sources

Participants in Study 2 were the same as Study 1. Participants included 18 groups of two or three (52 learners) from two ninth-grade classes taught by the same science teacher in a public high school. About half of the students were White and half were Latinx. The school enacted a one-to-one laptop policy. Learners had no prior experience with learning agents but had experience using technology to build concept maps. Because of the COVID-19 pandemic, learners were maintaining distance and primarily relying on the chat windows in group discussions.

Figure 3.1

Kibot Interface and Example Talks. Notes: Breakdowns Happened in the First 3 Lines



Data sources came from the group chat logs. The analyses were at the group level (involving 2-3 learners) due to the small number of utterances for individual learners. Analyses of chats at this level also allowed us to link group discourse to the concept maps they co-constructed.

Analytical Procedures

To answer RQ1 about communication repair strategies, I analyzed the chat data in two iterations. In the first iteration, I manually identified instances of communication breakdowns where the components or mechanisms in the group chats did not appear in the system-generated concept maps, even though those mechanisms would be conceptually correct. These instances would indicate circumstances where the agent's natural language understanding did not perform satisfactorily. During such circumstances, after five talk turns where the agent could not parse users' chats, the agents would prompt the users to provide further information, with statements such as "I didn't catch that. Can you explain to me what you mean?" (the expert version), or "I'm still learning. Can you explain to me what you mean?" (the less-knowledgeable-peer version). These agents' utterances provided another way to identify when conversational breakdowns happened (e.g., 5 chat turns before agents' prompts for further information).

Before coding, to avoid researchers' bias, I removed the agent conditions from the chat (e.g., marking a chat as Group 1/ Session 1 instead of Group 1/ Expert). I drew on the extant research in interactions with customer service agents to identify additional codes for user reactions to agents' failures: reframing to continue the conversation and quitting (Kvale et al., 2019). Within each category, there were finer-grained reactions, such as *repeating*, *rephrasing*, or *rephrasing and explaining*. I coded 5% of the data to validate the initial categories. In this stage, I also identified additional codes, such as *reframing by imitating* a prior agent's nudge or expressing *frustration*. The former code suggests that learners adopt the agent's scaffolds as models of productive talk (Dyke, Adamson, et al., 2013), while the latter indicates instances where breakdowns hinder student-agent interactions (Feine et al., 2019; Kumar et al., 2010).

In the second iteration, I coded 20% of the data to examine code stability. To establish inter-rater reliability, I trained a second coder on the coding scheme. The coders separately coded the same data subset and achieved an acceptable agreement with the author's codes (Cohen's κ = .77). I coded the rest of the data. Table 3.1 presents the codebook.

Table 3.1

Coding Scheme for Reaction to Agent's Communication Breakdowns

Codes	Definitions	Examples
Repeat (reframe)	Repeat a prior utterance	S1: Overfishing decreases fish. S2: Hello? Overfishing decreases fish.
Rephrase (reframe)	Reframe an utterance by changing the wording or adding words (without changing overall meaning)	S1: Overfishing decreases fish. S1: When humans hunt fish and fish too much, the fish population decreases.
Rephrase by imitating the agent's talk (reframe)	Reframe an utterance using similar sentence structures to the agents' nudges	Kibot: S2, what would happen if we overfish? S1: Overfishing decreases fish. S3: What would happen to the fish populations if we overfish?
Rephrase and explain (reframe)	Provide additional explanations to clarify intent	S1: Overfishing decreases fish. S1: When we overfish, fish populations may not have as much time to reproduce. So the fish population decreases overall.
Frustration (quit)	Express frustration at the breakdowns	S1: Urgh, why doesn't it connect overfishing to fish? OR S2: Kibot you're dumb.
Change topics (quit, switch subject)	Switch to a new idea unrelated to the current ideas	S1: Overfishing decreases fish. S2: Phytoplankton increases zooplankton.

Table 3.2

Coding Scheme for Systems Thinking in Groups' Chat

Codes	Definitions	Examples
C-C	Identify relationship between components	Fish lives in coral reefs.
C-M	Identify mechanisms by elaborating on connections between components	Kelp relies on photosynthesis.
C-P	Identify structures in relation to phenomena	Regulation reduces overfishing.
C-M-P	Identify components, links among them, and connections to phenomena	Regulation reduces overfishing. If there is less overfishing, there will be more biodiversity.

To link repair strategies to group learning, I analyzed the chat data for evidence of systems thinking. The chat data evidenced group knowledge construction because all learners were providing systems thinking statements within the chat sessions, with small variations between individuals. The data here included all students' interactions with the agents, instead of just utterances related to conversational breakdowns. Each group had two artifacts for analysis: one iteratively formed from chats with the less knowledgeable peer agent, and one with the expert version. Utterances were coded for Components, Mechanisms, and Phenomena, building on prior work with similar contexts around biodiversity (e.g., Hmelo-Silver et al., 2017).

The codes included *Component-Component* (C-C) to suggest simple, structural links between two systems elements. A statement coded as *Component-Mechanism* (C-M) or *Component-Phenomena* (C-P) would indicate links between components and the underlying mechanisms or phenomena of biodiversity, respectively (Table 3.2). An utterance that shows a coherent link among *Components, Mechanisms, and Phenomena* (C-M-P; as opposed to mentioning only components) would indicate attention to macro-level output and deeper systems thinking. To establish inter-rater reliability, the same coders from RQ1 coded all data separately. We reached a high agreement (Cohen's $\kappa = .90$) and resolved differences through discussion.

To answer RQ1 about how learners reacted to communication breakdowns, I calculated the frequencies of each repair strategy. I then linked the frequencies of repair strategies to counts of systems thinking statements by calculating Pearson's correlations between the repair strategy frequencies and the counts of C-C, C-M, C-P, and C-M-P connections.

To understand how the repair strategies differed between agent conditions (RQ2), I used the Wilcoxon signed-rank tests (for paired observations) to compare the repair strategies and counts of systems thinking codes between agent conditions. The Wilcoxon test was more appropriate when the data were non-normal. Briefly speaking, this test ranks the differences between pairs of observations for the same subject (i.e., learner group), and the test statistic (V) denotes the sum of ranks assigned to positive differences. To account for multiple comparisons, I used the Benjamini-Hochberg procedure to calculate the adjusted p values.

Results

RQ1: Learners' Repair Strategies are Correlated with Sophisticated Systems Thinking

Overall, 15% of the sessions included an attempt to fix the communication (*n* = 271 out of 1,764 total messages). The most frequent repair strategies were rephrasing (*n* = 124; 46%), followed by repeating (*n* = 44; 16%) and providing explanations (*n* = 43; 16%). In comparison, there were few instances of imitating the agent's talks (*n* = 21; 8%), switching topics (*n* = 26; 10%), and frustration (*n* = 13; 5%). For example, to reframe a statement such as "sea urchins bring down the population of kelp", students might say "sea urchins are bad for kelp". Meanwhile, to provide explanations for the same statement, students might add that "kelp is the food source of urchins". To examine the appropriateness of the group-level analyses, I further checked the proportions of repair moves that involved an individual versus multiple students. Thirty-eight percent of the repairs happened when multiple individuals joined in to correct the agents, indicating that exchange with the agents involved groups as well as individuals.

Consistent with prior frameworks (Chi, 2009; Davis, 2010), I found that group interactions that included a higher number of explanation repairs had more C-M-P statements, or statements with coherence among Components, Mechanisms, and Phenomena, $r(34) = .53$, $p = .01$. The following excerpt can help to contextualize the finding.

S1: ocean acidification decreases biodiversity.

S1: does ocean acidification decrease biodiversity?

S1: how do I

Kibot (peer): I'm learning so much from you! If there is something I did not catch, you can explain it to me?

S2: ocean aci [acidification] decreases biodiversity because it kills a lot of species

In this excerpt, the students S1 and S2 were trying to build a connection between ocean acidification and biodiversity. Because this link did not exist in the underlying expert map, Kibot did not parse the connection correctly. S1 attempted to reframe the connection (from a statement in the first line to a question in the second line), while S2 provided an explanation for this link in response to Kibot ("… because it kills a lot of species"). This excerpt illustrates how explaining behaviors serve as a form of constructive knowledge building, where learners link the components and bring in additional reasoning to make their ideas more coherent.

To examine whether explaining behaviors only occurred in conversation breakdowns or throughout the interactions with the agents, I conducted a qualitative examination of the chat logs for learners in groups with a high number of C-M-P statements. I observed that learners in these groups used explanations throughout and not just when the agent failed to respond. This suggests that groups that employed explanations to discuss systems' components, mechanisms, and phenomena throughout their interactions might have also used the same strategy to repair breakdowns with the agents.

RQ2: Groups Used more Reframing and Explaining in Response to the Peer Agent

To examine the differences between agent conditions, I ran a series of Wilcoxon signed-rank tests to compare how the counts of repair strategies differed when the same learner groups interacted with the peer versus the expert agents (Table 3.3). Groups used more reframing and explaining in interaction with the less-knowledgeable-peer agent, compared to the expert agent (reframing: M expert = 1.50, SD = 1.54; M peer = 4.09, SD = 3.60, p = .01; explaining: M expert = .80, SD = 1.67; M peer = 1.17; SD = 1.06, p = .02). This may indicate a higher tolerance for the less-knowledgeable-peer agent's mistakes and willingness to correct conversational breakdowns, compared to the expert agent.

Next, I compared the counts of systems thinking statements (C-C, C-M, C-P, or C-M-P) in interactions with the agents throughout the group conversations. There were more statements that connected Components-Mechanisms-Phenomena (*C-M-P*) in interactions with the less knowledgeable peer version than with the expert agent (*M* expert = 1.73; *SD* = 2.10; *M* peer = 3.36; *SD* = 1.75, $p < .01$).

These findings echo results from RQ1 about the positive correlation between explaining and evidence of systems thinking. Groups might have used more explanations in interactions with the less knowledgeable peer agent, and such explanations were correlated with a higher number of complex systems thinking statements.

Table 3.3

Repair Strategies Within Groups Within a Chat Session

Code	Expert		Peer				
	M	*SD*	*M*	*SD*	*V*	*p*	*Adjusted p*
Reframe							
Repeat	.95	1.32	1.09	1.38	218	.76	.86
Rephrase	1.50	1.54	4.09	3.60	130	.01*	.045*
Imitate	.65	.75	.35	.57	280	.16	.38
Explain	.80	1.67	1.17	1.06	142	.02*	.045*
Frustration	.30	.73	.30	.88	235	.86	.86
Switch topics	.70	1.17	.52	.67	221	.81	.86

Notes. * $p < .05$; ** $p < .01$; *** $p < .001$.

Discussion

Repair Strategies Can Enrich Learning

Failure to respond to user intent and maintain dialogues is a common challenge that may hinder user engagement and trust (Kvale et al., 2019; Luger & Sellen, 2016). The low abandonment rate in the current work is thus encouraging, with evidence that learner groups most often rephrased, repeated, and elaborated on their ideas.

Overall, I found that student groups most often employed reframing and repeating prior utterances to repair breakdowns with the agents. These behaviors are also observed in the usage of voice interfaces (Myers et al., 2018). However, repeating information may not adequately promote learning by itself (Teasley, 1995). Findings thus raise questions about how we can design agents and their responses to encourage repair strategies that are more conducive to learning. Prompting for users' explanations is one strategy that has proven effective (Kvale et al., 2019). Other studies have also explored the utility of rewards (M. K. Lee et al., 2010) or increasing feedback transparency by showing the underlying agent's algorithms (Ashktorab et al., 2019). These design ideas offer promising ground for future research in collaborative learning.

Additionally, I found a positive correlation between providing explanations during communication breakdowns and higher-level systems thinking throughout the group interactions. This finding illuminates how explaining to the agents to clarify misunderstanding can offer learners the opportunities to reason about systems concepts, particularly probing into the mechanisms they have not thought about (Danish et al., 2017). These processes can contribute to deeper learning by prompting individuals to identify gaps in their knowledge (Chi, 2009; Jordan et al., 2013). Such processes also help to make systems thinking more coherent, building towards a hierarchical understanding of visible and hidden processes instead of simple, linear links among components (Jordan et al., 2013; Snapir et al., 2017).

Explanations can be either self-initiated or part of a group's knowledge building efforts. In these instances, the interactions extend from one-to-one (i.e., agent and an individual learner) and human-human exchange (i.e., group discussion) to interactive group-computer knowledge building. These findings contribute to the notion of human-AI collaboration to consider the interplay between an AI agent (the agent) and multiple group members, where each party takes action in lieu of the responses from their partners. Positioning the agents as conversational partners can broaden the design scope of their appearances and dialogues. For example, in addition to giving nudges for the articulation of concepts and learner participation, the agents can also foster relationships with learner groups and contribute to constructing groups' knowledge artifacts.

Agent Designs Foster Different Repair Strategies

Conversational breakdowns are common due to the complex nature of human language. In response, researchers have developed language processing pipelines and responsive agent designs to increase users' tolerance of and willingness to continue engaging with the technology (Ashktorab et al., 2019; Chaves & Gerosa, 2019; Yu et al., 2016). The current work examined the affordances of design tweaks in agents' appearances and linguistic styles to facilitate different repair strategies.

It is possible that if learners were too focused on fixing their exchange with the agents, they would miss opportunities to develop their concept maps and deepen systems thinking. For this, I compared the counts of systems thinking statements in the overall groups' chat logs. Within groups, the conversations with the two agent designs did not differ significantly in the number of links between Components and Mechanisms or Components and Phenomena. However, conversations between agents differed in the counts of Components, Mechanisms, and Phenomena statements, suggesting more complex systems thinking with the less knowledgeable peer agent.

These findings can be linked to the learn-by-teaching paradigm, which creates opportunities for learners to seek information to teach an agent and identify gaps in their knowledge through testing the agent (Biswas et al., 2010). Prior work has found that a learn-by-teaching agent can promote systems learning more than an expert agent that provides feedback (Biswas et al., 2016). Findings from the current study echo this past literature and illustrate that unintentional moments of conversational breakdowns are not necessarily detrimental to learning.

These moments can support idea elaboration and knowledge construction, if the agent's language and designs can prime learners for certain cues, such as taking on the role of an explainer.

Limitations & Future Work

The limitations of the current study can provide directions for future work. First, due to the small sample size, analyses were for groups and not at the individual level. While these analyses align with the theoretical framework for knowledge building and co-evolution of group idea artifacts (Scardamalia & Bereiter, 1993), future work can examine the relations between repair strategies and learning artifacts for individuals and include covariates such as learner demographics and baseline understanding. Such analyses can reveal the extent to which repair strategies and learning outcomes differ for different student groups.

Analyses can also explore the pathways between agent condition, repair strategies, and evidence of enhanced learning. Follow-up work can explore the interplay between these variables in groups with mixed preferences. For instance, what would learning look like in groups where one individual focuses on providing explanations to the agents, while others just ignore it?

Second, the interactions between learner groups and the agents were constrained within the agent designs and the brief time of one class period. It is possible that if conversational breakdowns persist in extended interactions, learners may show different behaviors, such as more frustration rather than attempting to fix the interactions. Replication studies that vary the length of agent interactions, student populations, subject matters, or agent designs (beyond peer and expert personalities) can help to explain how repair strategies may vary with different study settings.

Conclusion

In this study, I explored the strategies that learner groups utilized to fix communication breakdowns with two agents in science learning domains. Findings highlight the positive link between providing explanations and deeper systems thinking. I found that groups varied in interactions with the agent designs. They provided more reframing and explaining to the less knowledgeable peer agent, compared to the expert agent. These findings suggest that communication strategies are related to the agents' characteristics, particularly if such designs can prime learners for social interactions. Results illustrate affordances of learning technology designs to engage learners to explain and co-construct knowledge, even during conversational breakdowns with the technology. They serve as a starting point for future work on the interactions of designs, user preferences, and learning.

References

Adamson, D., Dyke, G., Jang, H., & Rosé, C. P. (2014). Towards an agile approach to adapting dynamic collaboration support to student needs. *International Journal of Artificial Intelligence in Education*, *24*(1), 92–124. https://doi.org/10.1007/s40593-013-0012-6

Ashktorab, Z., Jain, M., Vera Liao, Q., & Weisz, J. D. (2019). Resilient chatbots: Repair strategy preferences for conversational breakdowns. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3290605.3300484

Biswas, G., Jeong, H., Kinnebrew, J. S., Sulcer, B., & Roscoe, R. O. D. (2010). Measuring self-regulated learning skills through social interactions in a teachable agent environment. *Research and Practice in Technology Enhanced Learning*, *5*(02), 123–152.

Biswas, G., Segedy, J. R., & Bunchongchit, K. (2016). From Design to Implementation to Practice a Learning by Teaching System: Betty's Brain. *International Journal of Artificial Intelligence in Education*, *26*(1), 350–364. https://doi.org/10.1007/s40593-015-0057-9

Brennan, S. E. (1998). The grounding problem in conversations with and through computers. *Social and Cognitive Approaches to Interpersonal Communication*, 201–225.

Chaves, A. P., & Gerosa, M. A. (2019). How should my chatbot interact? A survey on human-chatbot interaction design. *ArXiv Preprint ArXiv:1904:02743*. https://doi.org/10.1080/10447318.2020.1841438

Chi, M. T. H. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. *Advances in Instructional Psychology*, *5*, 161–238.

Chi, M. T. H. (2009). Active-Constructive-Interactive: A Conceptual Framework for Differentiating Learning Activities. *Topics in Cognitive Science*, *1*(1), 73–105. https://doi.org/10.1111/J.1756-8765.2008.01005.X

Clark, H. H., & Brennan, S. E. (1991). *Grounding in communication.*

Danish, J., Saleh, A., Andrade, A., & Bryan, B. (2017). Observing complex systems thinking in the zone of proximal development. *Instructional Science*, *45*(1), 5–24. https://doi.org/10.1007/S11251-016-9391-Z/TABLES/9

Davis, E. A. (2010). Scaffolding students' knowledge integration: prompts for reflection in KIE. *Http://Dx.Doi.Org/10.1080/095006900412293*, *22*(8), 819–837. https://doi.org/10.1080/095006900412293

Dyke, G., Adamson, D., Howley, I., & Rose, C. P. (2013). Enhancing scientific reasoning and discussion with conversational agents. *IEEE Transactions on Learning Technologies*, *6*(3), 240–247. https://doi.org/10.1109/TLT.2013.25

Dyke, G., Howley, I., Adamson, D., Kumar, R., & Rosé, C. P. (2013). Towards Academically Productive Talk Supported by Conversational Agents. In *Productive Multivocality in the Analysis of Group Interactions* (pp. 459–476). Springer US. https://doi.org/10.1007/978-1-4614-8960-3_25

Engelhardt, S., Hansson, E., & Leite, I. (2017). Better Faulty than Sorry: Investigating Social Recovery Strategies to Minimize the Impact of Failure in Human-Robot Interaction. *WCIHAI@ IVA*, 19–27.

Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2019). A Taxonomy of Social Cues for Conversational Agents. *International Journal of Human-Computer Studies*, *132*, 138–161. https://doi.org/10.1016/J.IJHCS.2019.07.009

Følstad, A., Nordheim, C. B., & Bjørkli, C. A. (2018). What Makes Users Trust a Chatbot for Customer Service? An Exploratory Interview Study. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11193 LNCS*, 194–208. https://doi.org/10.1007/978-3-030-01437-7_16

Gnewuch, U., Morana, S., & Mädche, A. (2017). Towards Designing Cooperative and Social Conversational Agents for Customer Service. *ICIS 2017 Proceedings*. https://aisel.aisnet.org/icis2017/HCI/Presentations/1

Graesser, A. C. (2016). Conversations with AutoTutor Help Students Learn. *International Journal of Artificial Intelligence in Education*, *26*(1), 124–132. https://doi.org/10.1007/s40593-015-0086-4

Hmelo-Silver, C. E., Jordan, R., Eberbach, C., & Sinha, S. (2017). Systems learning with a conceptual representation: a quasi-experimental study. *Instructional Science*, *45*(1), 53–72. https://doi.org/10.1007/S11251-016-9392-Y/FIGURES/5

Hmelo-Silver, C. E., & Pfeffer, M. G. (2004). Comparing expert and novice understanding of a complex system from the perspective of structures, behaviors, and functions. *Cognitive Science*, *28*(1), 127–138. https://doi.org/10.1207/S15516709COG2801_7

Hogan, K., Nastasi, B. K., & Pressley, M. (1999). Discourse Patterns and Collaborative Scientific Reasoning in Peer and Teacher-Guided Discussions. *Cognition and Instruction*, *17*(4), 379–432. https://doi.org/10.1207/S1532690XCI1704_2

Howley, I., Kumar, R., Mayfield, E., Dyke, G., & Rosé, C. P. (2013). Gaining Insights from Sociolinguistic Style Analysis for Redesign of Conversational Agent Based Support for Collaborative Learning. In *Productive Multivocality in the Analysis of Group Interactions* (pp. 477–494). Springer US. https://doi.org/10.1007/978-1-4614-8960-3_26

Jacobson, M. J., & Wilensky, U. (2006). Complex Systems in Education: Scientific and Educational Importance and Implications for the Learning Sciences. *Http://Dx.Doi.Org/10.1207/S15327809jls1501_4*, *15*(1), 11–34. https://doi.org/10.1207/S15327809JLS1501_4

Jordan, R. C., Hmelo-Silver, C., Liu, L., & Gray, S. A. (2013). Fostering reasoning about complex systems: using the aquarium to teach systems thinking. *Applied Environmental Education & Communication*, *12*(1), 55–64.

