Designing an AI Partner for Jigsaw Classrooms

Jie Cao, Rachel Dickler, Marie Grace, Jeffrey Bush, Alessandro Roncone, Leanne M. Hirshfield, Marilyn A. Walker, Martha S. Palmer

Abstract The Jigsaw is a cooperative learning model that is widely used to organize small group classroom activities so that students have to be dependent on each other to succeed. There are a number of challenges, however, in supporting individual students and groups of students during jigsaw activities. In this paper, we present an AI partner designed to support jigsaw activities in middle school classrooms through dialogue. First, we present examples of a jigsaw activity taught in our partner classrooms, which focuses on block programming with sensors. We then share the design of a state space and example dialogue flow for a dual-function AI partner (supporting both group-level conversations and individual student answers) to guide young students during jigsaw activities. We look forward to discussing how the child-agent interactions that result from this design might improve both group dynamics and student knowledge gained during cooperative learning activities such as Jigsaws.

1 Introduction

Researchers of educational contexts have long agreed on the value of collaborative learning through interactions with one's peers [8, 13]. Achieving this vision of classroom learning remains a challenge in many educational settings. One proposed strategy is to structure collaborative activities to create an environment where students each play a unique role, with interdependent contributions, during small group activities [22, 21]. Specifically, the Jigsaw is one such form of cooperative learning where students first work in a small group to become experts in a particular content area, and then are regrouped with other students with different areas of expertise [14]. Each student in the second group is now the expert for their specific content area and can help to teach other students [2]. Jigsaw activities provide a valuable opportunity for social learning, however, it can be challenging for teachers to monitor and support each group of students (and individual students within each group [25]). While online

Jie Cao, University of Colorado Boulder, Boulder, CO 80302, e-mail: jie.cao@colorado.edu

tools have been developed to help with organizing Jigsaw materials (e.g., [10]), these tools do not yet provide group and individual-level support to students while they engage in real-time in the activity.

For this research, we explore examples of student interactions during a Jigsaw activity in a middle school classroom that is part of an ongoing research practice partnership [17]. This partnership integrates co-designed storylined curriculum [20, 16] and programmable sensor technology to empower youth to conduct personally meaningful investigations while programming and wiring their own sensor systems in a curriculum called Sensor Immersion [7]. During this curriculum, students learn block-programming to control 3 sensors: a **sound sensor** that detects the volume and intensity of sound in the room, a **moisture sensor** that detects levels of water in the soil or other material, and an **environmental sensor** that detects the temperature and level of carbon dioxide in the air. Students first learn how to program just one sensor (either the sound, moisture, *or* environmental sensor) in a small group. The students are then regrouped such that there is an "expert student representative" for each sensor within the new group. Students in the new group complete a worksheet together with questions about the specific sensors (i.e., closed questions with only one correct answer) as well as questions that build up towards inventing a new technology with all three sensors that could solve a real-world problem in their community (i.e., open questions with multiple possible answers that integrate all sensors).

The Jigsaw activity in this curriculum was added as a part of a collaboration with the NSF National AI Institute for Student-AI Teaming (iSAT) [15] which uses Sensor Immersion curriculum as a context for developing AI agents to support collaborative learning. While the jigsaw activity here was designed to catalyze cooperative learning, it is new and can benefit from additional support. Our case study examines how these activities can benefit from additional scaffolds to achieve their maximum potential. We theorize that there is an opportunity for AI to support these Jigsaw interactions. In the next sections, we present case studies on Jigsaw classrooms and then outline the design of an AI partner to support students during Jigsaw activities.

2 Background and Motivation

To show the potential for an AI partner to improve the collaboration during authentic jigsaw classroom implementations, we use data from classrooms that used the sensor immersion curriculum, but without constraining teachers from partner schools to a specific implementation of the curriculum. Our case study covers 56 jigsaw activities, implemented by 18 teachers across 12 schools. We start by analyzing the worksheets that students complete during the jigsaw activity, then we analyze videos randomly sampled from different classrooms to explore variations in teacher implementations. Based on the case study, in Section 3, we propose an Options Framework to hierarchically reinforce our dual-AI partners in Jigsaw classrooms.