Kim, S., Eun, J., Oh, C., Suh, B., & Lee, J. (2020). Bot in the Bunch: Facilitating Group Chat Discussion by Improving Efficiency and Participation with a Chatbot. *Conference on Human Factors in Computing Systems - Proceedings*, 1–13. https://doi.org/10.1145/3313831.3376785

Kim, Y., & Baylor, A. L. (2006). A social-cognitive framework for pedagogical agents as learning companions. *Educational Technology Research and Development*, *54*(6), 569–596. https://doi.org/10.1007/s11423-006-0637-3

Kim, Y., & Baylor, A. L. (2016). Research-Based Design of Pedagogical Agent Roles: A Review, Progress, and Recommendations. *International Journal of Artificial Intelligence in Education*, *26*(1), 160–169. https://doi.org/10.1007/s40593-015-0055-y

Kumar, R, Beuth, J., & Rose, C. (2011). Conversational Strategies that Support Idea Generation Productivity in Groups. In H. Spada, G. Stahl, N. Miyake, & N. Law (Eds.), *Connecting Computer-Supported Collaborative Learning to Policy and Practice: CSCL2011 Conference Proceedings. Volume I - Long Papers* (pp. 398–405). International Society of the Learning Sciences. https://repository.isls.org/handle/1/2475

Kumar, R., Ai, H., Beuth, J. L., & Rosé, C. P. (2010). Socially capable conversational tutors can be effective in collaborative learning situations. *Lecture Notes in Computer Science*, *6094 LNCS*(PART 1), 156–164. https://doi.org/10.1007/978-3-642-13388-6_20

Kvale, K., Sell, O. A., Hodnebrog, S., & Følstad, A. (2019). Improving Conversations: Lessons Learnt from Manual Analysis of Chatbot Dialogues. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11970 LNCS*, 187–200. https://doi.org/10.1007/978-3-030-39540-7_13

Lee, M. K., Kiesler, S., Forlizzi, J., Srinivasa, S., & Rybski, P. (2010). *Gracefully mitigating breakdowns in robotic services*. 203–210. https://doi.org/10.1109/HRI.2010.5453195

Lee, S., Lee, S., & Kim, S. (2019, May 2). "What does your agent look like?" A drawing study to understand users' perceived persona of conversational agent. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3290607.3312796

Luger, E., & Sellen, A. (2016). "Like having a really bad pa": The gulf between user expectation and experience of conversational agents. *Conference on Human Factors in Computing Systems - Proceedings*, 5286–5297. https://doi.org/10.1145/2858036.2858288

Myers, C., Furqan, A., Nebolsky, J., Caro, K., & Zhu, J. (2018). Patterns for how users overcome obstacles in Voice User Interfaces. *Conference on Human Factors in Computing Systems - Proceedings*, *2018-April*. https://doi.org/10.1145/3173574.3173580

Nguyen, H., & Santagata, R. (2020). Impact of computer modeling on learning and teaching systems thinking. *Journal of Research in Science Teaching*, tea.21674. https://doi.org/10.1002/tea.21674

Roscoe, R. D., & Chi, M. T. H. (2007). Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors' explanations and questions. *Review of Educational Research*, *77*(4), 534–574. https://doi.org/10.3102/0034654307309920

Scardamalia, M., & Bereiter, C. (1993). Technologies for knowledge-building discourse. *Communications of the ACM*, *36*(5), 37–41. https://doi.org/10.1145/155049.155056

Snapir, Z., Eberbach, C., Ben-Zvi-Assaraf, O., Hmelo-Silver, C., & Tripto, J. (2017). Characterising the development of the understanding of human body systems in high-school biology students – a longitudinal study. *Http://Dx.Doi.Org/10.1080/09500693.2017.1364445*, *39*(15), 2092–2127. https://doi.org/10.1080/09500693.2017.1364445

Teasley, S. D. (1995). The Role of Talk in Children's Peer Collaborations. *Developmental Psychology*, *31*(2), 207–220. https://doi.org/10.1037/0012-1649.31.2.207

Yoon, S. A., Anderson, E., Klopfer, E., Koehler-Yom, J., Sheldon, J., Schoenfeld, I., Wendel, D., Scheintaub, H., Oztok, M., & Evans, C. (2016). Designing computer-supported complex systems curricula for the Next Generation Science Standards in high school science classrooms. *Systems*, *4*(4), 38.

Yu, Z., Nicolich-Henkin, L., Black, A. W., & Rudnicky, A. (2016). A wizard-of-oz study on a non-task-oriented dialog systems that reacts to user engagement. *Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 55–63.

CHAPTER 4 CASE STUDY OF GROUP COMPOSITIONS

An earlier version of this chapter has appeared in:
Nguyen H. (2022) Exploring Group Discussion with Conversational Agents Using Epistemic Network Analysis. In: Wasson B., Zörgő S. (eds) Advances in Quantitative Ethnography. ICQE 2021. Communications in Computer and Information Science, vol 1522. Springer, Cham. https://doi.org/10.1007/978-3-030-93859-8_25
The final authenticated version is available at: https://doi.org/10.1007/978-3-030-93859-8_25

Abstract

Conversational agents—dialogue systems that provide learning support to students in real time—have shown promise in facilitating science discussions. To enact conversation norms, the appearances, personalities, or tones of the agents often resemble personas that students are familiar with, such as peers or mentors. This study uses epistemic network analysis (ENA) to explore how students interacted with two agent designs, a less knowledgeable peer and an expert, in collaborative settings. ENA visualizes co-occurring discussion moves in student interactions as networks and compares the differences in discussion networks between the two agents. Data came from the chat logs of three student groups with emergent, mixed, and expanding prior domain knowledge. The groups interacted with both prototypes. The chat logs received qualitative codes for discussion types, including claim-making, reasoning, building on prior ideas, and responsiveness to the agents. ENA visualized the differences in discussions between groups and between the agent conditions. Overall, the expanding groups engaged in more claim-making in tandem with building on prior ideas when interacting with the less knowledgeable peer agent, compared to the expert agent. Meanwhile, the emergent group showed more syntheses of previous ideas when responding to the expert agent. Findings illuminate the design space for adapting agent designs to group settings to facilitate productive exchange, for example, fading scaffolds in groups with higher levels of prior knowledge.

Introduction

A fundamental challenge in science education in the United States is the promotion of sustained interest and participation in scientific practices (Zacharia & Barton, 2004). Engaging in collaborative discussion offers ways for students to build on the knowledge of others, gain agency, and develop interest in science (Hoadley & Kilner, 2005; Scardamalia & Bereiter, 1993; Stahl et al., 2014). Conversational agents such as the text-based chatbots in this study have shown particular promise in facilitating discussion (Dyke et al., 2013; Kumar et al., 2011; Tegos & Demtriadis, 2017). These agents use natural language understanding of text and speech to provide suggestions that enrich idea generation, argumentation, and conceptual knowledge.

A key goal in designing conversational agents is to make them more conducive to discussion and subsequent learning (Kim & Baylor, 2016; Seering et al., 2019; Walker et al., 2014). The agent's appearances, tones, and gestures may enact human characteristics such as friendliness, competence, or sociability, to prime users to demonstrate conversational norms as if they were interacting with human partners (Moon & Nass, 1996; Nass et al., 1994). Agents can adopt the appearance and lingo of a mentor, teacher, or peer in educational contexts to simulate classroom norms (Kim & Baylor, 2006, 2016; Graesser, 2016; Liew et al., 2013). Learners with emergent understanding may seek help from a mentor agent, whereas learners with expanding knowledge may deepen their knowledge through giving help to a peer agent (Graesser, 2016).

However, considerations of agent designs have mostly been applied to individual interactions between a student and an agent and not to collaborative contexts. In the latter case, students' interactions with the agents may vary with group dynamics (Oliveira & Sadler, 2008). This is because the hints and questions from the agents may get ignored in groups where elaboration is not the norm (Kumar et al., 2011). A factor that may influence such group dynamics is knowledge composition (Saleh et al., 2005; Webb & Farivar, 1994). For example, students with emerging or average knowledge may not participate as much when the conversation is dominated by peers with expanding knowledge.

The current case study explores how students' interactions with conversational agents vary in three student groups of emergent, mixed, and expanding prior knowledge within a high school science curriculum. The group classification is based on students' pretest of science domain: emergent (all three students scored below average), mixed (at least one above average), and expanding (all above average). This exploratory study serves to illuminate the dynamics in discussion with an agent, and not to recommend a way to group students based on baseline understanding. I draw from prior agent designs in individual tutoring settings to test two prototypes: a less knowledgeable peer and an expert agent (Graesser, 2016; Kim & Baylor, 2016). In a within-subject design, student groups (each consisting of two to three 9th graders) interacted with both prototypes in randomized orders, as they built a concept map of the marine ecosystems. The following questions guided the research:

RQ1. How do students' collaborative discussion patterns vary with different group compositions, based on prior domain knowledge?

RQ2. Within student groups, how do discussion patterns differ when interacting with the less knowledgeable peer versus the expert agents?

Overall, students in the mixed and expanding groups showed more connection between making claims, providing reasoning, and building on prior ideas, compared to the group with emergent knowledge. The groups also showed distinct patterns when interacting with the two agents. Students in groups with expanding levels of prior knowledge appeared to engage in more reasoning and claim-making when interacting with the less knowledgeable peer version, compared to the expert version. Meanwhile, those in the group with emergent knowledge levels showed more reasoning and elaboration on prior ideas when responding to the expert agent.

These findings illuminate how responses to agent designs vary with group composition. This understanding has important design and learning implications. From a design perspective, findings contribute to research on developing agent personas that can be perceived as natural social partners and encourage productive conversational norms. From a learning perspective, designers of education technology and facilitators may examine the diverse patterns that student groups display. Such considerations help to identify patterns that are conducive to knowledge construction and assist the development of adaptive learning scaffolds for different students.

Background

Conversational Agents to Promote Collaboration

Engaging in knowledge building efforts, where students collaboratively share ideas and build on one another's knowledge, is an important process to enrich science understanding (Hoadley & Kilner, 2005; Scardamalia & Bereiter, 1993). Through social exchange, individuals outline their personal and sociocultural experiences, discuss, negotiate, and develop a shared understanding of learning phenomena (Muhonen et al., 2017; Stahl et al., 2014).

Students engage in a range of knowledge building discussion moves, from stating claims to elaborating on and expanding shared knowledge (Gillies et al., 2014). This wide array of discussion moves presents opportunities for teachers to support different participation structures. For example, teachers can decide which students may initiate questions or propose actions, ask questions to redirect student attention, or propose tasks for students to build on one another's ideas and artifacts (Chi et al., 2001; Hewitt & Scardamalia, 1998). These participation structures shift knowledge building from teachers as the sole transmitter of knowledge to students and the tools that they engage with (Scardamalia & Bereiter, 1993; Hewitt & Scardamalia, 1998).

Conversational agents can foster knowledge building communities, similar to teachers' facilitation. These agents use natural language understanding to process students' talks and provide in-time nudges for conceptual understanding and participation. More than tools to enrich student discourse, agents can serve as active participants in group conversations. Agents can propose prompts for the groups to explain, contrast, and support their ideas, thus fostering different discussion moves (Dyke et al., 2013).

Less Knowledgeable Peer and Expert Agent Designs

Researchers have underscored the importance of designing agents to facilitate natural human-computer interactions. Designs that replicate human behaviors or social norms can influence users' perceptions. In turn, users display social behaviors when interacting with the computer agents (Nass et al., 1994). Users may engage in social norms such as gender stereotyping or reciprocating help with the machines, even with small interface cues or when they acknowledge that machines may not have underlying feelings (Moon & Nass, 1996).

Two distinct agent profiles have emerged in learning contexts: peer and expert agent. The peer agent profile is based on the similarity-attraction effect (Byrne & Nelson, 1965), which suggests that learners would be attracted to agents that parallel them in appearance, knowledge, or interest (Liew et al., 2013). In a study where researchers assigned students to work with computers whose characteristics either aligned with or mismatched students' personalities, students agreed more with the personality that was similar to their own and found it more attractive and intelligent (Nass et al., 1995). A related design is the less knowledgeable agent (Biswas et al., 2016), where users are responsible for explaining target concepts to an agent with presumably limited knowledge. From these interactions, students may deepen reasoning and understanding of the target concepts.

Meanwhile, students may attribute more trust to agents whom they perceive to possess higher knowledge levels, in ways that are similar to how they would interact with a teacher or domain expert (Chaiken & Maheswaran, 1994). This conjecture has led to the designs of expert-like agents—characters that are more knowledgeable and employ talk moves that simulate the human instructors (Heidig & Clarebout, 2011).

User perceptions and responses to agent designs can be leveraged in designing agents to produce optimal learning and interests (Chen et al., 2020; Graesser, 2016). For example, agent designs can adapt to learners' domain knowledge (Graesser, 2016). Learners with emergent understanding and skills may benefit from interacting with the expert-like agent. Meanwhile, learners with expanding knowledge may attempt to teach the peer agent. Such interactions provide opportunities for learners to acquire alternative viewpoints and deepen knowledge.

Agent Designs Influence Human-Agent Interactions

The content of user replies also varies with how users perceive the agent (Kim & Baylor, 2006, 2016; Rosenberg-Kima et al., 2008; Kim et al., 2019; Chen et al., 2020). For example, users rated a casual tone agent as friendlier and provided more elaborate survey responses to the agent than one that relied on a formal questionnaire style (Kim et al., 2019). In an exchange with learning robots, learners who interacted with a robot that behaved as a peer showed more affective displays, compared to a robot that resembled a tutor (Chen et al., 2020).

Research from human-human tutoring interaction lends further insights into the potential differences in the content of students' responses to the different agent profiles. In inquiry-driven classrooms, teachers tend to use questions to elicit students' thinking, and students reply to teachers' elicitations, elaborate on their thinking, or offer argumentation (Chin, 2006). In peer tutoring, rather than only focusing on providing explanations, students may also express emotions such as triumph, anger, and confusion around the learning problems (Agne & Muller, 2019). They also pose provoking questions to their partners (King et al., 1998). Thus, we may expect different discussion patterns with the agent profiles. For example, students may engage in more explanation and argumentation with the expert agent, while being open to brainstorming ideas when it comes to discussing with the peer agent.

In addition, embedding the agents in group discussions introduces an additional, interesting dynamic. Students with different levels of domain knowledge may interact with each other and with the agents in varied ways. In mixed-knowledge groups, students with higher levels of domain knowledge tend to become the "mentors" who provide help and explanations to other students. In turn, the peers with emergent understanding can benefit from spontaneous help-seeking (Saleh et al., 2005; Webb & Farivar, 1994). We can thus explore whether students with expanding knowledge will dominate the conversations in interactions with a less knowledgeable peer agent, or if other students (e.g., those with emergent knowledge) will also see the agent with limited domain knowledge and engage in elaborative talk with the agent.

How can we analyze the content of group discussions to reveal differences in interactions between agent profiles? Education researchers have focused on the types of argumentative discourse students make, the vocabulary and conversational topics they engage in, and their contributions to group discussions (Howley et al., 2013; Rosé et al., 2008). Researchers of group discourse have also examined how discussion moves can co-occur within the same conversational windows (Gašević et al., 2019; Shaffer et al., 2016). Analyses of co-occurring moves can reveal key insights to formulate the epistemic structure of the discussions. For example, co-occurrences between claim and evidence making within the same message may suggest a more coherent argument than a message with only claims. To understand broader patterns of students' interactions, the current research thus analyzes discussion patterns as co-occurring moves, instead of separate counts of talks.

Hypotheses

In this work, I compared discussion patterns with two conversational agents (a less knowledgeable peer and an expert) within and between groups with different domain knowledge. My hypotheses were built on prior work on interactions between individual students and agents (Biswas et al., 2016; Graesser, 2016). I hypothesized that students with emergent knowledge would likely engage in more help-seeking from the expert agent, whereas those in the expanding knowledge groups would provide more help-giving to the less knowledgeable peer agent.

H1. Between groups, groups with expanding content knowledge will engage in more explaining to the less knowledgeable peer agent than groups with emergent content knowledge.

H2. Within groups, the expanding group will be more responsive to the less knowledgeable peer agent than the expert agent, similar to behaviors observed in peer tutoring where students with expanding knowledge dominate the discussions. In contrast, the emergent group will be more responsive to the expert agent than the less knowledgeable peer agent.

Methodology

This study presents a qualitative discourse analysis of students' discussion moves with one another and the conversational agents. The analyses focused on three purposefully sampled student groups of emergent, mixed, and expanding prior knowledge in science. I applied ENA (Shaffer et al., 2016) to study the differences between group compositions and agent conditions. ENA visualizes student interactions as networks, where each discussion move is a node and co-occurring discussion moves are connected. The choice of ENA builds on an understanding of collaborative learning not as isolated elements, but as relationships among those elements to form systemic understanding (Shaffer et al., 2016).

Study Settings

This research is situated in a high school marine biology program that is part of a multi-year partnership between a local state park, education and biology researchers, and local school districts in the southwestern United States. The state park has been working for several years to study how a coastal marine protected area responds to reduced human impacts. Early in the curriculum, students were asked to brainstorm the elements and processes that might affect the marine habitats. In groups of two to three, students engaged in a virtual chat with each other and with the agents. The goal of the chat was to build a concept map of the marine ecosystem and to reason through the connections in the concept map.

Agent Design

The agent designs are detailed in Study 1. The agents applied natural language processing to parse students' messages into subject-object pairs that represented relationships within the marine ecosystem, and compared students' answers to an expert concept map. The main goals of the agents' prompts were to (1) help students think about missing relationships, and (2) encourage the less active participants to partake in the discussion.

In this study, similar to procedures in Studies 1 and 2, student groups interacted with both the less knowledgeable peer and the expert agents. The interaction order with the agents was randomized. Half of the student groups started with the peer agent, while the other half started with the expert agent, and switched halfway through the activity.

The Less Knowledgeable Peer Agent. The agent was designed to resemble a peer with lower knowledge levels. This agent was not presented as learning from the chat. The agent used colloquial expressions to ask the students to "explain" concepts and discuss with peers to teach the agent. Similar to work in one-on-one tutoring (Kim & Baylor, 2016), the peer agent expressed emotions through changing its facial expressions, e.g., holding a lightbulb when suggesting ideas about connections in the students' concept maps.

The Expert Agent. The expert agent was designed to resemble a scientist with mastery of the content knowledge. This agent used a formal tone. Contrary to the less knowledgeable peer agent that changed its expressions, the expert agent kept the same expression throughout the chats. The agent asked students to explain concepts to prove their understanding.

Participants

Participants were from the same sample as Studies 1 and 2. The sample involved ninth-grade students in two classes taught by the same science teachers in a public high school in the southwestern U.S. The school served a diverse student population that was majorly White and Hispanic/Latino in the 2019-20 school year. The students were participating in the marine science program during their normal environmental science class. The school had a one-to-one laptop policy, and students were familiar with using chat windows to converse with one another. Students had experiences with collaborative group work in their science class but reported limited experiences with learning chatbots before the lessons.

Prior to seeing the agent interface or interacting with the agents in groups, all participants individually answered a pretest. The pretest aimed to capture students' science domain knowledge of the MPA. In the pretest, students answered three open-ended questions about the marine ecosystem and the role of fishing regulation. The questions were as follows:

(1) How do marine protected areas affect fish populations and other things in the ocean? List all the connections you can think of.

(2) What are some ways we can improve ocean biodiversity?

(3) After the Marine Protected Area is introduced, we see an increase in the number of fish, particularly sheepshead and kelp bass. What do you think contributes to this increase? Explain your answer.

The pretest's scoring scheme was consistent with the curriculum's focus on systems thinking, or understanding of relationships and functions of complex systems (detailed in Study 2). Student responses were scored as the sum of correct statements about the systems' components, mechanisms (relationships among components), and central phenomena of biodiversity. The scores were then ranked, and groups were categorized as emergent (all three students scored below average), mixed (at least one above average), and expanding (all above average).

Using purposeful random sampling (Patton, 2014), I randomly selected one group from each prior knowledge category for this study's analysis. The *Emergent Group* consisted of a female and two male students. The *Mixed* group had three female students, and the *Expanding* group had one female and two male students. The goal of the analysis is to understand potential variances in groups' interactions and to inform future design iterations of the agents.

Data Sources

The main data source came from the three groups' chat logs with the agents. Each chat per prototype (peer or expert) lasted 12.5 minutes on average (Emergent: 12 minutes with peer; 11 minutes with expert; Mixed: 14 minutes with expert; 12 minutes with peer; Expanding: 14 minutes with peer; 12 minutes with expert).

The unit of code was a single chat message (total $N = 407$ utterances; each group had 136 messages on average, SD = 23.5 messages). I used a priori codes for the types of contribution to the group's knowledge building (Kumpulainen & Wary, 2002; Muhonen et al., 2017). Additionally, I conducted a close reading of the group chats to devise emergent codes. An additional code (*Respond to Kibot*) emerged in this stage to note whether a student's utterance was self-initiated or in response to the agent's nudges. Although there were other codes included in prior frameworks, such as emotion (Muhonen et al., 2017) or questioning (Study 1), for this study's sample, these codes had fewer than three occurrences per agent and were thus excluded from the final analyses. Table 4.1 provides examples of the codes.

To establish reliability, a research assistant and I separately coded 20% of the chat and showed substantial agreement with the original codes, average Cohen's $\kappa = .96$. I coded the rest of the data. Each chat message received a dichotomous code for whether the code was present (coded as 1) or not (coded as 0). This means that a message can receive more than one code if multiple discussion moves existed, e.g., if students built on friends' ideas while making claims.

Procedures

In this study, I applied Epistemic Network Analysis (ENA) in the rENA package (Marquart et al., 2018) to examine the co-occurrences of codes within a moving chat window of three. Discussion types were considered to be associated if they appeared in the same sliding text window (e.g., a window of size three measures co-occurrences within three consecutive texts).

In ENA, co-occurrences of discourse types formed a binary matrix (1: occur; 0: not occur). Then, the package normalized the co-occurrence matrix and applied singular value decomposition (SVD) to reduce the dimension of vectors to two dimensions that explained the most variance in the data. ENA allowed for visualizations of the discussion network for each group. Each code (e.g., claim, reasoning, response to Kibot) became a node in the diagram, with lines indicating that the two nodes co-occurred within the chat windows. ENA also allowed for visualizing subtraction networks, that is, the differences between the epistemic networks for each condition per group. The visualization subtracted the connection weight of each network node and indicated a larger connection between discourse types by showing thicker, darker lines.

Table 4.1.

Coding Scheme for Student Discussion

Code	Definition	Example
Claim	Statement links systems elements	Kelp provides habitat.
Reasoning	Statement draws from pre-existing knowledge, scientific facts and data, or evidence in the lesson	plastic pollution reduces fish population **because it contaminates their habitat**
Build on self (Externalize)	Statement draws on previous ideas a student has stated	A: Sea urchins increase kelp. A: Sea urchins decrease kelp because they eat a lot of kelp.
Build on friend (Transactive)	Statement draws on previous ideas that peers have stated	A: Plastic pollution should be banned. B: Yes, because it harms the animals.
Respond to Kibot	Statement in response to the agent's nudges	Kibot: What do you think, A? Do you agree or disagree? A: I agree because a ban should let fewer people into that area and there will be less overfishing.

To answer RQ1 about the potential differences in students' interactions with the conversational agents between the emergent, mixed, and expanding groups, I first ran ENA with students in groups as the unit of analyses for all utterances (i.e., both agent prototypes). Analyses plotted the subtraction networks for group pairs (e.g., emergent-mixed, mixed-expanding, expanding-emergent) to highlight differences between groups' discussion networks.

To answer RQ2 about how students' discussions might differ when interacting with the agent prototypes, I conducted ENA within each group. Individual students' chat occurrences within groups within conditions constituted the unit of analyses.

Results

Between-group Differences in Discussion Network

The first research question examines the extent to which group discussion networks varied prior domain knowledge of systems thinking, as indicated by students' pretest scores. I created subtraction networks (Figure 4.1) and compared the connections between discussion types between groups. Each dot in the figure represents an individual student, the squares denote the centroid (or the mean of the group in high-dimensional space), and the colors indicate the groups they belong to (i.e., black for Emergent, purple for Mixed, and blue for Expanding).

For example, consider the top right panel comparing the Mixed and Expanding groups in Figure 4.1. The lines are the "leftovers" after subtracting the networks of the two groups. The purple line between "reasoning" and "buildonFriend" (transactive exchange) suggests that when comparing the Mixed and Expanding groups, the Mixed group had more connections between the two discussion moves. Meanwhile, the blue lines suggest that the Expanding group had relatively more co-occurrences between "claim" and "buildonFriend", "claim" and "respond to Kibot", and "claim" and "buildonSelf" (externalize).

Overall, students in the expanding groups appeared to show more complex discussion moves. For example, when comparing groups Mixed and Emergent (left panel, Figure 4.1), the purple line in Figure 4.1 suggests a link between *Reasoning* and *Transactive exchange* (build on friends' ideas). This indicates that individuals in group Mixed (the group with relatively higher domain knowledge) showed more connections in these discussion moves. Similarly, the blue lines in the central and right panel (Figure 4.1) show that group Expanding showed more connections between *Claims* and *Externalize* (build on self's ideas) compared to the other groups. To illustrate, consider the following excerpt from the Emergent group.

Kibot: What would happen if kelp increases?

S1: Fish increase/zooplankton increase/phytoplankton increase.

S2. Pollution increases/ fish increases.

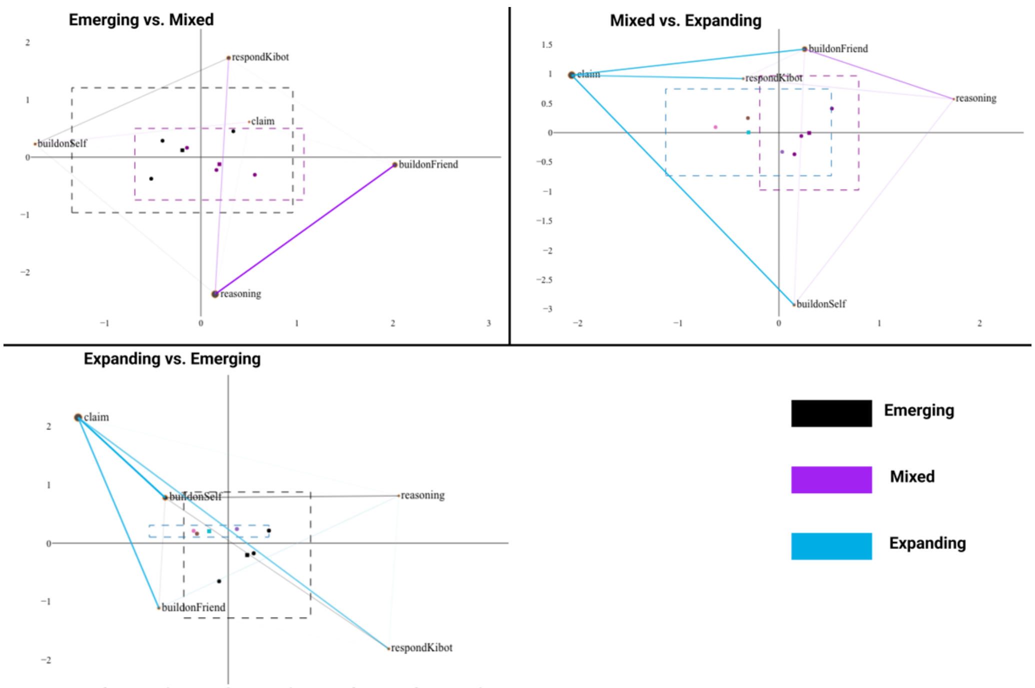
S1: Global warming increases/ water temperature increases.

Kibot: Why do you think so?

S3: When the temperature rises it affects the fish.

Figure 4.1

Between-group Comparisons (ENA Subtraction Networks) for group Emergent (black), Mixed (purple), and Expanding (blue).



Notes. Dots present individual students, and lines present co-occurrences between discussion moves. The darker colors of the lines show differences in co-occurrences between groups.

In this exchange, students were mostly making claims without providing reasoning for their answers. They also did not explicitly mention previous ideas to build on. In contrast, the following excerpt from the Expanding group illustrates how S5 built on his friend's idea ("regulation increases fish") to give explanations and inquired about a related relationship (breeding-fish).

Kibot: What would happen if regulation increases?

S4: Regulation increases fish.

S5: Regulation increases fish because it protects the fish from overfishing.

S5: Does breeding increase fish?

In sum, group comparisons reveal that students in expanding groups appeared to show more connection between making statements, reasoning, and engaging in transactive exchange, compared to the emergent group.

Within-group Differences in Interactions with the Agents

The observed differences in interactions between groups suggest that there may exist variations with group composition. The second research question explores these differences using another grain size: within-group shifts in interactions between the less knowledgeable peer and the expert agent. To answer this question, I used ENA to compare the discussion networks between the agent conditions for each group (Figures 4.2-4.4; red lines indicate more connections for the expert agent; blue lines indicate the less knowledgeable peer agent).

Overall, the most noticeable difference between the less-knowledgeable-peer and expert discussion networks for the Emergent group is the connection between *Externalize* and *Respond to Kibot* in the expert condition, compared to the less-knowledgeable-peer condition (Figure 4.2). This observation is based on the red line in the right panel of Figure 4.2, showing that the expert group (red lines) had more co-occurrences of “buildonSelf” and “respondKibot” when subtracting the two networks in the left panel. While this group mostly stated simple claims, within a 3-utterance chat window, students more frequently built on prior claims when responding to nudges from the expert agent than those from the less-knowledgeable-peer agent.

Meanwhile, the Expanding group (Figure 4.3) demonstrated more discussion connections in the less knowledgeable peer condition, compared to the expert agent. Individuals made more links between *Claim – Respond to Kibot; Claim – Externalize;* and *Respond to Kibot – Transactive*, as highlighted by the blue lines for the peer condition. Take the following excerpt from the Expanding group as an example of co-occurrences for *Respond to Kibot – Transactive*:

S6: Regulation decreases plastic pollution.

Kibot (peer): Ah, regulation can reduce plastic pollution! S5, what other ways do humans influence the ecosystem?

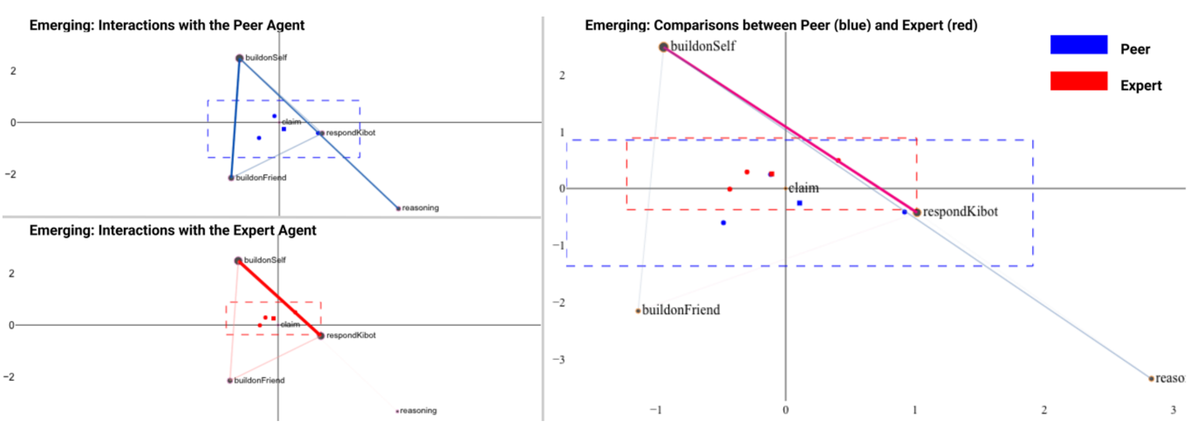
S4: CO_2 emissions increase global warming.

S5: Other ways we influence the ocean ecosystem is burning of fossil fuels, so there should be regulation for that too.

In this excerpt, the less knowledgeable peer agent was encouraging S5 to build on S6's idea around regulation. In response, S4 and S5 brought up related concepts ("CO_2 emissions", "global warming", "fossil fuel burning"). S5 specifically linked the ideas under discussion ("regulation") in her response.

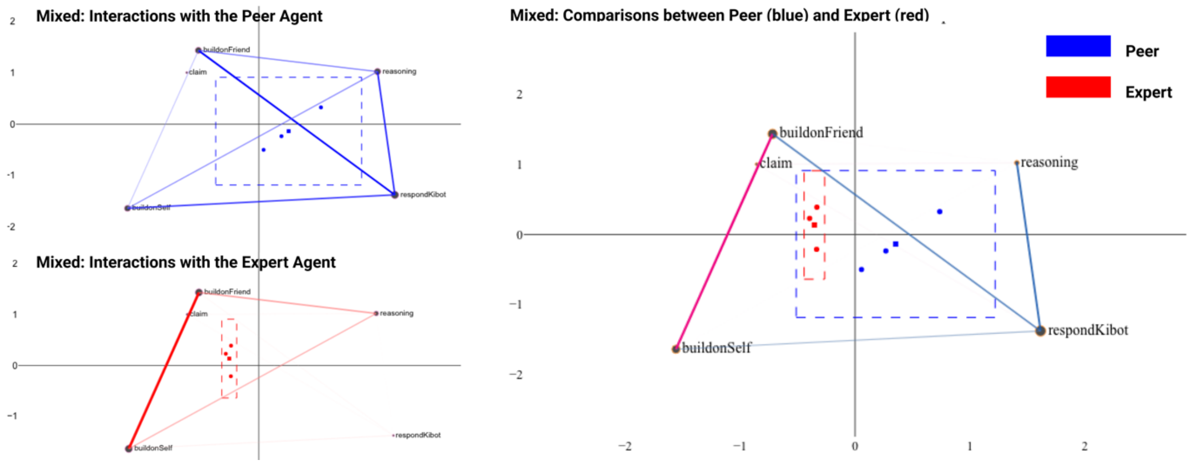
Figure 4.2

Comparisons of Emergent group between Expert (Red) and Peer (Blue) Agents



Notes. Dots present individual students, and lines present co-occurrences between discussion moves. Right panel: The darker colors of the lines show differences in co-occurrences between agent conditions.

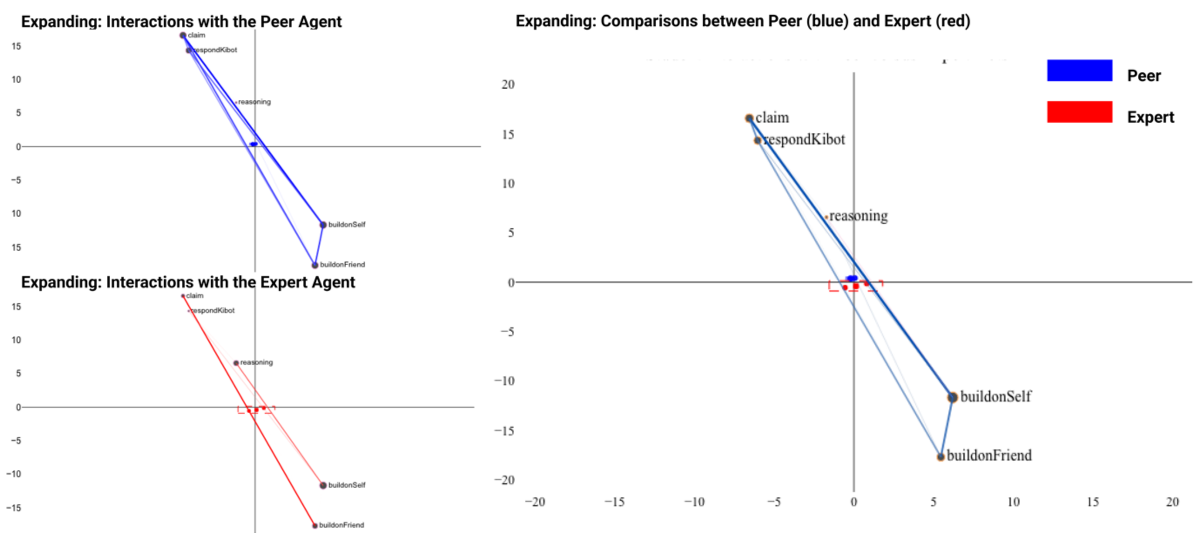
Figure 4.3
Comparisons of Expanding group between Expert (Red) and Peer (Blue) Agents



Notes. Dots present individual students, and lines present co-occurrences between discussion moves. Right panel: The darker colors of the lines show differences in co-occurrences between agent conditions.

Figure 4.4

Comparisons of Mixed group between Expert (Red) and Peer (Blue) Agents



Notes. Dots present individual students, and lines present co-occurrences between discussion moves. Right panel: The darker colors of the lines show differences in co-occurrences between agent conditions.

Interestingly, the Mixed group (Figure 4.4) showed mixed patterns in comparisons of the agent prototypes. Similar to the Expanding group, this group showed more connections between *Respond to Kibot – Transactive* and *Respond to Kibot – Reasoning* in the less-knowledgeable-peer condition. At the same time, individuals in this group showed more links between *Externalize – Transactive* in the expert condition (right panel, Figure 4.4). Notably, the coordinates of individual students for the expert agent (red dots) were close to one another. This shows that interactions with this agent might have been similar for individuals within the group.

Discussion

Grouping Arrangements Showed Different Discussion Patterns

In this research, I examined grouping based on prior domain knowledge as one factor that may be associated with group interactions. Overall, students in groups with expanding prior domain knowledge showed more reasoning and claim-making, in combination with elaboration on prior ideas. Transactive exchange, where students build on each other's knowledge to co-construct ideas, is a key practice in knowledge building (Hewitt & Scardamalia, 1998). Such exchange allows students to discuss the concepts at hand in depth to advance individual learning (Van Boxtel et al., 2000; Yang et al., 2016). Note that the focus of the study on knowledge composition is not to recommend one grouping model over others (e.g., heterogeneous versus homogenous grouping). Rather, I was interested in exploring the potential differences in group interactions to provide insights for future agent designs and research directions.

Findings of the different interaction patterns between the groups may suggest the design needs to provide further scaffolds on knowledge building for certain groups, instead of providing nudges at similar intervals across groups. Scaffolds can be reflective prompts that explicitly guide students to think about how they are putting their ideas together and present the group as a community (Yang et al., 2016). These prompts may help students to focus on knowledge building goals and improve the quality of the group discussions over time.

Another direction is to provide prompts that adapt to groups' ongoing discourse. The Kibot agents gave conceptual hints that adapted to the groups' knowledge states, through comparing their ongoing concept maps with an underlying expert map. To emphasize equal participation, the agents constantly directed their nudges at the less active participants. However, students' uptake of the nudges varied. It is possible that the agents' nudges were not opportune or sufficient to prompt groups with emergent understanding to engage in discussions.

Other systems have attempted to address this issue by using natural language processing to categorize students' chats and dynamically select the agents' messages from candidate talk moves (Adamson & Rosé, 2012; Adamson et al., 2014). Researchers have found the promise of such dynamic conversational systems among different age groups, including high school and college students (Adamson et al., 2014). The adaptive talk moves can also be combined with fading when students monitor their learning to decide on the gradual removal of learning scaffolds (Wecker & Fischer, 2011). Fading talk moves can help students to internalize collaborative discussion strategies (Vogel et al., 2017). Future empirical work can explore how conversational agents can be integrated into these adaptive systems to support learners within the same classrooms or discussion groups.

Group Interactions Diverged with the Agent Designs

Findings around grouping arrangements warrant the need for investigating how student groups' exchange varies with agent designs. Both designs are assumed to elicit scientific discourse from students. Each agent's introductions frame students' role as an explainer (expert agent: "*Show me what you know*"; less knowledgeable peer agent: "*I'm here to learn with you.*"). Explicit framing of students' role discussion may encourage engagement and improved outcomes across groups with different domain knowledge (Saleh et al., 2005).

Interestingly, findings from the current study reveal that interactions with the two agent designs were not similar across groups with different domain knowledge compositions. Students in the emergent group more frequently provided reasoning when responding to the expert version. Meanwhile, those in the expanding groups engaged in more transactive exchange with the less knowledgeable peer agent. These findings overlap with hypotheses around student-agent interactions in one-on-one tutoring systems (Graesser, 2016). Students with expanding levels of understanding may engage in deeper elaboration of ideas when they are trying to provide help for a less capable peer agent, and those with emergent domain knowledge may more frequently respond to nudges from the expert agent.

Analyses of data from the mixed group suggest that these patterns were likely taken up by the group, as opposed to being instigated by an individual. This means that the students with emergent knowledge also became explainers to the less knowledgeable peer agent. Thus, combining the two agent designs may result in a richer set of discussion patterns in the mixed group. A follow-up analysis can follow a larger sample of heterogeneous groups and explore which types of social dynamics may support shared interaction norms. Future work can also formally test the pathways between agent conditions, student interactions, and learning outcomes while accounting for different group compositions.

Conclusion

Prior work has explored design paradigms in learning exchange between individual students and conversational agents (Kim & Baylor, 2016; Liew et al., 2013). Interactions likely diverge in group settings, where individuals are influenced by the participation norms from others in the group. The current study illustrates how interactions are embedded within social structures such as group compositions of prior knowledge. Analyses suggest the affordances of ENA in comparing group interactions across dimensions of group compositions and agent designs.

A limitation to the current research is that it only presents a small sample of cases. Future research can examine whether these patterns exist in a larger sample, particularly in groups with mixed knowledge bases. Iterations of this work will apply other research designs, such as between-subject designs, to examine the pathway between varied group interactions and student learning. These analyses can also account for variables such as students' gender, social status, and participation.

In addition, student interactions with the conversational agents were brief because they were constrained within one class period. A direction for future exploration is to track how the observed interactive patterns evolve, as students gain more exposure to the agents' nudges. Another direction is to examine whether students transfer the discussion moves to another task without the appearance of the agents. Finally, the current research focuses on students' prior domain knowledge. Future work can consider adaptive designs that account for evolving group interactions and students' knowledge states.

Overall, this study reveals interaction dynamics when collaborative agents are embedded in group discussions. The patterns that this research uncovers broaden the design space for learning systems. They also suggest the learning moments that facilitators such as teachers and conversational agents may consider to develop appropriate support for productive discourse across group compositions.

References

Adamson, D., Dyke, G., Jang, H., Rosé, C. P. (2014). Towards an agile approach to adapting dynamic collaboration support to student needs. *International Journal of Artificial Intelligence in Education 24*(1), 92-124.

Adamson, D., Rosé, C. P. (2012). Coordinating multi-dimensional support in collaborative conversational agents. In Cerri, S. A., Clancey, W., Papadourakis, G. & Panourgia, K. (Eds.), *Proceedings of the 11th international conference on Intelligent Tutoring Systems* (pp. 346-351). Springer, Berlin, Heidelberg.

Agne, R. R., Muller, H. L. (2019). Discourse strategies that co-construct relational identities in STEM peer tutoring. *Communication Education 68*(3), 265–286.

Biswas, G., Segedy, J. R., Bunchongchit, K. (2016). From design to implementation to practice a learning by teaching system: Betty's brain. *International Journal of Artificial Intelligence in Education*, *26*(1), 350–364.

Byrne, D., Nelson, D. (1965). Attraction as a linear function of proportion of positive reinforcements. *Journal of Personality and Social Psychology, 1*(6), 659–663.

Chaiken, S., Maheswaran, D. (1994). Heuristic processing can bias systematic processing: Effects of source credibility, argument ambiguity, and task importance on attitude judgment. *Journal of Personality and Social Psychology 66*(3), 460–473.

Chen, H., Park, H. W., Breazeal, C. (2020). Teaching and learning with children: Impact of reciprocal peer learning with a social robot on children's learning and emotive engagement. *Computers and Education*, *150*, 103836.

Chi, M. T. H., Siler, S. A., Jeong, H., Yamauchi, T., Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science, 25*(4), 471–533.

Chin, C. (2006). Classroom interaction in science: Teacher questioning and feedback to students' responses. *International Journal of Science Education, 28*(11), 1315–1346.

Dyke, G., Howley, I., Adamson, D., Kumar, R., Rosé, C. P. (2013). Towards academically productive talk supported by conversational agents. In Suthers, D. D., Lund, K., Rosé, C. P., Teplovs, C., & Law, N. (Eds.), *Productive multivocality in the analysis of group interactions* (pp. 459–476). Springer, US.

Gašević, D., Joksimović, S., Eagan, B. R., Shaffer, D. W. (2019). SENS: Network analytics to combine social and cognitive perspectives of collaborative learning. *Computers in Human Behavior, 92*, 562–577.

Gillies, R. M., Nichols, K., Burgh, G., Haynes, M. (2014). Primary students' scientific reasoning and discourse during cooperative inquiry-based science activities. *International Journal of Educational Research, 63*, 127-140.

Graesser, A. C. (2016). Conversations with AutoTutor help students learn. *International Journal of Artificial Intelligence in Education*, *26*(1), 124–132.

Heidig, S., Clarebout, G. (2011). Do pedagogical agents make a difference to student motivation and learning? *Educational Research Review*, *6*(1), 27–54. Elsevier.

Hewitt, J., Scardamalia, M. (1998). Design principles for distributed knowledge building processes. *Educational Psychology Review, 10*(1), 75-96.

Hoadley, C. M., Kilner, P. G. (2005). Using technology to transform communities of practice into knowledge-building communities. *ACM SIGGROUP Bulletin, 25*(1), 31–40.

Howley, I., Kumar, R., Mayfield, E., Dyke, G., Rosé, C. P. (2013). Gaining insights from sociolinguistic style analysis for redesign of conversational agent based support for collaborative learning. In Suthers, D. D., Lund, K., Rosé, C. P., Teplovs, C., & Law, N. (Eds.), *Productive multivocality in the analysis of group interactions* (pp. 477–494). Springer.

Kim, S., Lee, J., Gweon, G. (2019). Comparing data from chatbot and web surveys effects of platform and conversational style on survey response quality. In Brewster, S., Fitzpatrick, G., Cox, A., & Kostakos, V. (Eds.), *Proceedings of the conference on human factors in computing systems* (pp. 1–12). ACM, New York.

Kim, Y. (2007). Desirable characteristics of learning companions. *International Journal of Artificial Intelligence in Education, 17*(4), 371–388.

Kim, Y., Baylor, A. L. (2006). A social-cognitive framework for pedagogical agents as learning companions. *Educational Technology Research and Development, 54*(6), 569–596.

Kim, Y., Baylor, A. L. (2016). Research-based design of pedagogical agent roles: A review, progress, and recommendations. *International Journal of Artificial Intelligence in Education, 26*(1), 160–169.

King, A., Staffieri, A., Adelgais, A. (1998). Mutual peer tutoring: Effects of structuring tutorial interaction to scaffold peer learning. *Journal of Educational Psychology, 90*(1), 134–152.

Kumar, R., Beuth, J., Rose, C. (2011). Conversational strategies that support idea generation productivity in groups. In Spada, H., Stahl, G., Miyake, N., Law, N. (Eds.), *Connecting computer-supported collaborative learning to policy and practice*: *CSCL2011 conference proceedings* (pp. 398–405). International Society of the Learning Sciences.

Kumpulainen, K., Wray, D. (2002). Classroom interaction and social learning: From theory to practice. *Psychology Press*.

Liew, T. W., Tan, S.-M., Jayothisa, C. (2013). The effects of peer-like and expert-like pedagogical agents on learners' agent perceptions, task-related attitudes, and learning achievement. *Educational Technology & Society, 16*(4), 275–286.

Marquart, C. L., Zachar, S., Collier, W., Eagan, B, Woodward, R., Shaffer, D.W. (2018). rENA: Epistemic Network Analysis. https://cran.r-project.org/web/packages/rENA/index.html

Moon, Y., Nass, C. I. (1996). Adaptive agents and personality change. In Tauber, M. (Ed.), *Conference companion on human factors in computing systems common ground - CHI '96* (pp. 287–288).

Muhonen, H., Rasku-Puttonen, H., Pakarinen, E., Poikkeus, A. M., Lerkkanen, M. K. (2017) Knowledge-building patterns in educational dialogue. *International Journal of Educational Research, 81*, 25-37.

Nass, C., Moon, Y., Fogg, B. J., Reeves, B., Dryer, D. C. (1995). Can computer personalities be human personalities? *International Journal of Human-Computer Studies, 43*(2), 223–239.

Nass, C., Steuer, J., Tauber, E. R. (1994). Computers are social actors. In Adelson, B., Dumais, S., & Olson, J. (Eds.), *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 72–78). ACM, New York.

Nguyen, H., Santagata, R. (2020). Impact of computer modeling on learning and teaching systems thinking. *Journal of Research in Science Teaching, 58*(5), 661-688.

Oliveira, A. W., Sadler, T. D. (2008). Interactive patterns and conceptual convergence during student collaborations in science. *Journal of Research in Science Teaching, 45*(5), 634-658.

Patton, M. Q. (2014). *Qualitative research & evaluation methods: Integrating theory and practice*. Sage.

Rosé, C. P., Wang, Y. C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., Fischer, F. (2008). Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning, 3*(3), 237–271.

Rosenberg-Kima, R. B., Baylor, A. L., Plant, E. A., Doerr, C. E. (2008). Interface agents as social models for female students: The effects of agent visual presence and appearance on female students' attitudes and beliefs. *Computers in Human Behavior, 24*(6), 2741–2756.

Saleh, M., Lazonder, A. W., De Jong, T. (2005). Effects of within-class ability grouping on social interaction, achievement, and motivation. *Instructional Science, 33*(2), 105-119.

Scardamalia, M., Bereiter, C. (1993). Technologies for knowledge-building discourse. *Communications of the ACM, 36*(5), 37–41.

Seering, J., Luria, M., Kaufman, G., Hammer, J.. (2019). Beyond dyadic interactions: considering chatbots as community members. In Brewster, S., Fitzpatrick, G., Cox, A., & Kostakos, V. (Eds.), *Proceedings of the conference on human factors in computing systems* (pp. 1–13).

Shaffer, D. W., Collier, W., Ruis, A. R. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics, 3*(3), 9-45.

Stahl, G., Cress, U., Law, N., Ludvigsen, S. (2014). Analyzing the multidimensional construction of knowledge in diverse contexts. In *International Journal of Computer-Supported Collaborative Learning*, *9*(1), 1–6. Springer, New York.

Tegos, S., Demetriadis, S. (2017). Conversational agents improve peer learning through building on prior knowledge. *Educational Technology & Society, 20*(1), 99–111.

Van Boxtel, C., Van der Linden, J., Kanselaar, G. (2000). Collaborative learning tasks and the elaboration of conceptual knowledge. *Learning and Instruction, 10*(4), 311-330.

Vogel, F., Kollar, I., Ufer, S., Reichersdorfer, E., Reiss, K., & Fischer, F. (2016). Developing argumentation skills in mathematics through computer-supported collaborative learning: the role of transactivity. *Instructional Science*, *44*(5), 477–500. https://doi.org/10.1007/s11251-016-9380-2

Walker, E., Rummel, N., Koedinger, K. R. (2014). Adaptive intelligent support to improve peer tutoring in algebra. *International Journal of Artificial Intelligence in Education, 24*(1), 33–61.

Webb, N. M., & Farivar, S. (1994). Promoting helping behavior in cooperative small groups in middle school mathematics. *American Educational Research Journal, 31*(2), 369-395.

Wecker, C., & Fischer, F. (2011). From guided to self-regulated performance of domain-general skills: The role of peer monitoring during the fading of instructional scripts. *Learning and Instruction*, *21*(6), 746–756. https://doi.org/10.1016/J.LEARNINSTRUC.2011.05.001

Yang, Y., van Aalst, J., Chan, C. K., Tian, W. (2016). Reflective assessment in knowledge building by students with low academic achievement. *International Journal of Computer-Supported Collaborative Learning, 11*(3), 281-311.

Zacharia, Z., Barton, A. C. (2004). Urban middle-school students' attitudes toward a defined science. *Science Education, 88*(2), 197-222.

CHAPTER 5 SYSTEMS THINKING WITH AGENTS

Abstract

Conversational agents can facilitate learning discussions, by applying natural language understanding to process students' discourse. Agents can assume the personalities of figures such as peers or mentors, to promote help-giving actions similar to human interactions. In this study, I explore how and for whom different agent designs can facilitate discussion patterns and systems thinking in small-group discussions. Participants included 172 students in 9th grade (ages 13-14). Participants were randomly assigned to groups of five students and interacted with no agent, an expert agent, or a less knowledgeable peer agent. Results suggest that both agents facilitated learning of systems mechanisms by enhancing transactive exchange, where students built on prior ideas. I also found differences in the agents' effects on discussion and learning outcomes based on groups' variation in systems thinking pretest. These findings have implications for the design and application of collaborative learning agents.

Keywords. collaborative learning, systems thinking, conversational agents, multilevel model

Introduction

Small-group discussions have been associated with positive attitudes and learning outcomes (Lou et al., 1996). Educators frequently employ group discussions to teach about complex systems that involve multiple components and interrelationships (Jordan et al., 2013; Nguyen & Santagata, 2020; Yoon, 2007). However, students do not always participate equitably in group discussions. Discussions are often dominated by students who are more active or possess a higher level of domain knowledge (Saleh et al., 2007; Yun & Kim, 2015). In addition, unfacilitated group discussions can veer off-task (Hogan et al., 1999). In response, researchers have developed conversational agents (also known as pedagogical agents) that can provide real-time nudges for participation and on-task knowledge development (Adamson et al., 2014; Dyke, Adamson, et al., 2013; Tegos & Demetriadis, 2017).

Agents can pose as companions, assistants, or mentors to promote learning interactions (Kim et al., 2020; Kim & Baylor, 2006, 2016; Nguyen, 2021). This line of research with individual learners assumes that intentional agent designs can prime users to show responses such as help-giving and help-seeking, similar to human-human conversations (Nass et al., 1995). Productive interactions from agent-facilitated discussions can facilitate students' learning (Howley et al., 2013; Kinnebrew et al., 2014; Ogan, Finkelstein, Mayfield, et al., 2012). However, there is little research that has examined how varying agent designs might impact interactions and learning outcomes in small-group discussions. This understanding can provide insights into the mechanisms in which technologies such as conversational agents can support productive discussion moves and learning.

In this study, I explore how different conversational agent designs can facilitate discussion patterns and learning about marine ecosystems. The two designs include an expert and a less knowledgeable peer agent. Drawing on prior evidence that students with different baseline understanding may interact in group discussions in varied ways (Nguyen, 2021), I examine how the paths between agent conditions, discussion patterns, and learning outcomes differed along different values of individuals' and groups' baseline understanding. To answer these questions, I employ a randomized experimental design with 172 students, where students randomly joined groups and interacted as a group with different agent conditions (58 in no-agent, 59 in expert, 55 in less-knowledgeable-peer condition).

Background

This study focuses on the link between conversational agent designs, discussion patterns, and learning. To ground this discussion, I draw from research in science education to outline the learning outcome: enhanced systems thinking to understand a marine ecosystem. I further review the discussion moves that contribute to this outcome and how agents may support these moves.

Systems Thinking

Systems thinking entails knowledge of systems' structures, the relationships between those structures, and the functions that they serve (Hmelo-Silver et al., 2017; Liu & Hmelo-Silver, 2009; Snapir et al., 2017). Systems thinking is important because systems are ubiquitous at both micro-level such as cells and macro-level such as ecosystems (Yoon et al., 2016). For example, interactions between organisms in habitats and across species can be described as systems of multiple components, interrelations, and processes (Hmelo-Silver, Duncan, et al., 2007; Wilensky & Resnick, 1999). Changes in a component, such as a surge in the amount of human-induced carbon dioxide (CO_2) in the atmosphere, can lead to ocean acidification and negatively impact other marine organisms.

Researchers have conceptualized systems thinking to help students understand key concepts and processes (Danish et al., 2017; Hmelo-Silver et al., 2017; Hmelo-Silver, Marathe, et al., 2007; Liu & Hmelo-Silver, 2009). Hmelo-Silver et al. (2017) proposed a framework that includes Components (elements within the system), Mechanisms (the processes and interactions happening between components), and Phenomena (the system's macro patterns). Within the context of marine ecosystems, examples of Components may include descriptions of components and their locations (e.g., "habitat contains fish"). Mechanisms describe the relationships between multiple components, for example, photosynthesis will increase with an increase in phytoplankton. Finally, Phenomena may refer to biodiversity, for instance, connecting processes of habitat destruction, invasive species, and overfishing to biodiversity loss.

Developing a coherent understanding of Components, Mechanisms, and Phenomena can be challenging. Learners tend to emphasize the visible components of a system more often than invisible, interrelated processes and focus on linear, simple causal mechanisms (Gellert, 1962; Wilensky & Resnick, 1999). Complex causal links, in comparison, may involve several implicit mechanisms that impact components in different directions (Jacobson & Wilensky, 2006). For instance, ocean acidification increases photosynthesis (through increases in carbon) and phytoplankton. At the same time, acidification prevents certain phytoplankton from creating their calcium carbonate skeletons, thus reducing the number of those organisms.

Group Discussions Promote Systems Thinking

Group discussions can promote systems thinking by facilitating rich interactions about systems with peers and experts (Jacobson & Wilensky, 2006; Paus et al., 2012; Scardamalia & Bereiter, 1993; Yoon, 2007). Complex systems behaviors emerge through iterative knowledge construction, where students examine others' ideas to refine their assumptions and contribute to high-level ideas. Researchers have identified discussion moves that contribute to systems thinking, with a particular focus on idea formation, transactive exchange, and questioning (Cohen, 1994; Dyke, Howley, et al., 2013; Michaels et al., 2008; Oliveira & Sadler, 2008; Resnick et al., 2010; van Aalst, 2009; Yoon, 2007).

Students create *claims* and *hypotheses* to articulate their understanding and simulate relationships within systems. Students provide evidence from scientific facts and personal experiences to justify those claims (Hogan et al., 1999; Jacobson & Wilensky, 2006), and extend and challenge peers' statements to revise their claims (Oliveira & Sadler, 2008; Webb, 1982). In the process of building on peers' ideas, students engage in transactive exchange, enriching prior group reasoning through collective efforts.

Transactive exchange differentiates knowledge construction from knowledge sharing and contributes to more coherent systems understanding (Dyke, Howley, et al., 2013; Fu et al., 2016; Teasley, 1997; van Aalst, 2009). Idea synthesis requires elaboration on previous ideas by justifying these ideas with logical reasoning and evidence (Michaels et al., 2008). Students engage with peers' ideas with more complexity and show enhanced learning when iteratively probed to explain their thinking (Michaels et al., 2008; Paus et al., 2012). Resnick et al. (2010) have identified talk moves that promote transactive exchange, for instance, asking students to elaborate or add to a peer's contribution (e.g., "Say more") and to explain why they agree or disagree with another statement (e.g., "Do you agree or disagree? Why?"). Such discussion moves sustain inquiry-driven discourse and produce relevant knowledge for all group members (Fu et al., 2016).

Discussions can further benefit from *questions* that prompt students to consider the underlying mechanisms that lead to the observed systems' behaviors (King, 1999; Mäkitalo et al., 2005; Webb, 1989). Reflective questions are open-ended, to elicit information and generate alternative perspectives for follow-up discussions (Aguiar et al., 2010; Chin & Osborne, 2010; Gillies, 2016; Hmelo-Silver & Barrows, 2008; Hogan et al., 1999; King, 1999). Questions also highlight gaps in the groups' knowledge, as students attempt to connect with others' ideas. In one study, Chin & Osborne (2010) analyzed the written arguments of 29 student groups and found that successful groups focused on asking questions about key concepts during discussions. These authors argued that questions might contribute to more structured group understanding by probing into alternative perspectives, eliciting elaboration, and clarifying uncertainties.

Prior domain knowledge might influence the dynamics of group discussions (Cohen, 1994; Gillies, 2003; Lou et al., 1996; Peterson et al., 1981). Compared to students with average knowledge, students with emergent and expanding prior knowledge tend to benefit more from spontaneous peer tutoring in small-group interactions (Cohen, 1994; Peterson et al., 1981). Students with emergent prior knowledge may learn from peers' explanations, while those with expanding knowledge may benefit from restructuring their understanding to provide explanations to others (Cohen, 1994; Wittrock, 1989). Further, students with expanding knowledge more frequently lead discussions (Yun & Kim, 2015). Researchers have thus developed learning systems and task structures to promote more equitable participation and support specific discussion moves that contribute to learning for all students.

Conversational Agents Enrich Discussions

Conversational agents can facilitate systems thinking (Biswas et al., 2016; Kinnebrew et al., 2014; Tegos & Demetriadis, 2017). For example, Biswas and colleagues (2010) developed two agents (a tutee and a mentor) in a learning system called Betty's Brain to help students construct concept maps. In one-on-one interactions with the agents, students iteratively built the concept maps, quizzed the tutee agent, and received feedback from the mentor agent to refine incorrect connections. Other researchers have developed more open-ended systems. Instead of giving feedback, the agents prompted students to elaborate on ideas and provide argumentation based on peers' ideas (Dyke, Adamson, et al., 2013; Dyke, Howley, et al., 2013).

These agents' prompts can promote discussions. As an example, Adamson et al. (2014) integrated into collaborative settings an agent that asked students to revoice ideas and found that the agent precipitated discussions where students revoice assertions more frequently. Scholars have developed agents to promote transactive exchange for students to build on and challenge each other's ideas (Howley et al., 2014), foster relationships (Ogan, Finkelstein, Mayfield, et al., 2012; Rosé et al., 2015), and regulate individual and group efforts (Harley et al., 2012, 2017).

Previous work has primarily explored agents' impact from cognitive perspectives, such as the extent to which the agents promote learning (Adamson et al., 2014; Liew et al., 2013). Scholars have also applied research in human-computer interaction to consider the design of student-agent interactions. This line of work primarily builds on the Computers as Social Actors (CASA) principle to position agents as social partners (Moon & Nass, 1996; Nass et al., 1995; Nass & Moon, 2000). This principle suggests that subtle design cues can prime users to show conversational norms with computer systems in ways that resemble human-human interactions.

For example, researchers have hypothesized that students with expanding knowledge may provide more explanation to teach a less competent agent, compared to an expert prototype (Graesser, 2016). In one study, female engineering students demonstrated higher interest and efficacy in engineering after one-on-one interactions with an agent that resembled a peer than a knowledgeable agent (Rosenberg-Kima et al., 2008). The authors argued that these students potentially related better to the peer agent's values and knowledge states.

There are limited examples of varying agent designs in collaborative settings. Nguyen (Study 1) analyzed high school students' interactions with two agents (an expert and a less knowledgeable peer). Student groups showed different interaction sequences with the agents. Groups more frequently posed questions to test a less experienced peer agent's knowledge and then built on prior group exchange to explain concepts, compared to exchange with an expert agent. This research illustrates that different designs might promote varied discussion dynamics.

In sum, conversational agents have shown promise in promoting discussion moves. However, limited research has linked agent designs to discussion moves and examined the learning implications of these associations. Discussion patterns can influence students' learning (Jacobson & Wilensky, 2006; Kim et al., 2020). Exploratory discourse, such as posing questions and building on prior ideas, can deepen knowledge over time (King, 1999; Mäkitalo et al., 2005; Webb, 1989). Responding to the agents' nudges—similar to how students may interact with a teacher—can help to construct more complex reasoning if such interactions span beyond superficial exchange (Hogan et al., 1999). It is thus crucial to explore the link between agent designs, discussion moves, and learning outcomes in group settings.

Research Questions

This study employed a cluster-randomized design to explore the effectiveness of different agent designs on learning. Students were randomly assigned to groups, and the groups chatted with no agent, an expert agent, or a less knowledgeable agent. I examined the following questions:

RQ1. To what extent do conversational agents promote systems understanding?

RQ2. To what extent do conversational agents facilitate discussion patterns such as idea formation, transactive exchange, and questioning? To what extent is there a difference between the expert and less-knowledgeable-peer agents in facilitating discussion patterns?

RQ3. Do discussion patterns serve as a mechanism through which conversational agents affect systems understanding?

RQ4. How do agents' effects on systems thinking and discussion patterns differ with group compositions, based on students' prior domain knowledge?

I hypothesized that students interacting with the agents would show higher levels of systems understanding, particularly in terms of systems mechanisms (H1). This hypothesis drew from prior work on the positive association between agent integration and students' learning (Dyke et al., 2013; Tegos & Demetriadis, 2017), and on the specific focus on systems relationships of the agents in this study (Nguyen, 2021). Furthermore, students who interact with the agents would engage in more transactive exchange and ideation (H2a; Dyke et al., 2013; Rosé et al., 2015). Students who interact with the less-knowledgeable-peer agents would pose more questions, compared to the other conditions (H2b, based on sequences of quizzing the peer agent in prior investigations; Study 1). For RQ3, I hypothesized that discussion patterns would mediate the relationship between agents and systems thinking (H3). These patterns have been identified as key factors to promote productive discussion and learning (Hogan et al., 1999; Oliveira & Sadler, 2008; Wilensky & Resnick, 1999). RQ4 presents an exploratory analysis that may provide insights into the designs of agents to support different group dynamics.

Methodology

Study Settings

The context of this work was a high school environmental science program, representing a multi-year partnership between a local state park, education and biology researchers, and local school districts in the southwestern United States. The state park had been studying how reduced human impacts in a marine protected area might increase biodiversity. The environmental science program was an educational outreach to deepen students' understanding of local biodiversity phenomena. The program consisted of eight lessons, where students learned about the marine protected area (MPA) and explored the elements and processes that might affect biodiversity. Students collected and analyzed data on fish and water quality to give recommendations to the park on the effectiveness of the MPA. The learning tasks with the agents took place in lesson 3, where students worked in small groups to build concept maps of the marine ecosystem. The agents helped students reason through the connections in their maps.

Agent Designs

The agent designs were detailed in Study 1. Using natural language processing, the agents grouped students' chats into subject-object pairs to represent connections within the marine ecosystem (e.g., "kelp"-"photosynthesis"), compared students' maps to an expert map, and provided hints to promote articulation of systems mechanisms and participation in the discussions. In a between-subject design, student groups interacted with an expert agent, a less knowledgeable peer agent, or no agent. Both agents were text-based and appeared as robots with human characteristics. The agents' prompts had the same underlying functions, but they used different visual and verbal cues. Figure 5.1 shows the interface and the experiment conditions.

The Expert Agent. The expert agent had a "scientist" tag on its coat, and introduced itself as a bot who "accompanies scientists in the field, and has learned about ocean-related concepts". This agent challenged students to show what they knew about ocean ecosystems. For example, the agent framed its elaborative prompts as "Tell me more to make sure you understand" or "Why don't you talk to your friends to see if you both agree?" The agent referred to the knowledge bar that showed how many connections the groups had built, and sent prompts to acknowledge students' efforts, for example, "Keep going" or "Good progress".

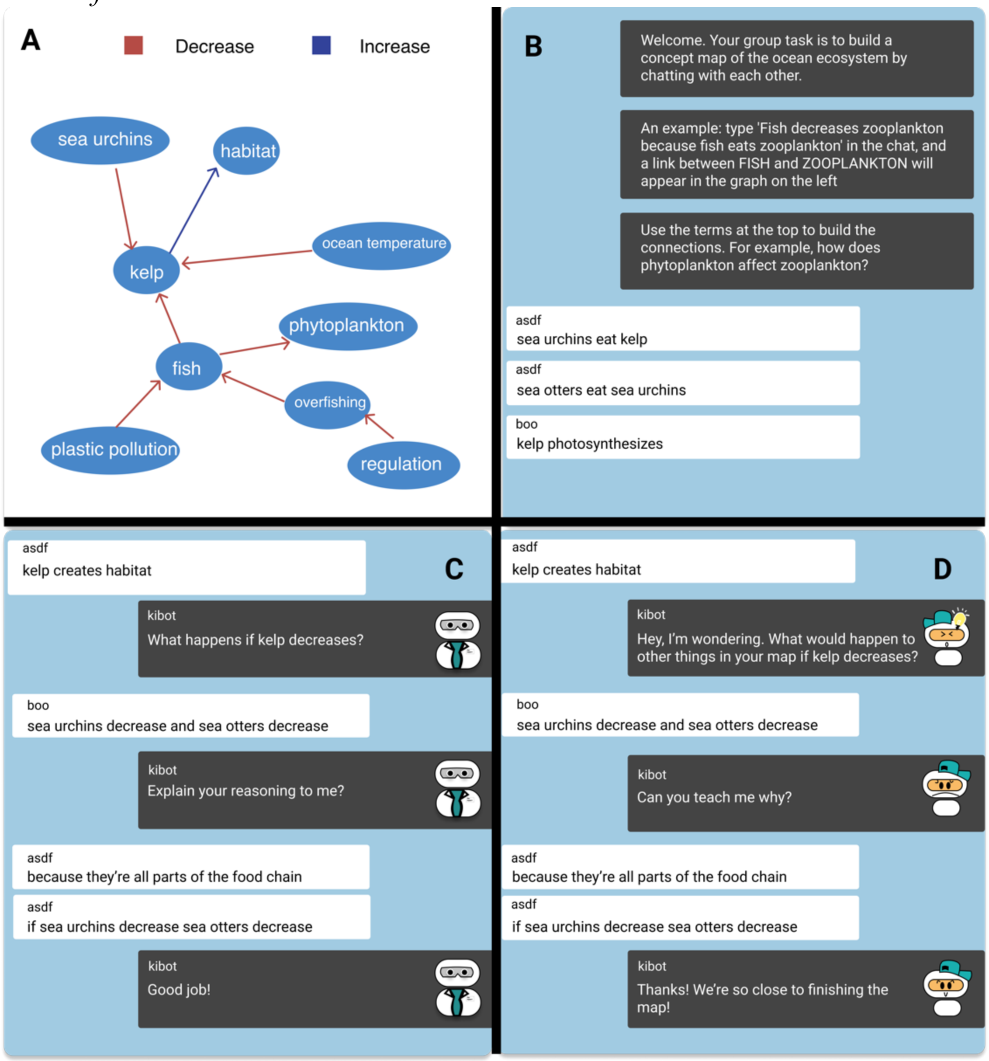
The Less Knowledgeable Peer Agent. Meanwhile, the less knowledgeable peer agent resembled a peer who has "never been to the ocean" and "is ready to learn with the group". This agent framed its prompts as "I don't understand. Can you explain more?" or "Can you build on what your friends just said to help me understand?" The agent referred to the knowledge bar with prompts such as "You're doing great! The more connections you name, the more I learn". Similar to prior work with one-on-one tutoring systems to animate the peer agents (e.g., Kim & Baylor, 2016), the agent changed its facial expressions when it was confused ("frowning"), came up with an idea ("holding a lightbulb"), or acknowledged the group ("smile").

In sum, the agents sent the same conceptual nudges to students at similar frequencies. However, their design and vocabulary varied to give impressions of different levels of competence. The agent designs were validated through user testing and semi-structured focus groups with 17 participants (convenience sample of high school students, researchers, and college graduates). All participants identified the agents' roles, competence, and emotions as intended.

Participants

Participants came from six 9th grade classes in a high school in the southwestern U.S. All participants were taught by the same marine science teacher, who had a long-term partnership with the state park in the MPA program. The school had a diverse student population (47.2% White, 38.1% Hispanic, 8% Asian; 23.5% on free and reduced lunch in the 2020-21 school year). Students completed the task in small groups (average size of five), with roughly equal numbers of students assigned to each condition. Three students had the pretest missing and were excluded from analyses, final N = 36 groups (172 students); no agent: 11 groups (58 students); expert: 12 groups (59 students); and less knowledgeable peer: 13 groups (55 students). Female students accounted for 52% of the sample.

Figure 5.1.
The Agents' Interface.



Notes. Panel A: Concept maps of the ecosystem, dynamically created from students' chats; panel B: no-agent condition with just introduction of the task; panel C: expert; panel D: less-knowledgeable-peer

Power Analyses

Before data collection, I conducted a power analysis using the PowerUp tool (Dong & Maynard, 2013) for a 2-level cluster random assignment design (5 students per group, group effects are random). Because students were nested within discussion groups, there might be unobserved effects at the group level that violated the independence of observation assumptions and overestimated the coefficients in the linear models (Raudenbush & Bryk, 2002). Thus, I fitted multilevel models to predict individual learning. I specified power of .80, α = .05, proportion of variance in outcome between groups (ICC) of .10. The power analysis suggested that we would need 51 groups (255 students) to detect a minimum relevant effect size (MRES) of .30 (i.e., difference in posttest learning of at least 30%) and 20 groups (100 students) for an MRES of .50. Thus, the study's sample was sufficient to detect medium to large effects but could be underpowered to detect smaller effects.

Procedures

The day before the conversational agent (CA) activity, students completed a 15-minute pretest on systems thinking. During the CA activity, students were randomly assigned to a group chat with others. Students engaged in a 25-minute chat with each other (no agent; control condition) or with other students plus the agent as an expert (treatment 1) or a less knowledgeable peer (treatment 2). The goal of the discussion was to build a concept map to describe the marine ecosystem. After 25 minutes, a pop-up window guided students to a posttest on systems thinking. The posttest was similar to the pretest.

Instruments

Data from this study came from students' pretest and posttest on systems thinking, group chat logs, and a survey on students' group investment.

Systems Thinking. The pretest and posttest captured students' systems thinking through four open-ended questions about the effects of the marine protected areas on components, interrelations among those components, and ways in which students thought they could improve biodiversity. An example is: "After the Protected Area is introduced, we see an increase in the number of kelp bass. What do you think contributes to this increase? Explain your answer."

I developed a coding scheme for systems thinking based on the Components-Mechanisms-Phenomena framework (Danish et al., 2017; Hmelo-Silver et al., 2017).

Components indicate elements or properties of elements (e.g., "fish lives in coral reef"). *Mechanisms* describe the interactions between system elements and involve mechanistic reasoning (e.g., "if there is more fish, the zooplankton will get eaten"). Similar to Danish et al. (2017), I did not include *Phenomena* as a separate category. This is because the assessment explicitly mentioned the central phenomenon of biodiversity, and students frequently described phenomenon-related mechanisms in their responses (e.g., "Monitoring habitat improves biodiversity"). Thus, descriptions of mechanisms related to biodiversity phenomena were counted toward *Mechanisms*. Table 5.1 shows the codebook. The scores for each component were the sum of statements that counted towards the components.

To establish inter-rater reliability, the author and a second coder went through three iterations. The agent conditions were removed from the coding sample to avoid researchers' bias. The first iteration was based on 5% of the data. Upon resolving disagreements and iterating with another 5% of the data to establish the coding scheme, the two coders reviewed another 20% of the data in the third iteration. We established high agreement in the third iteration (Components: Cohen's $\kappa = .96$; Mechanisms: $\kappa = .93$). The author coded the rest of the data.

Discussion Patterns. The group chats received codes for idea formation, transactive exchange, elaborative questions, procedural questions, and off-task talks (Table 5.2). These codes were developed through two iterations of going through data samples and building on prior work (Chin & Osborne, 2010; Fiacco & Rosé, 2018; Hmelo-Silver & Barrows, 2008; Roschelle & Teasley, 1995; Teasley, 1997; van Aalst, 2009; Weinberger & Fischer, 2006).

Idea formation indicates how students create claims or conjectures about systems connections. Claims are statements that assert facts or describe components and relationships within the marine ecosystems, for instance, "more fossil fuel burning releases CO_2". Conjectures are hypotheses that students propose, for example, “I guess fish will not increase right away, because it takes time to reproduce”.

Transactive exchange focuses on knowledge construction efforts (Roschelle & Teasley, 1995; Weinberger & Fischer, 2006). Students may build on their previous moves (*externalize*) or build on peers’ ideas (*transactive*). Transactive acts occur when students assume a partner’s perspectives or modify previous ideas with elaboration, challenges, and reasoning (Weinberger & Fischer, 2006). Such elaboration can incorporate the use of evidence from observations, personal experiences, and facts (Nguyen & Santagata, 2020). Codes for transactive exchange do not overlap with idea formation. An utterance would be coded under transactive exchange if it is related to a prior idea in the discussion, and “claim” if it is a new idea.

Elaborative questions occur when students elicit elaboration (“What happens next?”). Meanwhile, procedural questions request confirmation about stated connections (*confirmatory,* e.g., “Does fish decrease kelp?”) or the learning tasks (*task;* e.g., “How to make the lines appear?”). Questions were separated into different categories, because elaborative questions may be more conducive to high-level reasoning and enriched learning (Hogan et al., 1999).

Finally, the off-task talks capture instances when the chats were not related to the concept map building task. Off-task discussions can distract students from the task (Hogan et al., 1999). At the same time, off-task conversations such as complaining and exclamations can build rapport in tutoring contexts in both human-human and human-agent exchange (Ogan, Finkelstein, Mayfield, et al., 2012; Ogan, Finkelstein, Walker, et al., 2012).

Researchers have included metacognitive codes such as monitoring progress or setting new goals (Fu et al., 2016; van Aalst, 2009). However, these codes had few occurrences in the sample (12 occurrences, .4% of total chat utterances). It is possible that the short learning task did not enable planning as we might observe in multi-session learning activities. Thus, these codes were not included in the final analysis. The author and a second coder achieved acceptable agreement across dimensions on a sample of 11% of the data (Table 5.2). The coders divided the data in half for coding and resolved uncertain cases through weekly discussions over one month.

Table 5.1.
Systems Thinking Coding Scheme

Category	Definition	Example
Components	Describe location of components	Fish lives in coral reefs.
	Describe properties of components	Kelp forest consists of multiple kelp.
Mechanisms	Describe properties of components in relation to mechanisms	Larger kelp forest allows fish to build larger habitat.
	Identify components that take part in a mechanism	Oxygen production is due to photosynthesis from phytoplankton.
	Describe the mechanisms in which two components interact	When there are more predators more fish will get eaten, so the number of fish goes down.
	Describe the sequences of mechanisms (chain-forward)	With ocean acidification there is more carbon being dissolved in the water. Phytoplankton use the carbon for photosynthesis.

Table 5.2.
Codes for Group Chat included in the Analysis

Code	Description	Example	N	κ
Idea Formation				
Claim	State an idea or concept	Fish decreases plankton	1886	.91
Conjectures	Propose hypotheses about systems connections	I don't think fish will increase right away.	12	.76
Elaborative Question				
Elaborative	Ask for elaboration on connections within systems	What would happen to phytoplankton if there is more CO_2 in the ocean?	38	.72
Procedural Questions				
Confirmatory	Inquire about stated systems connections to resolve uncertainties	Does more fish mean more planktons?	44	.73
Task-related	Gather information about the learning task/ the interface	How do I get the arrows to appear?	18	1
Transactive				
Externalize	Elaborate on one's own prior thoughts to groups	As I mentioned, I think we should also regulate CO_2 emissions.	245	.95
Transactive	Integrate or challenge peers' prior ideas to expand responses	Example 1: What she said. Whales decrease fish because fish is whales' s food source. Whales eat plankton. Example 2: I don't think so. It's kelp that increases photosynthesis, not the other way around.	351	.88
Off-task				
Off-task	State something unrelated to the task, complain, or react to off-task exchange	Example 1: Do you believe in aliens? Example 2: Ughhh boring Example 3: lol	172	1

The codes were dichotomous (1 for occurrence, 0 for non-occurrence). For the analysis, I calculated the sum of items under each discussion pattern. The relatively high correlations of subcodes under each discussion pattern, e.g., $r(170)$ for externalize -transactive code = .84, $p <$ .001 (Appendix A5.1) provide evidence of convergent validity to group the items.

Analytic Strategies

In this study, students were nested within discussion groups and may thus be correlated with one another. Multilevel models that consider groups' random effects can account for this nesting. To determine whether multilevel models were necessary, I calculated the intraclass correlation coefficients (ICCs) for outcome variables of systems thinking and discussion patterns in unconditional models (just group random effects; no covariates). ICCs (range 0, 1) describe the proportion of variance between individuals that can be attributed to their discussion groups, with values closer to 1 suggesting higher variance between groups. The ICCs were small for the systems thinking posttest (.02 for Components, .08 for Mechanisms) and had a wide range for the discussion patterns (.08 for idea formation, .19 for elaborative questions, .10 for procedural questions, .04 for transactive exchange, and .32 for off-task). The larger ICCs justified the decision for the multilevel models, since ignoring the nested structure with even small ICCs can lead to increased Type I errors (Siddiqui et al., 1996).

To answer RQ1 about the effects of agent designs on systems thinking, I fitted multilevel models to predict posttest scores of Components and Mechanisms. At the student level, the models included students' gender, pretest scores for Components and Mechanisms, and the classes they were in (e.g., Period 1-6). At the group level, the model included the agent conditions, with the no-agent condition as the reference group. I modeled the group-level coefficients as random effects to indicate that individuals within groups were not independent from one another. The model equation took the form:

$$y_{ij} = \beta_0 + \beta_1 \text{Agent\_Condition}_{ij} + \beta_k X_{kij} + U_j + R_{ij} \ (1)$$

In equation (1), *i* refers to individual students and *j* refers to the discussion groups. X is a vector of covariates (e.g., gender, pretest scores for Components and Mechanisms, class). U captures the random effect of the discussion groups with estimated variance τ^2, and R denotes the student-level error with estimated variance σ^2. Additionally, I fitted a model with interaction terms between individuals' pretest Mechanisms scores and the agent conditions. This model examined whether the agent had differential effects depending on baseline mechanistic understanding. All continuous variables were z-score standardized before entering the models.

Similar to RQ1, to answer RQ2 about the link between agent conditions and discussion patterns, I fitted multilevel models with the agent conditions as the predictor for each pattern: idea formation, elaborative question, procedural question, transactive exchange, and off-task. An additional model included the interaction between pretest Mechanisms scores and agent conditions. This model examined potential variances in the impact of agent conditions based on baseline Mechanisms understanding. Covariates for the models were similar to those in RQ1.

RQ3 formally examined whether discussion patterns mediated the impact of agents on systems thinking. Researchers have found mediating effects of collaborative elaboration on individuals' conceptual understanding (Paus et al., 2012). For RQ3, I fitted a multilevel mediation model that included the agent conditions with direct paths to systems thinking posttest scores and indirect paths through transactive exchange. Transactive exchange was selected because this discussion pattern was moderately correlated with systems thinking scores (Table

5.3). The model included random effects of discussion groups. Students' gender, pretest scores, and classes were covariates.

RQ4 presented an exploratory analysis of the relationship between group compositions (based on expanding and emergent prior knowledge), agent conditions, and learning outcomes. I built on the multilevel models in RQ1 to fit a model that included an interaction term between agent conditions and groups' standard deviation at pretest. Groups with higher deviation may present more heterogeneous grouping based on prior domain knowledge. This analysis aimed to understand how agents' impact on discussion and learning might differ with group compositions.

For robustness checks, I considered alternative models that also allowed the individuals' pretest systems understanding to vary (i.e., random slope). This means that the effect of independent variables such as baseline understanding on posttest scores varied with discussion groups. However, a series of likelihood ratio tests comparing the random slope and random intercept models suggested that the random slope models did not significantly improve model fits. As a result, I reported results from the simpler random-intercept models as the final models (Appendix A5.2). I adopted Lorah's (2018) guidelines for reporting variance in multilevel models. I calculated R^2 as the proportional reduction in variance at the student level when comparing the unconditional model and the tested models (Snijders & Bosker, 2012). From the R^2 values, I calculated the effect size related to the variance explained by the models (Lorah, 2018), with .02, .15, .35 suggesting small, medium, and large effects (Cohen, 1992). Following Selya et al. (2012), the effect size of the agent condition predictor can be estimated as:

$$f_b^2 = \frac{R_{ab}^2 - R_a^2}{1 - R_{ab}^2} \quad (2)$$

In this equation, R_{ab}^2 represents the proportion of variance of outcome explained by all predictors in the full model (including agent condition), and R_a^2 is the proportion of variance of outcome in a reduced model (similar to the full model, excluding the effect of agent conditions). R^2_{ab}-R^2_a thus denotes the additional proportion of variance of outcome explained by agent conditions.

Because the analyses tested multiple hypotheses, I calculated adjusted p-values using the Benjamini and Hochberg procedure, which controlled for the proportions of all significant tests that are false. The raw p-values obtained from the hypotheses tested (agent conditions and interaction terms) were combined into a vector, ranked and ordered ascendingly, and multiplied by m/k where m is the number of independent tests and k is the position of the p-value in the list.

I used the R packages "lmerTest" and "nlme" (Kuznetsova et al., 2015; Pinheiro et al., 2017) to fit the multilevel models and "emmeans" (Lenth et al., 2019) to conduct pairwise comparisons of agent conditions from the fitted models. The p-values in the multilevel models were calculated using Satterthwaite approximations. This approximation approach produced Type-I error rates close to the acceptable threshold of .05 for smaller samples (Luke, 2017). The mediation analyses used "mediate" (Tingley et al., 2014) which tested the significance of indirect effects through discussion patterns using bootstrapping procedures. The analyses averaged the indirect effects and their confidence intervals from 1,000 bootstrapped samples.

Extensions. I explored the impact of group solidarity, or how individuals identified their closeness with the discussion groups. Group solidarity might impact how individuals approach tasks, for example, affecting how often they engage in the discussion (Leach et al., 2008; Ogan et al., 2011). Because group solidarity is not the focus of this study, this exploratory analysis is presented in Appendix A5.3 as potential extensions in future work.

Furthermore, because Components and Mechanisms posttests were moderately correlated (Table 5.3), I fitted a multivariate multilevel model that accounted for the covariance between these two dependent variables (Appendix A5.4). Overall, results from the univariate and multivariate models did not substantially differ. A similar, multivariate multilevel model to simultaneously predict discussion patterns failed to converge.

Results

Descriptive Statistics

Table 5.3 presents the descriptive statistics for systems thinking and discussion patterns for the overall sample, as well as the correlations between these variables. Students included about 2 statements that described components at post-test ($M = 1.97$, $SD = .92$). On average, students mentioned 4.11 mechanisms ($SD = 1.78$). Students generally sent more chat messages to establish ideas ($M = 10.72$, $SD = 7.97$) than providing transactive exchange ($M = 3.37$, $SD = 3.50$), procedural questions ($M = .35$, $SD = .84$), or elaborative questions ($M = .22$, $SD = .68$). There were also few instances of off-task chats ($M = .97$, $SD = 4.09$).

In terms of correlations, Components and Mechanisms posttests were positively correlated, $r(170) = .55$, $p < .001$. Idea formation and transactive exchange were also positively correlated, $r(170) = .48$, $p < .001$. Transactive exchange was significantly correlated with both Components, $r(170) = .24$, $p < .01$ and Mechanisms, $r(170) = .30$, $p < .001$. More idea formation utterances were linked to higher Components posttest scores, $r(170) = .15$, $p = .04$. More off-task utterances were associated with a lower score for Mechanisms, $r(170) = -.18$, $p = .02$.

Table 5.3.

Descriptive Statistics and Correlations of Posttest and Outcome Measures

	C	M	Idea	EQ	PQ	Trans	OT	S	*M*	*SD*	Range
Component I	1								1.97	.92	0, 4
Mechanism (M)	.55***	1							4.11	1.78	0, 9
Idea	.15*	.12	1						10.72	7.97	0, 42
Elaborative Question (EQ)	.09	.07	-.04	1					.22	.68	0, 5
Procedural Question (PQ)	-.01	-.05	.06	.30***	1				.35	.84	0, 5
Transact (Trans)	.24**	.30***	.48***	.11	.05	1			3.37	3.50	0, 15
Off-task (OT)	-.08	-.18*	.02	.16*	.12	-.08	1		.97	4.09	0, 32

Note. * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 5.4.
Systems Thinking and Discussion Patterns by Agent Conditions

	No Agent	Expert	Peer	Kruskal-Wallis test *p*
Components				
M (SD)	1.97 (.91)	2.13 (.87)	1.78 (.83)	.08
Median (Min, Max)	2.00 (0, 4)	2.00 (1, 4)	2.00 (0, 4)	
Mechanisms				
M (SD)	3.69 (1.59)	4.46 (1.62)	4.18 (2.06)	.06
Median (Min, Max)	4 (0, 8)	4 (2, 9)	4 (0, 9)	
Ideation				
M (SD)	12.14 (9.21)	10.18 (7.36)	9.80 (7.08)	.50
Median (Min, Max)	10 (1, 42)	8 (0, 29)	9.50 (0, 36)	
Elaborative Question				
M (SD)	.22 (.56)	.16 (.55)	.27 (.90)	.56
Median (Min, Max)	0 (0, 3)	0 (0, 3)	0 (0, 5)	
Procedural Question				
M (SD)	.56 (1.07)	.21 (.61)	.29 (.73)	.06
Median (Min, Max)	0 (0, 5)	0 (0, 3)	0 (0, 4)	
Transactive Exchange				
M (SD)	2.49 (3.18)	3.53 (3.53)	4.11 (3.64)	.01*
*Median (Min, Max)	2 (0, 15)	2.5 (0, 14)	3 (0, 15)	
Off-task talks				
M (SD)	2.22 (6.61)	.13 (.46)	.59 (2.16)	.17
Median (Min, Max)	0 (0, 32)	0 (0, 3)	0 (0, 12)	

Note. * $p < .05$, ** $p < .01$, *** $p < .001$.

RQ1. Effects of Agent Conditions on Systems Thinking

Table 5.4 presents the descriptive statistics for systems thinking posttests across the agent conditions. To evaluate the effect of agent conditions, I fitted two univariate multilevel models to predict Components and Mechanisms posttests. The first models included agent condition as the predictor variable (Table 5.5). The ICC values were .04, indicating that discussion group membership accounted for 4% of the variance in Components and Mechanisms posttest scores. The f^2 values of .52 and .23 mean that the models explained 52% and 23% of the variance in posttest Components and Mechanisms, respectively, suggesting a large and medium effect size.

Table 5.5 presents the results. The intercept presents the average standardized scores for a student in Period 1, who was a male with average standardized pretest scores and did not interact with an agent. I found positive effects of the agent conditions for Mechanisms posttest for both the expert (β = .48, *SE* = .18, *t* = 2.75, *p* = .01, adjusted *p* = .04) and less-knowledgeable-peer conditions (β = .50, *SE* = .20, *t* = 2.64, *p* = .01, adjusted *p* = .04). The f^2 values of agent condition is .06 (small effect), suggesting that 6% of the variance in posttest Mechanisms, beyond what was accounted for by group membership, can be attributed to agent conditions. Pairwise comparisons of the marginal means for Mechanisms with Tukey adjustment suggested differences between the agents and the control condition, but did not provide evidence for differences between the two agents; control: *M* = -.35, *SE* = .14, expert: *M* = .13, *SE* = .14, peer: *M* = .18, *SE* = .13, expert-control *p* = .03, peer-control *p* = .04, expert-peer *p* = .97. There was no significant difference in Components posttests for agent conditions, relative to the no-agent condition (expert: β = .18, *SE* = .17, *t* = 1.07, *p* = .30, peer: β = -.06, *SE* = .18, *t* = -.31, *p* = .75).

I further explored the interaction between agent conditions and Mechanisms pretests (Table 5.5). The coefficients for the interaction terms were not significant for Components and Mechanisms posttests. The model fit did not significantly improve when adding the interaction terms, Components $\chi2(2)$ = 3.81, *p* = .58, Mechanisms $\chi2(2)$ = 1.04, *p* = .59. Thus, results did not provide evidence that the effects of agents on posttest differed by baseline understanding.

Table 5.5.
Effects of Agent Conditions on Systems Thinking

	Components	Mechanisms	Components	Mechanisms
Fixed effect	Coeff. (SE)	Coeff. (SE)	Coeff. (SE)	Coeff. (SE)
Intercept	-.25 (.23)	-.53 (.25)*	-.15 (.21)	-.45 (.22)*
Expert	.18 (.17)	.48 (.18)*	.20 (.15)	.44 (.17)**
Peer	-.06 (.18)	.50 (.20)*	-.05 (.17)	.48 (.19)**
Pre-Components	.45 (.10)***	.15 (.11)	.45 (.10)***	.14 (.11)
Pre-Mechanisms	.22 (.10)*	.35 (.11)**	.28 (.13)*	.29 (.14)*
Pre-Mechanisms*Expert			-.08 (.16)	.08 (.18)
Pre-Mechanisms*Peer			-.15 (.16)	.18 (.18)
Random effect	Coeff. (SE)	Coeff. (SE)	Coeff. (SE)	Coeff. (SE)
Between-group var(U_j)	.03 (.01)	.03 (.01)	.02 (.01)	.03 (.01)
Within-group var(R_{ij})	.63 (.04)	.78 (.05)	.64 (.04)	.78 (.05)
f^2	.52	.23	.52	.23

Note. * *p* < .05, ** *p* < .01, *** *p* < .001. The coefficients are standardized. Reference group = No Agent. For all models, covariates included individuals' gender and classrooms' fixed effects

RQ2. Effects of Agent Conditions on Discussion Patterns

Similar to RQ1, I fitted univariate multilevel models that included the discussion patterns as outcome variables. The estimates and standard errors are reported in Table 5.6. The between and within-group variances suggest that individual differences accounted for a larger proportion of variance in the outcome variables than group differences. The higher ICC values (e.g., .23 for off-task exchange and .17 for elaborative questions) suggest the potential existence of unobserved contextual factors at the discussion group level for those variables. The f^2 values indicate a small effect size of the models in explaining the variance at the student level for ideation, transactive exchange, and off-task talks.

Table 5.6.

Effects of Agent Conditions on Discussion Patterns

	Ideation	Elaborative Ques	Procedural Ques	Transactive	Off-Task
Fixed effect	Coeff. (SE)	Coeff. (SE)	Coeff. (SE)	Coeff. (SE)	Coeff. (SE)
Intercept	-1.30 (.32)***	-.09 (.32)	.01 (.29)	-.73 (.25)**	.18 (.31)
Expert	-.09 (.18)	.01 (.25)	-.37 (.22)	.36 (.18)*	-.41 (.25)
Peer	.11 (.19)	.22 (.27)	-.37 (.24)	.56 (.19)**	-.30 (.27)
Random effect	Est. (SE)	Est. (SE)	Est. (SE)	Est. (SE)	Est. (SE)
Between-group var(U_j)	.02 (.01)	.17 (.02)	.08 (.02)	.01 (.01)	.23 (.04)
Within-group var(R_{ij})	.88 (.07)	.83 (.05)	.90 (.05)	.90 (.05)	.58 (.04)
f^2	.12	.01	.01	.10	.11

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. The coefficients are standardized. Reference group = No Agent. For all models, covariates included individuals' gender, Mechanisms pretest, and classrooms' fixed effects.

Table 5.7.

Interaction Effects with Individual Pretest Scores on Transactive Exchange

	Model 1	**Model 2**	**Model 3**
	Coeff. (SE)	Coeff. (SE)	Coeff. (SE)
Intercept	-.73 (.25)**	-.67 (.25)**	-.75 (.24)**
Expert	.36 (.18)*	.23 (.18)	.26 (.18)
Less knowledgeable peer (Peer)	.56 (.19)**	.53 (.20)**	.53 (.20)*
Pretest (Mechanisms)	.28 (.12)*	-.08 (.15)	.36 (.08)***
Expert x Pretest		.37 (.19)$^+$	
Peer x Pretest		.16 (.19)	
Solidarity			
Expert x Solidarity			
Peer x Solidarity			
Group Variation			-.25 (.14)$^+$
Expert x Group Variation			.44 (.20)*
Peer x Group Variation			.20 (.21)
Between-group Var	.01 (.01)	.01 (.01)	.01 (.01)
Within-group Var	.90 (.05)	.90 (.05)	.92 (.05)
f^2	.10	.10	.10

Note. $^+ p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$. Coefficients are standardized. Reference group = No Agent. Covariates included individuals' gender, classrooms, and pretest. Models accounted for groups' random effects.

There were positive effects for transactive exchange for students who interacted with the less-knowledgeable-peer agent, relative to the no-agent condition, β = .56, *SE* = .19, *t* = 2.83, *p* = .01, adjusted *p* =.04. There were also positive associations between transactive exchange and the expert agent condition, compared to the no-agent condition, β = .36, *SE* = .18, *t* = 1.96, *p* = .05, adjusted *p* = .15. The f^2 values of agent condition is .02 (small effect), showing that 2% of the variance in transactive exchange was attributable to agent conditions. Pairwise comparisons of estimated means suggest no significant differences between agent conditions (expert: *M* = .07, *SE* = .13, less knowledgeable peer: *M* = .27, *SE* = .14, *p* = .38).

Using the no-agent condition as the reference group, I found no differences by agent conditions for the other discussion patterns. Pairwise comparisons of the estimated means of discussion patterns did not reveal significant differences between the agent conditions. Overall, students in different agent conditions did not differ significantly in their frequencies of forming ideas, posing questions, or going off-task.

Interaction with Baseline Understanding. Next, I explored whether the effects of agents varied with individuals' baseline understanding. Following the finding that agent conditions had a positive association with transactive exchange, I fitted a multilevel model with transactive exchange as the outcome variable (Model 2; Table 5.7). Predictors included Mechanisms pretest, agent conditions, and interaction term Mechanisms pretest*Condition. The coefficient for the interaction between Mechanisms pretest and the expert agent had a moderate effect size, β = .37, *SE* = .19, *t* = 1.95, *p* = .05, adjusted *p* = .15, suggesting that the effect of the expert agent might be larger for students with higher pretest scores. The model with the interaction term did not have improved fit, $\chi2(2) = 3.80$, *p* = .15. This analysis, therefore, does not provide evidence that agents' effects on transactive exchange differed with pretest.

RQ3. Mechanisms through which Agent Conditions Affect Systems Thinking

To examine the pathways between agent conditions, discussion patterns, and systems thinking, I fitted a multilevel model that included transactive exchange as the mediator. The no-agent condition was set as the reference group. I focused on Mechanisms as the outcome of interest, as there were significant effects of agent conditions on this variable (RQ1). The variable selection also aligned with the study's focus on enriching more complex mechanistic reasoning.

Results suggest that the effect of the agents on Mechanisms posttests was mediated via transactive exchange. There was a significant direct effect between the expert agent and the Mechanisms scores, β = .40 [.04, .78], p = .03. The indirect effect via transactive exchange for the expert agent was also significant, β = .06 [.00, .17], p = .04. The coefficients between the less-knowledgeable-peer condition and transactive exchange and between transactive exchange and Mechanisms post-test scores were significant, leading to an indirect effect of .10 [95% CI .02, .23], p = .01. There was a significant, moderate direct path from the peer agent condition to post-test scores, β = .39 [.05, .79], p = .04. Overall, these results suggest that the effects of the two agents on Mechanisms' understanding can be partially explained by the higher counts of transactive exchange. Figure 5.2 shows the mediation model. To illustrate this finding, consider the following excerpt from a group's interaction with the peer agent.

Kibot: what would happen to the components you've listed if kelp increases?

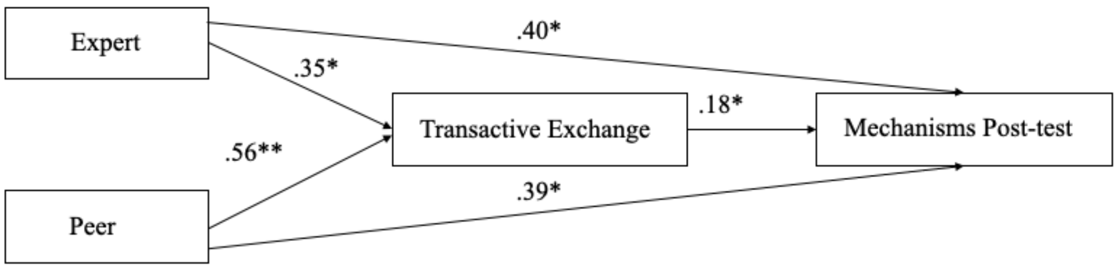
A.R.: photosynthesis decreases

Kibot: what do others think? Do you agree or disagree with what A.R. said?

C.D.: If there is an increase in kelp, there is also an increase in photosynthesis since kelp is a plant and a primary producer, it gets its energy from the sun and do photosynthesis,

C.D.: an increase in kelp and photosynthesis increases the levels of oxygen and pH

Figure 5.2.
Mediation Model of Agent Conditions, Transactive Exchange, and Mechanisms. * $p < .05$



Here, the agent was prompting students to think about components that related to kelp and provided a follow-up for transactive exchange ("what do others think"). Student C.D. responded by challenging her friend's claim that "photosynthesis decreases", providing reasoning that kelp served as a producer, and further elaborating on her ideas by listing the connections with O_2 and pH levels. The agent's prompts and students' responsiveness to such prompts provided the opportunity to approach systems mechanisms coherently (kelp to photosynthesis to O_2 and pH).

RQ4: Group Compositions, Systems Thinking, and Discussion

Groups composed of different levels of prior knowledge might show different interaction patterns with the agents (Nguyen, 2021). Thus, I examined how discussion groups' variance (as indicated by the standard deviation in baseline understanding) might moderate the effects of agent conditions on learning. I found that for Mechanisms, the positive impact of the expert agent on posttest scores was larger in groups with higher deviations in pretest scores, compared to the no-agent group, $\beta = .44$, SE = .19, $t = 2.36$, $p = .02$, adjusted $p = .08$ (Table 5.8). The model with the interaction term had improved fit compared to the one without, $\chi 2(2) = 5.97$, $p = .05$. The coefficients can be interpreted as follows: The average student in the no-agent condition had a standardized Mechanisms posttest of -.58 (the intercept). Those who interacted with the expert agent at the average group variation had a standardized score of -.10 (intercept +

coefficient for “expert”). Every standardized unit increase in group variation would be associated with a .44 standardized unit increase in Mechanisms posttest in the expert agent condition.

To contextualize these findings, I examined the transactive exchange in three groups with high deviations in baseline understanding (i.e., group’s SD in pretest above the mean of all groups). For these groups, transactive exchange often occurred in episodes of three to five chat turns. Transactive episodes were predominately initiated by students who scored above average at pretest (21/29 episodes, 72%; 6/6 in the first group, 12/18 in the second, 3/5 in the third). Consider the following excerpt at the beginning of a group chat:

A.B.: zooplankton eats phytoplankton

A.B.: sea otters eat fish

M.H.: sea otters eat sea urchins

S.S.: that [sea otters eating sea urchins] helps kelp because sea urchins eat kelp

M.H.: and that is good because kelp helps photosynthesis

This group consisted of two individuals with emergent prior knowledge (A.B. and S.D.) and three individuals with expanding prior knowledge (M.H., S.S., E.P.). The excerpt illustrates how the group discussed concepts coherently, as all components were related to one another. Students S.S. and M.H. demonstrated transactive exchange, building on M.H.’s claim about sea otters and sea urchins to connect the discussion with kelp. Here, the transactive occurrences were driven by M.H. and S.S., who scored higher at baseline. As this discussion progressed, students with emergent baseline knowledge also participated in transactive exchange, although at lower frequencies: S.D. (transactive exchange counts = 2), A.B. (1), compared to M.H. (9), S.S. (10), E.P. (4). In comparison, the effect of the less-knowledgeable-peer agent on Mechanisms posttest and transactive exchange did not differ with groups’ variation in posttest scores. To understand this result, I examined the discussion patterns in the groups with high variation (above mean of all groups) in the less-knowledgeable-peer agent condition.

Table 5.8.

Interaction with Group Compositions for Mechanisms Posttests

	Components	Mechanisms
Fixed effect	Coeff. (SE)	Coeff. (SE)
Intercept	-.29 (.22)	-.58 (.24)*
Expert	.19 (.16)	.48 (.17)**
Peer	-.08 (.17)	.46 (.19)**
Group Variation	.08 (.12)	-.19 (.13)
Expert*Group Variation	.00 (.17)	.44 (.19)*
Peer*Group Variation	.18 (.18)	.35 (.20)
Random effect	Coeff. (SE)	Coeff. (SE)
Between-group variance	.01 (.01)	.01 (.01)
Within-group variance	.61 (.04)	.71 (.05)
f^2	.61	.39

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. The coefficients are standardized. Reference group = No Agent. For all models, covariates included individuals' pretest, gender, and classrooms' fixed effects

Table 5.9.

Descriptive Statistics for Systems Thinking and Transactive Exchange, by Pretest

	No agent		Expert		Less knowledgeable peer	
(Group)	Low Var.	High Var.	Low Var.	High Var.	Low Var.	High Var.
Components	1.91 (.73)	2.04 (1.09)	2.21 (.80)	1.94 (1.00)	1.77 (.81)	1.80 (.89)
Mechanisms	3.88 (1.45)	3.44 (1.74)	4.21 (1.49)	5.06 (1.80)	4.20 (2.04)	4.15 (2.13)
Transactive	2.66 (3.29)	2.30 (3.09)	3.33 (3.25)	4.22 (4.15)	3.89 (3.59)	4.50 (3.79)
(Individual)	Lower Avg.	Higher Avg.	Lower Avg.	Higher Avg.	Lower Avg.	Higher Avg.
Components	1.59 (.71)	2.41 (.93)**	1.88 (.70)	2.50 (.95)**	1.56 (.64)	2.33 (1.05)*
Mechanisms	3.19 (1.51)	4.26 (1.51)**	4.00 (1.22)	5.15 (1.85)**	3.64 (1.77)	5.60 (2.20)**
Transactive	1.83 (1.84)	3.22 (4.17)	2.70 (2.32)	4.41 (4.00)*	3.42 (2.74)	6.00 (5.08)

Note. * $p < .05$, ** $p < .01$. Comparisons of values within conditions. Grouping variables included Higher and Lower than average for group SD (Group), and higher and lower than average for individual pretest (Individual).

Similar to interactions with the expert agent, transactive exchange occurred in episodes. Different from the expert condition, however, it appeared that students with emergent prior knowledge were as likely to initiate transactive episodes. The majority of the episodes were initiated by individuals who scored below average for the pretest (29/46 episodes, 63%; 3/5 in the first group, 12/17 in the second, 4/9 in the third, 10/15 in the fourth). An explanation for the differences between the agent conditions is that the less knowledgeable peer condition might encourage more balanced initiative taking in transactive exchange.

A follow-up analysis that focused on transactive exchange provided evidence for the qualitative observations. I fitted a univariate multilevel model predicting transactive exchange and included an interaction between groups' standard deviation in pretest and agent conditions. Results indicate that the positive effect of the expert agent on transactive exchange was larger among groups with higher variation in Mechanisms pretests, $\beta = .44$, SE = .20, $t = 2.26$, $p = .03$, adjusted $p = .09$ (Model 3, Table 5.7). The model with the interaction term also had improved fit relative to the one without, $\chi 2(2) = 5.98$, $p = .05$. Comparisons of the average values for transactive exchange suggest that there were no differences in the number of transactive exchange based on individuals' pretest (higher or lower than average) for the less knowledgeable peer condition (Table 5.9). Meanwhile, students who scored higher than average at pretest produced significantly more transactive exchange than their counterparts in the expert condition, $M = 4.41$, $SD = 4.00$, compared to $M = 2.70$, $SD = 2.32$, $p = .03$.

Discussion

Agents Support Mechanistic Understanding

Articulating relationships within systems is challenging for learners, who tend to focus on separate components (Gellert, 1962; Hmelo-Silver, Marathe, et al., 2007; Wilensky & Resnick, 1999). In this work, I explore the potential of two conversational agent designs, an expert and a less knowledgeable peer, to support students to engage in productive discussion and deepen understanding of relationships among components.

The first research question investigates the effects of the agents on students' understanding of Components and Mechanisms. Results suggest that the agents can support more articulations of systems mechanisms at posttests. Both agents have a positive association with the number of systems mechanisms students articulated at posttest, compared to no-agent groups.

This finding shows that intentionally designed prompts on conceptual understanding might have a positive effect on systems understanding. The agents pose explicit prompts for students to develop hypotheses about systems relationships, for example, "What would happen to other components in the system if regulation increases?" Furthermore, the agents' effects on Mechanisms posttests do not vary with individual Mechanisms pretest. This finding indicates that the agents appear to support students of varying baseline understanding.

The second question explores how agents might impact students' discussion patterns. Researchers have suggested that agents might promote productive talk moves, such as idea articulation, reasoning, and transactive exchange (Dyke, Adamson, et al., 2013; Dyke, Howley, et al., 2013; Kumar et al., 2010). Findings from the current work provide evidence that students in groups with agents (particularly the less knowledgeable peer agent) produce more transactive exchange. The agents' prompts focus on productive talk moves, asking students to elaborate on what they just stated or to contribute to another student's contribution (Resnick et al., 2010). For example, the less knowledgeable peer phrases its prompts as "I didn't understand that. Can you explain to me what you mean?" These prompts focus on reasoning and elaboration, which might have encouraged students to engage with ideas with more complexity. Findings of the positive impacts of the agents on transactive exchange are in line with prior work with agents with different designs, for instance, posing as humans or chat boxes (Dyke, Adamson, et al., 2013; Rosé et al., 2015; Tegos & Demetriadis, 2017). Thus, findings suggest the robustness of agents' transactive prompts on students' exchange with or without human representations.

The mediation analysis provides further evidence for the path through which agents may support systems understanding. Both agents have a significant, indirect effect on Mechanisms' posttest scores through transactive exchange. This means that the increased transactive exchange in interaction with the agents can enhance the articulation of systems mechanisms. Students who engage in transactive exchange build on previous ideas in the discussions, thereby making their understanding more structured over time (Oliveira & Sadler, 2008; Teasley, 1997).

Results do not provide evidence that the experimental conditions differ in other discussion patterns, namely idea formation, elaborative and procedural questions, or off-task exchange, after accounting for students' baseline understanding, gender, and classes. These results are potentially due to the prompt design of the agents. The agents focus on idea elaboration and group participation, instead of questioning or building rapport with users through off-task exchange. Claim-making (without building on previous utterances) is the most common discussion pattern across all conditions. Students use claims to build out their concept maps, with or without agents' prompts. This finding aligns with previous observations of group discussions, where sharing ideas is a common discussion move (van Aalst, 2009; Yun & Kim, 2015).

In addition, the number of elaborative questions is small and does not differ significantly across conditions. This finding is different from my hypothesis that students would engage in more quizzing with the less knowledgeable peer agent and thus provide more elaborative questions in this agent condition. Elaborative questions can foster explanation and deeper reasoning towards knowledge advancement (Hogan et al., 1999). An explanation is that elaborative questioning is not an explicit focus of the agents' prompts. Other researchers have developed collaboration scripts that ask students to pose questions to peers within discussions, instead of just responding to the agents' prompts (Radkowitsch et al., 2020, 2021; Vogel et al., 2017; Walker et al., 2014). Such explicit prompts may be related to a higher number of elaborative questions among students.

Finally, I explore how frequencies of off-task utterances might differ across agent conditions. I hypothesize that the agents could gear students back to the learning task in case of irrelevant utterances and improve learning. This hypothesis is partially supported by the negative correlations between off-task utterances and Mechanisms posttest, $r(170) = -.18$, $p = .02$. Students who engage in fewer off-task utterances have higher Mechanisms posttest scores.

Prior work also suggests a different dynamic, where off-task utterances smoothen interactions with friends and meaningfully contribute to individuals' learning over time (Ogan, Finkelstein, Mayfield, et al., 2012; Ogan, Finkelstein, Walker, et al., 2012). In this study, students were randomly assigned to groups and not all students used their real names as usernames. As a result, the use of off-task exchange might be less salient than when students interact with known friends. Furthermore, the average number of off-task utterances was small and had a large deviation among individuals and between groups, suggesting that the use of off-task exchange was not universal ($M = .97$, $SD = 4.09$; ICC = .32). Follow-up research with larger sample sizes and considerations of friendship in randomization (e.g., friends versus unknown individuals) can examine the role of off-task exchange as hindering or supporting discussions.

The Effect of Agent on Transactive Exchange Varied with Group Baseline Understanding

To understand the types of grouping arrangement that best support learning, I examine the interactions between agent conditions and groups' baseline understanding of systems. Findings provide suggestions for future work. Among groups who interact with the expert agent, higher group variations at baseline understanding (i.e., a mix of expanding and emergent prior knowledge) have a larger effect on transactive exchange and Mechanisms posttest scores, compared to groups with lower variation in the same condition.

A qualitative examination of the data suggests that students with expanding prior knowledge tend to initiate transactive exchange more frequently in the expert agent condition, compared to the less knowledgeable peer condition. These students might be more willing to initiate exchange and structure their own knowledge (Gillies, 2003; Peterson et al., 1981). Such interactions resemble exchange in teacher-guided queries, where students with higher prior knowledge tend to dominate the discussions (Cohen, 1994; Yun & Kim, 2015).

In comparison, group's variation in prior knowledge is not a significant moderator of the relationship between the less knowledgeable peer agent and Mechanisms posttest. This agent potentially invites more balanced participation. Contributions to the discussions may be affected by students' sense of how they are perceived by others. Whereas students with high rates of contribution to the discussions may associate their agency to contribute with "learning science", those with lower rates of contribution tend to be more concerned with "being right" (Clarke et al., 2016). These findings expand observations from prior work (Tegos & Demetriadis, 2017), which has focused on the link between agent usage, transactive exchange, and learning outcomes, but has not considered the impact of group compositions.

Overall, findings suggest the nuances in interactions with the agent designs and call for considerations of factors at the individual and group levels to promote equitable discussions. The posttest scores for systems thinking do not significantly differ between the two agent conditions. However, the agent conditions likely support different participation in transactive exchange in heterogeneous groups. Findings have practical implications for group arrangement and framing of learning tasks. In interactions with the expert agent, for instance, heterogeneous grouping might be beneficial. Students with expanding prior knowledge may initiate exchange and responsiveness with the agent, which are taken up by other members of the group. In comparison, the less knowledgeable peer agent shows promise in engaging learners at different baseline understanding to initiate transactive exchange. These nuances highlight the potential of task framing for teacher facilitators, such as "learning together" with a less knowledgeable peer, versus "show me what you know" with an expert.

Limitations

Findings from this study should be considered in light of its limitations. First, the study's sample size did not have enough power to detect small effects. Furthermore, the sample was limited to students from specific socioeconomic and demographic backgrounds within the same school. This sample size might have limited the ability to detect varying effects among subgroups, such as individuals' baseline understanding. Thus, future work that replicates the study designs with a larger sample and different populations will provide understanding about how agents might impact students differentially as a result of their baseline knowledge. Additionally, larger sample sizes will also accommodate methodologies such as multilevel structural equation modeling, which allow for simultaneously testing multiple paths between agent conditions, discussion patterns, and learning outcomes.

In addition, the models in this study included students' gender as a covariate, but I did not find significant differences in posttest systems thinking or most students' discussion patterns (except for off-task talks; Appendix A5.2). Thus, follow-up research can examine additional demographic variables (see work by Kim & Baylor, 2016) and the interaction of demographics variables at the group levels to examine additional dynamics in agent interactions.

Furthermore, the analyses with grouping variables were exploratory in nature. The potential differences in how heterogeneous groups may support different participation patterns between agent conditions open up questions for future research. For example, follow-up interviews with a subgroup of students may reveal insights into how marginalized students, such as those who do not identify with science or do not traditionally participate, are leading the discussions and participating in the conversations. Internalized perceptions of science participation, including questions about who can take part in the activities, might influence long-term identification with the discipline (Archer et al., 2017). Thus, follow-up insights are valuable in facilitating the design of learning tasks and technology to promote equitable participation.

Finally, the interactions with the agents were embedded in a single lesson that did not require intensive coordination with teachers. Emergent work has highlighted the importance of agent-teacher and agent-student collaboration in intelligent systems (Holstein et al., 2020). I intend to pursue this area in future work and examine how agents can be embedded into activities throughout the curriculum and provide additional insights to teachers, as well as incorporate teachers' feedback to support students' collaboration.

Conclusion

In this work, I explore the pathways in which different conversational agents (a less knowledgeable peer or an expert) can facilitate discussion and learning of complex systems, relative to the no-agent condition. Results highlight the potential of the two agents to deepen students' mechanisms understanding through transactive exchange. These findings illuminate the application of agents in collaborative contexts, to encourage students to use reasoning to build on others' ideas and foster complex systems thinking. They also call for future work on how agents' impacts on learning might vary with group compositions. Findings highlight the nuances of framing designs and discussion tasks. They call for considerations of research in human-centered interaction and collaborative discourse to identify other discussion moves that the conversational agents can promote (beyond the expert-peer paradigm), design these agents, and study how the agents may influence group dynamics and learning for different individuals and group arrangements.

References

Adamson, D., Dyke, G., Jang, H., & Rosé, C. P. (2014). Towards an agile approach to adapting dynamic collaboration support to student needs. *International Journal of Artificial Intelligence in Education*, *24*(1), 92–124. https://doi.org/10.1007/s40593-013-0012-6

Aguiar, O. G., Mortimer, E. F., & Scott, P. (2010). Learning from and responding to students' questions: The authoritative and dialogic tension. *Journal of Research in Science Teaching*, *47*(2), 174–193. https://doi.org/10.1002/TEA.20315

Archer, L., Dawson, E., DeWitt, J., Godec, S., King, H., Mau, A., Nomikou, E., & Seakins, A. (2017). Killing curiosity? An analysis of celebrated identity performances among teachers and students in nine London secondary science classrooms. *Science Education*, *101*(5), 741–764. https://doi.org/10.1002/SCE.21291

Biswas, G., Segedy, J. R., & Bunchongchit, K. (2016). From Design to Implementation to Practice a Learning by Teaching System: Betty's Brain. *International Journal of Artificial Intelligence in Education*, *26*(1), 350–364. https://doi.org/10.1007/s40593-015-0057-9

Chin, C., & Osborne, J. (2010). Students' questions and discursive interaction: Their impact on argumentation during collaborative group discussions in science. *Journal of Research in Science Teaching*, *47*(7), 883–908. https://doi.org/10.1002/TEA.20385

Clarke, S. N., Howley, I., Resnick, L., & Penstein Rosé, C. (2016). Student agency to participate in dialogic science discussions. *Learning, Culture and Social Interaction*, *10*, 27–39. https://doi.org/10.1016/J.LCSI.2016.01.002

Cohen, E. G. (1994). Restructuring the Classroom: Conditions for Productive Small Groups: *Review of Educational Research*, *64*(1), 1–35. https://doi.org/10.3102/00346543064001001

Danish, J., Saleh, A., Andrade, A., & Bryan, B. (2017). Observing complex systems thinking in the zone of proximal development. *Instructional Science*, *45*(1), 5–24. https://doi.org/10.1007/S11251-016-9391-Z/TABLES/9

Dong, N., & Maynard, R. (2013). *PowerUp!* : A Tool for Calculating Minimum Detectable Effect Sizes and Minimum Required Sample Sizes for Experimental and Quasi-Experimental Design Studies. *Journal of Research on Educational Effectiveness*, *6*(1), 24–67. https://doi.org/10.1080/19345747.2012.673143

Dyke, G., Adamson, D., Howley, I., & Rose, C. P. (2013). Enhancing scientific reasoning and discussion with conversational agents. *IEEE Transactions on Learning Technologies*, *6*(3), 240–247. https://doi.org/10.1109/TLT.2013.25

Dyke, G., Howley, I., Adamson, D., Kumar, R., & Rosé, C. P. (2013). Towards Academically Productive Talk Supported by Conversational Agents. In *Productive Multivocality in the Analysis of Group Interactions* (pp. 459–476). Springer US. https://doi.org/10.1007/978-1-4614-8960-3_25

Fiacco, J., & Rosé, C. (2018). Towards domain general detection of transactive knowledge building behavior. *Proceedings of the Fifth Annual ACM Conference on Learning at Scale*, 1–11.

Fu, E. L. F., van Aalst, J., & Chan, C. K. K. (2016). Toward a classification of discourse patterns in asynchronous online discussions. *International Journal of Computer-Supported Collaborative Learning*, *11*(4), 441–478. https://doi.org/10.1007/S11412-016-9245-3/TABLES/5

Gellert, E. (1962). Children's conceptions of the content and functions of the human body. *Genetic Psychology Monographs*, *65*, 293–405.

Gillies, R. M. (2003). Structuring cooperative group work in classrooms. *International Journal of Educational Research*, *39*(1–2), 35–49. https://doi.org/10.1016/S0883-0355(03)00072-7

Gillies, R. M. (2016). Dialogic interactions in the cooperative classroom. *International Journal of Educational Research*, *76*, 178–189. https://doi.org/10.1016/J.IJER.2015.02.009

Graesser, A. C. (2016). Conversations with AutoTutor Help Students Learn. *International Journal of Artificial Intelligence in Education*, *26*(1), 124–132. https://doi.org/10.1007/s40593-015-0086-4

Grilli, L., Pennoni, F., Rampichini, C., & Romeo, I. (2016). Exploiting TIMSS and PIRLS combined data: Multivariate multilevel modelling of student achievement. *The Annals of Applied Statistics*. https://doi.org/10.1214/16-AOAS988

Harley, J., Taub, M., Azevedo, R., & Bouchet, F. (2017). Let's set up some subgoals: Understanding human-pedagogical agent collaborations and their implications for learning and prompt and feedback compliance. *IEEE Transactions on Learning Technologies*, *11*(1), 54–66.

Harley, J., Taub, M., Bouchet, F., & Azevedo, R. (2012). A framework to understand the nature of co-regulated learning in human-pedagogical agent interactions. *11th International Conference on Intelligent Tutoring Systems, Crete*.

Hmelo-Silver, C. E., & Barrows, H. S. (2008). Facilitating Collaborative Knowledge Building. *Cognition and Instruction*, *26*(1), 48–94. https://doi.org/10.1080/07370000701798495

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). In *Educational Psychologist* (Vol. 42, Issue 2, pp. 99–107). Routledge. https://doi.org/10.1080/00461520701263368

Hmelo-Silver, C. E., Jordan, R., Eberbach, C., & Sinha, S. (2017). Systems learning with a conceptual representation: a quasi-experimental study. *Instructional Science*, *45*(1), 53–72. https://doi.org/10.1007/S11251-016-9392-Y/FIGURES/5

Hmelo-Silver, C. E., Marathe, S., & Liu, L. (2007). Fish Swim, Rocks Sit, and Lungs Breathe: Expert-Novice Understanding of Complex Systems. *Http://Dx.Doi.Org/10.1080/10508400701413401*, *16*(3), 307–331. https://doi.org/10.1080/10508400701413401

Hogan, K., Nastasi, B. K., & Pressley, M. (1999). Discourse Patterns and Collaborative Scientific Reasoning in Peer and Teacher-Guided Discussions. *Cognition and Instruction*, *17*(4), 379–432. https://doi.org/10.1207/S1532690XCI1704_2

Holstein, K., Aleven, V., & Rummel, N. (2020). A conceptual framework for human–AI hybrid adaptivity in education. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12163 LNAI*, 240–254. https://doi.org/10.1007/978-3-030-52237-7_20

Howley, I., Kumar, R., Mayfield, E., Dyke, G., & Rosé, C. P. (2013). Gaining Insights from Sociolinguistic Style Analysis for Redesign of Conversational Agent Based Support for

Collaborative Learning. In *Productive Multivocality in the Analysis of Group Interactions* (pp. 477–494). Springer US. https://doi.org/10.1007/978-1-4614-8960-3_26

Jacobson, M. J., & Wilensky, U. (2006). Complex Systems in Education: Scientific and Educational Importance and Implications for the Learning Sciences. *Http://Dx.Doi.Org/10.1207/S15327809jls1501_4*, *15*(1), 11–34. https://doi.org/10.1207/S15327809JLS1501_4

Jordan, R. C., Hmelo-Silver, C., Liu, L., & Gray, S. A. (2013). Fostering reasoning about complex systems: using the aquarium to teach systems thinking. *Applied Environmental Education & Communication*, *12*(1), 55–64.

Kim, S., Eun, J., Oh, C., Suh, B., & Lee, J. (2020). Bot in the Bunch: Facilitating Group Chat Discussion by Improving Efficiency and Participation with a Chatbot. *Conference on Human Factors in Computing Systems - Proceedings*, 1–13. https://doi.org/10.1145/3313831.3376785

Kim, Y., & Baylor, A. L. (2006). A social-cognitive framework for pedagogical agents as learning companions. *Educational Technology Research and Development*, *54*(6), 569–596. https://doi.org/10.1007/s11423-006-0637-3

Kim, Y., & Baylor, A. L. (2016). Research-Based Design of Pedagogical Agent Roles: A Review, Progress, and Recommendations. *International Journal of Artificial Intelligence in Education*, *26*(1), 160–169. https://doi.org/10.1007/s40593-015-0055-y

King, A. (1999). Discourse patterns for mediating peer learning. In *Cognitive perspectives on peer learning.* (pp. 87–115). Lawrence Erlbaum Associates Publishers.

Kinnebrew, J. S., Segedy, J. R., & Biswas, G. (2014). Analyzing the temporal evolution of students' behaviors in open-ended learning environments. *Metacognition and Learning*, *9*(2), 187–215. https://doi.org/10.1007/S11409-014-9112-4

Kiwanuka, H. N., Van Damme, J., Van Den Noortgate, W., Anumendem, D. N., Vanlaar, G., Reynolds, C., & Namusisi, S. (2016). How do student and classroom characteristics affect attitude toward mathematics? A multivariate multilevel analysis. *Http://Dx.Doi.Org/10.1080/09243453.2016.1201123*, *28*(1), 1–21. https://doi.org/10.1080/09243453.2016.1201123

Kumar, R., Ai, H., Beuth, J. L., & Rosé, C. P. (2010). Socially capable conversational tutors can be effective in collaborative learning situations. *Lecture Notes in Computer Science*, *6094 LNCS*(PART 1), 156–164. https://doi.org/10.1007/978-3-642-13388-6_20

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Package 'lmertest.' *R Package Version*, *2*(0), 734.

Leach, C. W., van Zomeren, M., Zebel, S., Vliek, M. L. W., Pennekamp, S. F., Doosje, B., Ouwerkerk, J. W., & Spears, R. (2008). Group-Level Self-Definition and Self-Investment: A Hierarchical (Multicomponent) Model of In-Group Identification. *Journal of Personality and Social Psychology*, *95*(1), 144–165. https://doi.org/10.1037/0022-3514.95.1.144

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2019). *Package 'emmeans.'*

Liew, T. W., Tan, S.-M., & Jayothisa, C. (2013). The Effects of Peer-Like and Expert-Like Pedagogical Agents on Learners' Agent Perceptions, Task-Related Attitudes, and Learning Achievement. *Educational Technology & Society*, *16*(4), 275–286. https://eric.ed.gov/?id=EJ1013641

Liu, L., & Hmelo-Silver, C. E. (2009). Promoting complex systems learning through the use of conceptual representations in hypermedia. *Journal of Research in Science Teaching*, *46*(9), 1023–1040. https://doi.org/10.1002/TEA.20297

Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B., & D'Apollonia, S. (1996). Within-Class Grouping: A Meta-Analysis: *Http://Dx.Doi.Org/10.3102/00346543066004423*, *66*(4), 423–458. https://doi.org/10.3102/00346543066004423

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, *49*(4), 1494–1502. https://doi.org/10.3758/S13428-016-0809-Y/TABLES/1

Mäkitalo, K., Weinberger, A., Häkkinen, P., Järvelä, S., & Fischer, F. (2005). Epistemic cooperation scripts in online learning environments: fostering learning by reducing uncertainty in discourse? *Computers in Human Behavior*, *21*(4), 603–622. https://doi.org/10.1016/J.CHB.2004.10.033

Michaels, S., O'Connor, C., & Resnick, L. B. (2008). Deliberative Discourse Idealized and Realized: Accountable Talk in the Classroom and in Civic Life. *Stud Philos Educ*, *27*, 283–297. https://doi.org/10.1007/s11217-007-9071-1

Moon, Y., & Nass, C. I. (1996). Adaptive agents and personality change. *Conference Companion on Human Factors in Computing Systems Common Ground - CHI '96*, 287–288. https://doi.org/10.1145/257089.257325

Nass, C., & Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues*, *56*(1), 81–103. https://doi.org/10.1111/0022-4537.00153

Nass, C., Moon, Y., Fogg, B. J., Reeves, B., & Dryer, D. C. (1995). Can computer personalities be human personalities? *International Journal of Human-Computer Studies*, *43*(2), 223–239. https://doi.org/10.1006/ijhc.1995.1042

Nguyen, H. (2021). Exploring Group Discussion with Conversational Agents Using Epistemic Network Analysis. *Communications in Computer and Information Science*, *1522 CCIS*, 378–394. https://doi.org/10.1007/978-3-030-93859-8_25

Nguyen, H., & Santagata, R. (2020). Impact of computer modeling on learning and teaching systems thinking. *Journal of Research in Science Teaching*, tea.21674. https://doi.org/10.1002/tea.21674

Ogan, A., Aleven, V., Jones, C., & Kim, J. (2011). Persistent Effects of Social Instructional Dialog in a Virtual Learning Environment. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *6738 LNAI*, 238–246. https://doi.org/10.1007/978-3-642-21869-9_32

Ogan, A., Finkelstein, S., Mayfield, E., D'Adamo, C., Matsuda, N., & Cassell, J. (2012). "Oh, dear Stacy!" Social interaction, elaboration, and learning with teachable agents. *Conference on Human Factors in Computing Systems - Proceedings*, 39–48. https://doi.org/10.1145/2207676.2207684

Ogan, A., Finkelstein, S., Walker, E., Carlson, R., & Cassell, J. (2012). Rudeness and Rapport: Insults and Learning Gains in Peer Tutoring. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7315 LNCS*, 11–21. https://doi.org/10.1007/978-3-642-30950-2_2

Oliveira, A. W., & Sadler, T. D. (2008). Interactive patterns and conceptual convergence during student collaborations in science. *Journal of Research in Science Teaching*, *45*(5), 634–658. https://doi.org/10.1002/TEA.20211

Paus, E., Werner, C. S., & Jucks, R. (2012). Learning through online peer discourse: Structural equation modeling points to the role of discourse activities in individual understanding. *Computers & Education*, *58*(4), 1127–1137. https://doi.org/10.1016/J.COMPEDU.2011.12.008
Peterson, P. L., Janicki, T. C., & Swing, S. R. (1981). Ability × Treatment Interaction Effects on Children's Learning in Large-group and Small-group Approaches: *Http://Dx.Doi.Org/10.3102/00028312018004453*, *18*(4), 453–473. https://doi.org/10.3102/00028312018004453
Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., Heisterkamp, S., Van Willigen, B., & Maintainer, R. (2017). Package 'nlme.' *Linear and Nonlinear Mixed Effects Models, Version*, *3*(1).
Radkowitsch, A., Sailer, M., Schmidmaier, R., Fischer, M. R., & Fischer, F. (2021). Learning to diagnose collaboratively – Effects of adaptive collaboration scripts in agent-based medical simulations. *Learning and Instruction*, *75*, 101487. https://doi.org/10.1016/J.LEARNINSTRUC.2021.101487
Radkowitsch, A., Vogel, F., & Fischer, F. (2020). Good for learning, bad for motivation? A meta-analysis on the effects of computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning*. https://doi.org/10.1007/s11412-020-09316-4
Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). sage.
Resnick, L. B., Michaels, S., & O'Connor, C. (2010). How (well structured) talk builds the mind. *Innovations in Educational Psychology: Perspectives on Learning, Teaching and Human Development*, 163–194.
Roschelle, J., & Teasley, S. D. (1995). The Construction of Shared Knowledge in Collaborative Problem Solving. In *Computer Supported Collaborative Learning* (pp. 69–97). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-85098-1_5
Rosé, C. P., Goldman, P., Zoltners Sherer, J., & Resnick, L. (2015). Supportive technologies for group discussion in MOOCs. *Current Issues in Emerging ELearning*, *2*(1), 5.
Rosenberg-Kima, R. B., Baylor, A. L., Plant, E. A., & Doerr, C. E. (2008). Interface agents as social models for female students: The effects of agent visual presence and appearance on female students' attitudes and beliefs. *Computers in Human Behavior*, *24*(6), 2741–2756. https://doi.org/10.1016/j.chb.2008.03.017
Saleh, M., Lazonder, A. W., & Jong, T. de. (2007). Structuring collaboration in mixed-ability groups to promote verbal interaction, learning, and motivation of average-ability students. *Contemporary Educational Psychology*, *32*(3), 314–331. https://doi.org/10.1016/j.cedpsych.2006.05.001
Scardamalia, M., & Bereiter, C. (1993). Technologies for knowledge-building discourse. *Communications of the ACM*, *36*(5), 37–41. https://doi.org/10.1145/155049.155056
Selya, A. S., Rose, J. S., Dierker, L. C., Hedeker, D., & Mermelstein, R. J. (2012). A practical guide to calculating Cohen's f 2, a measure of local effect size, from PROC MIXED. *Frontiers in Psychology*, *3*(APR), 111. https://doi.org/10.3389/FPSYG.2012.00111/BIBTEX
Siddiqui, O., Hedeker, D., Flay, B. R., & Hu, F. B. (1996). Intraclass Correlation Estimates in a School-based Smoking Prevention StudyOutcome and Mediating Variables, by Sex and Ethnicity. *American Journal of Epidemiology*, *144*(4), 425–433. https://doi.org/10.1093/OXFORDJOURNALS.AJE.A008945

Snapir, Z., Eberbach, C., Ben-Zvi-Assaraf, O., Hmelo-Silver, C., & Tripto, J. (2017). Characterising the development of the understanding of human body systems in high-school biology students – a longitudinal study. *Http://Dx.Doi.Org/10.1080/09500693.2017.1364445*, *39*(15), 2092–2127. https://doi.org/10.1080/09500693.2017.1364445
Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Sage Publications.
Teasley, S. D. (1997). Talking About Reasoning: How Important Is the Peer in Peer Collaboration? In *Discourse, Tools and Reasoning* (pp. 361–384). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-03362-3_16
Tegos, S., & Demetriadis, S. (2017). Conversational Agents Improve Peer Learning through Building on Prior Knowledge. *Educational Technology & Society*, *20*(1), 99–111.
Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). *Mediation: R package for causal mediation analysis*.
van Aalst, J. (2009). Distinguishing knowledge-sharing, knowledge-construction, and knowledge-creation discourses. *International Journal of Computer-Supported Collaborative Learning*, *4*(3), 259–287. https://doi.org/10.1007/s11412-009-9069-5
Vogel, F., Wecker, C., Kollar, I., & Fischer, F. (2017). Socio-Cognitive Scaffolding with Computer-Supported Collaboration Scripts: a Meta-Analysis. *Educational Psychology Review*, *29*(3), 477–511. https://doi.org/10.1007/S10648-016-9361-7/TABLES/4
Walker, E., Rummel, N., & Koedinger, K. R. (2014). Adaptive intelligent support to improve peer tutoring in algebra. *International Journal of Artificial Intelligence in Education*, *24*(1), 33–61. https://doi.org/10.1007/s40593-013-0001-9
Webb, N. M. (1982). Student Interaction and Learning in Small Groups: *Review of Educational Research*, *52*(3), 421–445. https://doi.org/10.3102/00346543052003421
Webb, N. M. (1989). Peer interaction and learning in small groups. *International Journal of Educational Research*, *13*(1), 21–39.
Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers and Education*, *46*(1), 71–95. https://doi.org/10.1016/j.compedu.2005.04.003
Wilensky, U., & Resnick, M. (1999). Thinking in Levels: A Dynamic Systems Approach to Making Sense of the World. *Journal of Science Education and Technology 1999 8:1*, *8*(1), 3–19. https://doi.org/10.1023/A:1009421303064
Wittrock, M. C. (1989). Generative processes of comprehension. *Educational Psychologist*, *24*(4), 345–376.
Yoon, S. A. (2007). An Evolutionary Approach to Harnessing Complex Systems Thinking in the Science and Technology Classroom. *Http://Dx.Doi.Org/10.1080/09500690601101672*, *30*(1), 1–32. https://doi.org/10.1080/09500690601101672
Yoon, S. A., Anderson, E., Klopfer, E., Koehler-Yom, J., Sheldon, J., Schoenfeld, I., Wendel, D., Scheintaub, H., Oztok, M., & Evans, C. (2016). Designing computer-supported complex systems curricula for the Next Generation Science Standards in high school science classrooms. *Systems*, *4*(4), 38.
Yun, S. M., & Kim, H. B. (2015). Changes in Students' Participation and Small Group Norms in Scientific Argumentation. *Research in Science Education*, *45*(3), 465–484. https://doi.org/10.1007/S11165-014-9432-Z/TABLES/6

Appendix

Appendix A5.1

Correlation of Discussion Patterns Codes and Subcodes

	I	EQ	PQ	CQ	TQ	Tr	Ex	T	OT	M	SD	Range
Ideation (I)	1									10.72	7.97	0, 42
Elaborative (EQ)	-.04	1								.22	.68	0, 5
Procedural (PQ)	.06	.30**	1							.35	.84	0, 5
Confirmatory (CQ)	.01	.26**	**.90*** **	1						.25	.75	0, 5
Task ques. (TQ)	.15*	.14	**.45*** **	.30**	1					.10	.36	0, 2
Transactive (Tr)	.48***	.11	.05	.02	.06	1				3.37	3.50	0, 15
Externalize (E)	.36***	.04	.00	-.01	.01	**.84*** **	1			1.99	2.02	0, 8
Build on friends (T)	.39***	.15	.08	.05	.10	**.80*** **	.44***	1		1.39	2.25	0, 11
Off-task (OT)	.01	.16*	.12	.01	.27**	-.08	-.02	-.12	1			

Note. Counts of claim and conjecture are summed because there were few instances of conjectures, M = .06, SD = .32, range 0-3.
* $p < .05$, ** $p < .01$, *** $p < .001$.

Appendix A5.2

Multilevel Models Predicting Systems Thinking and Discussion Patterns, Full Results

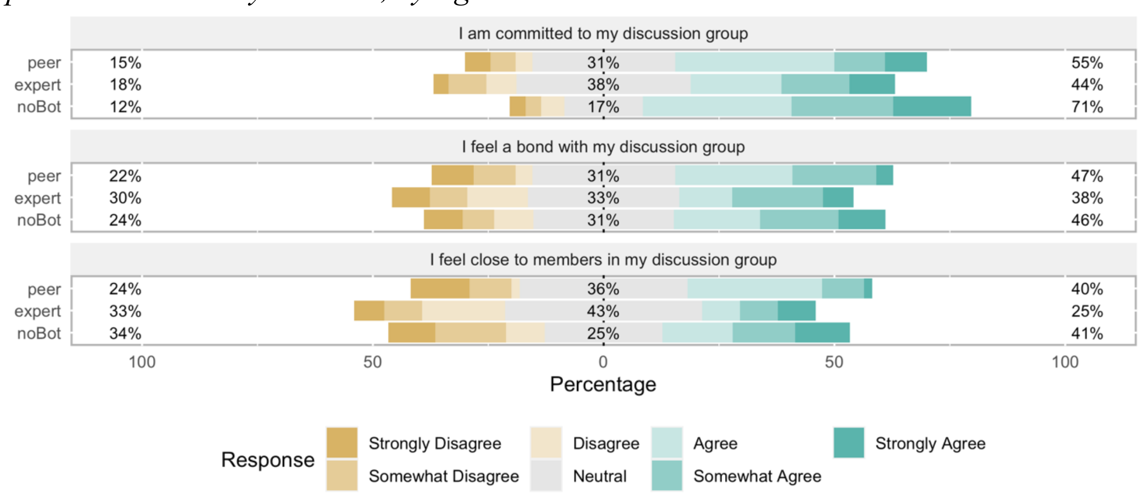
	Systems Thinking		Discussion Patterns				
	Components	Mechanisms	Idea	Elaborative Q.	Procedural Q.	Transactive	Off-task
	Coeff. (SE)	Coeff. (SE)	Coeff. (SE)	Coeff. (SE)	Coeff. (SE)	Coeff. (SE)	Coeff. (SE)
Intercept	-.25 (.23)	-.53 (.25)*	-1.30 (.32)***	-.09 (.32)	.01 (.29)	-.73 (.25)**	.18 (.31)
Expert	.18 (.17)	.45 (.18)*	-.09 (.18)	.01 (.25)	-.37 (.22)	.35 (.18)*	-.41 (.25)
Peer	-.06 (.18)	.49 (.20)*	.11 (.19)	.22 (.27)	-.37 (.24)	.56 (.20)**	-.30 (.27)
Pretest (C)	.45 (.10)***	.15 (.11)	.33 (.14)*	-.09 (.12)	.10 (.12)	.08 (.11)	.03 (.10)
Pretest (M)	.22 (.10)*	.35 (.11)**	.02 (.07)	.01 (.11)	-.18 (.12)	.28 (.12)*.	-.07 (.09)
Gender	-.003 (.13)	.09 (.14)	-.06 (.15)	-.18 (.15)	.005 (.16)	-.05 (.15)	-.37 (.13)**
Period 2	.41 (.24)	.22 (.26)	.41 (.25)	.61 (.35)	.36 (.31)	.29 (.26)	.67 (.35)
Period 3	-.05 (.26)	-.06 (.29)	.71 (.28)*	.42 (.37)	.30 (.34)	.49 (.29)	.30 (.36)
Period 4	.15 (.26)	.29 (.29)	1.24 (.28)***	-.16 (.36)	.23 (.33)	.80 (.28)**	-.01 (.36)
Period 5	.14 (.26)	.30 (.29)	.95 (.29)**	-.13 (.37)	.07 (.34)	.69 (.29)*	.13 (.36)
Period 6	.26 (.29)	.29 (.36)	.59 (.36)	-.29 (.45)	.39 (.42)	.54 (.36)	.04 (.43)
Between-group var.	.03 (.01)	.03 (.01)	.01 (.01)	.17 (.02)	.08 (.02)	.01 (.01)	.23 (.04)
Within-group var.	.63 (.04)	.78 (.05)	.88 (.07)	.83 (.05)	.94 (.05)	.90 (.05)	.58 (.04)

Note. * $p < .05$, ** $p < .01$, *** $p < .001$.

Appendix A5.3
Exploratory Analyses with Group Solidarity

Solidarity reflects psychological bonds with the group and can influence how individuals approach group activities (Ellemers et al., 1999). Following the discussion task, students filled in a survey on group solidarity, to reflect their commitment to fellow group members. The survey items were adapted from Leach et al.'s (2008) measures of in-group identification. To assess solidarity, students answered three items on a 7-point Likert scale (strongly disagree to strongly agree). An example item is "I feel a bond with my discussion group". I calculated the average of the items, with higher scores indicating higher group investment. The Cronbach's alpha coefficient is .86 (95% CI: .83, .90), suggesting high internal consistency of the survey items. Figure A5.4 shows the distribution of the survey answers. On average, there was no significant difference in the scores for solidarity across conditions (no agent: M = 4.5, SD = 1.46, expert: M = 4.19, SD = 1.37, peer: M = 4.24, SD = 1.37, Kruskal Wallis' test p = .41).

Figure A5.4
Group Investment Survey Answers, by Agent Conditions



Interaction with Group Solidarity.

I examined whether the relationship between agent conditions and transactive exchange varied with how students perceived their solidarity with the discussion groups. Perceived solidarity can impact how individuals engage in group tasks (Leach et al., 2008; Ogan et al., 2011). I fitted a multilevel model with transactive exchange as the outcome variable and included an interaction term between individuals' perceived solidarity and the agent conditions (Model 2, Table A5.5). Covariates included students' gender, pretest scores, and classes.

The model with the interaction term had improved model fit compared to the model without one, $\chi 2(2) = 6.85$, $p = .03$. The coefficient for the interaction term suggests that the positive effect of the less-knowledgeable-peer condition and transactive exchange was higher for individuals who perceived closer bonds with their groups, compared to those with lower solidarity in the same condition, $\beta = .43$, $SE = .18$, $t = 2.31$, $p = .02$. An increase in perceived bonds with the groups was associated with a higher number of elaborations on discussion ideas in the less knowledgeable peer agent condition.

Table A5.5

Interaction Effects with Individual Pretest Scores on Transactive Exchange

	Model 1	**Model 2**
	Coeff. (SE)	Coeff. (SE)
Intercept	-.73 (.25)**	-.67 (.25)**
Expert	.29 (.18)	.29 (.18)
Less knowledgeable peer (Peer)	.56 (.20)**	.62 (.20)**
Pretest (Mechanisms)	.28 (.12)*	.26 (.12)*
Expert x Pretest		
Peer x Pretest		
Solidarity		-.08 (.12)
Expert x Solidarity		.09 (.18)
Peer x Solidarity		.43 (.18)*
Between-group Var	.01 (.01)	.01 (.01)
Within-group Var	.90 (.05)	.88 (.05)

Note. $^{+}p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$. The coefficients are standardized. Reference group = No Agent. Covariates included individuals' gender, classrooms' fixed effects, and Components and Mechanisms pretest scores. Models accounted for discussion groups' random effects.

Mediation Analyses: Interactions with Group Solidarity

Based on findings that the effect of the agent conditions on transactive exchange varied with group solidarity for the less knowledgeable agent, I fitted an additional model to test the significance of the indirect effects of transactive exchange on Mechanisms understanding at differing levels of group solidarity (Table A5.6). Group solidarity was the moderator on the effect of agent conditions on Mechanisms, through transactive exchange. I compared the estimates for individuals high in group solidarity (1 SD above mean in solidarity) and low in solidarity (1 SD below mean). I did not find significant differences in the direct and indirect effects of agent conditions on Mechanisms posttests along high and low solidarity thresholds.

Table A5.6

Moderated Mediation Model between Conditions and Mechanisms

	Expert			Peer		
	High solidarity	Low solidarity	*p*	High solidarity	Low solidarity	*p*
Indirect effect (Agent to Mechanisms through Transactive)	.06 [-.02, .19]	.03 [-.05, .14]	.72	.17 [.01, .38]*	.03 [-.06, .14]	.12
Direct effect (Agent to Mechanisms)	.31 [-.15, .77]	.53 [.09, 1]*	.58	.39 [-.11, .94]	.48 [.02, .97]*	.95

Discussion & Future Directions

In sum, the exploratory analyses with group solidarity reveal potential differentiating effects of the less knowledgeable peer agent on transactive exchange based on a sense of solidarity. It is possible that the design of this agent heightened students' sense of solidarity and subsequently enhanced knowledge construction efforts. This pattern is similar to peer tutoring interactions where increased rapport generates more productive learning exchange (Ogan et al., 2011). Follow-up analyses, including interviews with individuals at high and low solidarity thresholds, may tease apart how individuals perceive their psychological bonds with the groups in each agent condition in more detail. Such analyses can also examine factors related to the learning task and the agent designs that affect these perceptions, and how perceptions of solidarity may influence individuals' perceived efforts in group discussions.

Appendix A5.4
Multivariate Multigroup Analyses
Method Overview

Multivariate analyses model multiple outcomes simultaneously and can be beneficial when outcomes are correlated. In addition, this approach allows for examining whether the effect of agents is stronger on some outcomes (e.g., Components) than others (e.g., Mechanisms). Researchers have applied multivariate multilevel models in education settings, for example, when students are nested within schools (Grilli et al., 2016; Kiwanuka et al., 2016; Snijders & Bosker, 2012). The fitted models followed the procedure for multivariate multilevel models outlined in Snijders and Bosker (2012). The values of the post-test Components and Mechanisms were stacked into a single outcome variable (y_h). An indicator variable (i.e., dummy-coded, d_h as 1 or 0) was added to indicate whether the data referred to Components or Mechanisms. The model equation takes the form:

$$y_{hij} = \beta_0 d_{hij} + \beta_1 \text{Agent\_Condition} d_{hij} + \beta_k X_{kij} d_{hij} + U_j d_{hij} + R_{ij} d_{hij} \quad (2)$$

In equation (2), the subscripts *h* indicates the index of the dependent variable (Components or Mechanisms), *i* refers to individual students, and *j* refers to the discussion groups. X is a vector of covariates (e.g., gender, pretest scores for Components and Mechanisms, class). U captures the random effect of the discussion groups, and R denotes the residual errors (within-person). Additionally, I fitted a model with interaction terms between individuals' pretest Mechanisms scores and the agent conditions. This model examined whether the agent had differential effects depending on baseline mechanistic understanding. All continuous variables were standardized.

Results

To evaluate the effect of agent conditions, I fitted a multivariate multilevel model to predict the posttest scores in Components and Mechanisms. The correlation coefficients were .66

at the group level and .60 at the student level. This particularly shows that the random group effects for Components and Mechanisms were substantially correlated, pointing at contextual factors at the group levels that might help explain both variables.

Overall, the magnitude and size of the estimates were similar between the univariate and multivariate models. There were positive effects of the agent conditions for Mechanisms posttest for both the expert (β = .49, *SE* = .17, *t* = 2.96, *p* = .003) and less-knowledgeable-peer conditions (β = .55, *SE* = .17, *t* = 3.14, *p* = .002). Follow-up pairwise comparisons of the marginal means for Mechanisms did not provide evidence for differences between the two agents; expert: *M* = .16, *SE* = .12, peer: *M* = .21, *SE* = .13, *p* = .99. There was no significant difference in Components posttests for agent conditions, relative to the no-agent condition (expert: β = .19, *SE* = .16, *t* = 1.14, *p* = .25, peer: β = .09, *SE* = .17, *t* = .54, *p* = .59).

The interaction model between agent conditions and Mechanisms pretests (Model 2) did not reveal significant interaction terms. The correlation coefficients at the group and student levels were still substantial (.58 and .47, respectively). Overall, this result does not provide evidence that the effects of agents on posttest differed by baseline understanding.

Table 5.5.
Effects of Agent Conditions on Systems Thinking

	Model 1		**Model 2**	
	Components	Mechanisms	Components	Mechanisms
Fixed effect	Coeff. (SE)	Coeff. (SE)	Coeff. (SE)	Coeff. (SE)
Intercept	-.11 (.24)	-.37 (.14)*	-.15 (.21)	-.45 (.22)*
Expert	.19 (.16)	.49 (.17)**	.20 (.15)	.44 (.17)**
Peer	.09 (.17)	.55 (.17)**	-.05 (.17)	.48 (.19)**
Pre-Components	.44 (.09)***	.08 (.09)	.45 (.10)***	.14 (.11)
Pre-Mechanisms	.18 (.09)*	.40 (.09)***	.28 (.13)*	.29 (.14)*
Pre-Mechanisms*Expert			-.08 (.16)	.08 (.18)
Pre-Mechanisms*Peer			-.15 (.16)	.18 (.18)
Random effect	Coeff. (SE)	Coeff. (SE)	Coeff. (SE)	Coeff. (SE)
Between-group var(U_{hj})	.02 (.01)	.006 (.002)	.02 (.01)	.006 (.002)
Covariance cov(U_{1j}, U_{2j})	.01 (.003)		.01 (.003)	
Within-group var(R_{hij})	.64 (.04)	.76 (.06)	.61 (.04)	.75 (.05)
Covariance cov(R_{1ij}, R_{2ij})	.42 (.03)		.32 (.03)	

Note. * *p* < .05, ** *p* < .01, *** *p* < .001. The coefficients are standardized. Reference group = No Agent. For all models, covariates included individuals' gender and classrooms' fixed effects

CHAPTER 6 CONCLUSION

This dissertation explores how students interact with and learn from conversational agents in scientific discussions. The explorations focus on two designs associated with facilitating learning interactions: an expert and a less knowledgeable peer agent (H. Chen et al., 2020; Kim & Baylor, 2016). Analyses of student interactions with one another and with the agents focus on occurrences of discussion moves, as well as their orders and co-occurrences towards larger knowledge construction patterns (B. Chen et al., 2017). Data draw from the chat logs from a total of 224 students, student surveys, and group artifacts (i.e., the concept maps students created). The studies come from two experiments: a within-subject design where students interacted with both agents (Studies 1-3), and a randomized cluster design where students interacted with no agent, an expert agent, or a less knowledgeable peer agent (Study 4).

The first study examines the **sequences of group interactions** with the agent designs. In a within-subject design, students in groups of two to three interacted with both the less-knowledgeable-peer and the expert-agent in a learning task (n = 52). There were no differences between agents in the frequencies of discussion moves. However, the sequence analyses revealed distinct patterns between the agents. Groups tended to show questioning and building on prior ideas with the less knowledgeable peer agent, whereas interaction sequences with the expert agent often involved responsiveness and reasoning. These patterns suggest that the agents might affect students' participation in discussion and subsequent learning in different ways.

In Study 2, I examine **how student groups fixed conversational breakdowns with the agents.** The study uses the same data sample and data sources as the first study. Breakdowns in conversations can negatively impact users' perceptions of and future usage of conversational agents. I found that learner groups generally attempted to fix the conversations during breakdowns, by repeating, reframing, and explaining their utterances. In particular, groups who more frequently employed explanations during breakdowns showed higher numbers of high-level systems thinking statements in their group chats. Furthermore, rephrasing and explanations were more common during interactions with the less knowledgeable peer agent.

Study 3 presents a case study to explore how **groups of emergent, mixed, and expanding knowledge interacted** with the agent designs. Data come from the chat logs of three student groups (*n* = 9) with emergent, mixed, and expanding knowledge of systems thinking. I found different patterns across groups. The expanding groups engaged in more transactive exchange and claim-making with the less knowledgeable peer agent. In comparison, the emergent group appeared to build on prior ideas more in response to the expert agent. These results illustrate the importance of considering group compositions in examining learning interactions and build towards Study 4.

Study 4 examines the **pathways between agent conditions, interactions with agents, and student learning**. Participants include 172 students ages 13-14 (36 groups). Students were randomly assigned to groups, where they built a concept map of the marine ecosystem with no agent, an expert agent, or a less knowledgeable peer agent. I found that the agents had a positive association with students' understanding of systems mechanisms, compared to the no-agent condition. This benefit was largely driven by students' increases in transactive exchange when interacting with the agents. Exploratory analyses further indicated potential differences in transactive exchange and Mechanisms understanding, along groups' variations in pretest scores. This finding highlights the need to consider designs that not only facilitate learning, but also equitable participation in discussion across student groups. I draw from these findings to discuss the study's implications and future directions.

Learning Implications

I found that the agents might deepen students' conceptual understanding, such as learning of systems mechanisms. A mechanism behind this effect is that the agents invite for productive talk moves, such as using reasoning and explanations to build on previous ideas in the group discussions. These findings contribute to emergent work on agents' interactions in collaborative contexts, by showing the pathway through which different agent designs might support learning (Adamson et al., 2014; Dyke, Howley, et al., 2013; Tegos & Demetriadis, 2017).

A direction for future work is to examine other pathways in which different agent designs can support discussions and learning in other domains, for example, in argumentation in English Language Arts or project-based design in Engineering. While the learning activity with the agents in this study took place in school settings, there is potential for future work to integrate the agents in informal learning activities such as museum simulations, to facilitate collaborative meaning-making of scientific phenomena.

Furthermore, the two agent designs might have supported different participation patterns in transactive exchange, with implications for how much students learned from the experience. In particular, students from different baseline knowledge more equally initiated and engaged in transactive exchange with the less knowledgeable peer agent, who framed the task as "learning together", compared to the expert agent with prompts such as "show me what you know". I hypothesize that the expert agent's prompts might have discouraged students who are more concerned with providing the correct responses (Clarke et al., 2016). These findings have implications for teachers and technology designers to frame the learning tasks to promote more equitable participation.

They also call for follow-up work that provides insights into how students from marginalized backgrounds, such as those with emergent knowledge or reservations about participation in scientific practices, perceive different agent designs and their contributions to the discussions. Such analyses may shed light on how learning designs can frame legitimate participation in scientific practices. Legitimate participation refers to opportunities for students to position themselves as active participants to engage in scientific practices and dialogues in classroom communities, as opposed to simply reciting facts and procedures from teachers (Lehrer & Schauble, 2006). To enact visions of instructional practices that legitimize students' participation (Kang et al., 2016; Stroupe, 2017), future work can examine the agent designs that effectively prompt students to discuss ideas, help students connect learning problems to lived experiences, and use students' ideas as resources to adapt participation prompts and learning tasks over time.

Design Implications

In this work, I explore the expert-less knowledgeable peer agents as one design paradigm that has been associated with productive help-seeking and help-giving in learning contexts (Graesser, 2016). Prior research has traditionally framed agents as facilitators or nudge providers (Dyke, Adamson, et al., 2013; Heidig & Clarebout, 2011; Rosé et al., 2015). However, findings from this dissertation illustrate that students might perceive agents as social partners in learning exchange. For example, student groups engaged in spontaneous questioning sequences with the less knowledgeable peer agent, and were generally tolerant of conversational breakdowns and attempted to fix such breakdowns. These findings illuminate how intentionally designed appearances and prompts for the agents can promote emergent social interactions.

Findings also call for considerations of design paradigms beyond expert and peer. Researchers have considered the feasibility of conversational agents as creative partners and co-storytellers (Ligthart et al., 2020; Sun et al., 2017). Applying these designs to collaborative contexts would require considerations of group dynamics within learning contexts, including how students perceive their relationships with peers and with the agents. Future work can turn to emergent research of agents in online communities. For example, agents can take on the role of an antagonist who uses off-task chats to build rapport with users, a storyteller that requires users to interact to advance plots, or an authority figure that constructs group norms (Seering et al., 2019). Such innovative designs might blur the positioning of agents as group outsiders and build towards more agile agent-human collaboration.

Implementation Implications

Conversational agents have gained increasing presence in the school and home environments. A follow-up question from the dissertation is how to implement this technology to support productive participation and learning at scale. In terms of technical design, the agents' interface and prompts can easily adapt to other contexts. The expert concept map that the agents base their prompts on can be swapped out with content from other domains to show relationships among concepts. Because the agents focus on facilitating idea elaboration rather than evaluation, existing transactive exchange and reasoning prompts are versatile in other learning situations.

More importantly, teachers' coordination needs to be considered for the agents to be integrated at scale. In this dissertation, the students and agents interacted in a single learning activity. Prior research with intelligent tutoring systems has demonstrated that such technology is most effective when it is used consistently and accounts for specific teacher routines (Holstein et al., 2019, 2017). Thus, an area of research that I will pursue is to understand how agents can provide ongoing feedback to teachers about students' discussions, coordinate with teachers to deploy collaborative prompts, evaluate the uptake and effectiveness of such prompts on students' learning, and encourage more equitable participation.

Together, the studies in this dissertation illustrate the potential of conversational agents in promoting productive discussion moves and scientific understanding. They also illuminate emergent interaction patterns associated with different agent designs. These findings reveal the need to design learning technology with considerations of the desired learning outcomes and interaction patterns, the unique social dynamics in different settings, and the interactions of individual and group factors for students to consider the agents as learning partners.

References

Adamson, D., Dyke, G., Jang, H., & Rosé, C. P. (2014). Towards an agile approach to adapting dynamic collaboration support to student needs. *International Journal of Artificial Intelligence in Education*, *24*(1), 92–124. https://doi.org/10.1007/s40593-013-0012-6

Chen, B., Resendes, M., Chai, C. S., & Hong, H.-Y. (2017). Two tales of time: uncovering the significance of sequential patterns among contribution types in knowledge-building discourse. *Interactive Learning Environments*, *25*(2), 162–175. https://doi.org/10.1080/10494820.2016.1276081

Chen, H., Park, H. W., & Breazeal, C. (2020). Teaching and learning with children: Impact of reciprocal peer learning with a social robot on children's learning and emotive engagement. *Computers and Education*, *150*, 103836. https://doi.org/10.1016/j.compedu.2020.103836

Clarke, S. N., Howley, I., Resnick, L., & Penstein Rosé, C. (2016). Student agency to participate in dialogic science discussions. *Learning, Culture and Social Interaction*, *10*, 27–39. https://doi.org/10.1016/J.LCSI.2016.01.002

Dyke, G., Adamson, D., Howley, I., & Rose, C. P. (2013). Enhancing scientific reasoning and discussion with conversational agents. *IEEE Transactions on Learning Technologies*, *6*(3), 240–247. https://doi.org/10.1109/TLT.2013.25

Dyke, G., Howley, I., Adamson, D., Kumar, R., & Rosé, C. P. (2013). Towards Academically Productive Talk Supported by Conversational Agents. In *Productive Multivocality in the Analysis of Group Interactions* (pp. 459–476). Springer US. https://doi.org/10.1007/978-1-4614-8960-3_25

Graesser, A. C. (2016). Conversations with AutoTutor Help Students Learn. *International Journal of Artificial Intelligence in Education*, *26*(1), 124–132. https://doi.org/10.1007/s40593-015-0086-4

Heidig, S., & Clarebout, G. (2011). Do pedagogical agents make a difference to student motivation and learning? In *Educational Research Review* (Vol. 6, Issue 1, pp. 27–54). Elsevier. https://doi.org/10.1016/j.edurev.2010.07.004

Holstein, K., McLaren, B. M., & Aleven, V. (2019). Designing for Complementarity: Teacher and Student Needs for Orchestration Support in AI-Enhanced Classrooms. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11625 LNAI*, 157–171. https://doi.org/10.1007/978-3-030-23204-7_14

Holstein, K., McLaren, B. M., & Aleven, V. (2017). Intelligent tutors as teachers' aides: Exploring teacher needs for real-time analytics in blended classrooms. *ACM International Conference Proceeding Series*, 257–266. https://doi.org/10.1145/3027385.3027451

Kang, H., Windschitl, M., & Stroupe, D. (2016). *Designing, Launching, and Implementing High Quality Learning Opportunities for Students That Advance Scientific Thinking Publication Date*. https://doi.org/10.1002/tea.21329

Kim, Y., & Baylor, A. L. (2016). Research-Based Design of Pedagogical Agent Roles: A Review, Progress, and Recommendations. *International Journal of Artificial Intelligence in Education*, *26*(1), 160–169. https://doi.org/10.1007/s40593-015-0055-y

Lehrer, R., & Schauble, L. (2006). *Scientific Thinking and Science Literacy.*

Ligthart, M. E. U., Neerincx, M. A., & Hindriks, K. V. (2020). Design patterns for an interactive storytelling robot to support children's engagement and agency. *ACM/IEEE International Conference on Human-Robot Interaction*, 409–418. https://doi.org/10.1145/3319502.3374826

Rosé, C. P., Goldman, P., Zoltners Sherer, J., & Resnick, L. (2015). Supportive technologies for group discussion in MOOCs. *Current Issues in Emerging ELearning*, *2*(1), 5.

Seering, J., Luria, M., Kaufman, G., & Hammer, J. (2019). Beyond dyadic interactions: considering chatbots as community members. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–13.

Stroupe, D. (2017). Ambitious Teachers' Design and Use of Classrooms as a Place of Science. *Science Education*, *101*(3), 458–485. https://doi.org/10.1002/SCE.21273

Sun, M., Leite, I., Lehman, J. F., & Li, B. (2017). Collaborative storytelling between robot and child: A feasibility study. *IDC 2017 - Proceedings of the 2017 ACM Conference on Interaction Design and Children*, 205–214. https://doi.org/10.1145/3078072.3079714

Tegos, S., & Demetriadis, S. (2017). Conversational Agents Improve Peer Learning through Building on Prior Knowledge. *Educational Technology & Society*, *20*(1), 99–111.