Jigsaw Outcome: Worksheet. Table 1 summarizes the goals of the lesson plan: (1) Consensus understanding of different sensors, (2) Creative thinking for future

Goals and Questions	Answer Analysis
Goal 1: How are the sensors and data displays similar and different? (**Consensus Understanding**)	Consensus formed within a group, but Informal, Vague, Incorrect
(1) What data can the sensor collect? (Closed)	
(2) What wires and connections did you use? (Closed)	
(3) Explain your code to your group. What blocks did you use? How does it work? (Semi-Closed)	
(4) How did your sensor system display the data (music, lights, numbers, letters)? (Semi-Closed)	
(5) What challenges did the group encounter with this sensor? (Semi-Closed)	
Goal 2: how can sensor systems help us in our scientific investigations? (**Creative Thinking**)	Some are Broad and Interesting, but Low Diversity, Not Creative
(6) What investigations in your home, school, or community can we do with the sensors? (Open)	
(7) What questions can we answer with data from this sensor? (Open)	
(8) What problems in your community could you use the sensors to help solve? (Open)	
(9) What kinds of scientific questions could the sensors help you answer? How could the sensors help improve your community? (Open)	

Table 1: Goals of Jigsaw Classroom and Worksheets (Sensor Learning). Please refer to the Appendix for more details.

project planning. A worksheet consists of two sets of questions designed to help students organize their thoughts. We categorize the questions for each goal into closed questions and open questions. For example, the first 5 questions have expected answers (with only a little freedom to customize, e.g., the data display and code). These closed or semi-closed questions are designed to help students build a consensus understanding of how to use the sensors. The questions with open answers are designed to help students brainstorm future project ideas on how to use those sensors in the real world. Our analysis revealed that the worksheet is supporting students in recognizing one anothers' expertise and that students are both listening to their peers and recording what their peers say on the worksheet. However, there is still room for improved student activity on this worksheet. As shown in Table 1, our analysis of the Jigsaw worksheets shows that: (1) for closed questions, a consensus is formed within a group, while the answers are **generally informal, vague, and sometimes incorrect**. (2) for open questions, some answers are broad and interesting, however, many are **vague, repetitive and obsolete**. This shows that students could benefit from additional support helping them increase the formality, , relevance and specificity of their responses.

Conversational Flow in a Jigsaw Classroom. To identify the reasons for the shortfalls of the Jigsaw implementation, we also analyze the classroom videos. Fig. 1 shows a typical conversation flow within each jigsaw group. The students first establish who is the expert for each sensor, and then move to answer each question. When a question explicitly splits into 3 columns for each sensor, the expert for each sensor will lead the **sharing** first, and then all the students will share their comments, questions, and **discuss** that sensor. Meanwhile, they will **fill** in the answer. The same flow extends to each sensor for that question. For questions on integrating multiple sensors, the students are free to brainstorm in any order. While analyzing the conversations, we discovered three themes for areas where the collaborative learning could be improved. (1) **necessary intervention and encouragement:** some group discussions are low-energy and often off-task, resulting in half-baked worksheets.

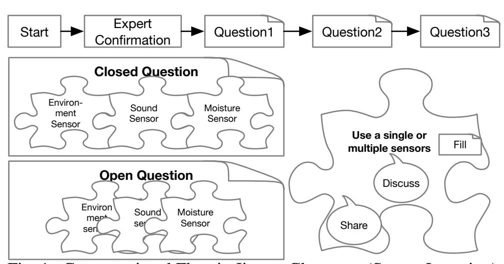

Fig. 1: Conversational Flow in Jigsaw Classroom(Sensor Learning).

In contrast, when there is a teacher or student leading the group, the flow in Fig. 1 is more efficient. (2) **enhanced content support:** during the **sharing** phase, some students get blocked, and will either ask his/her previous expert group members for help, or feel unsure about sharing. We posit that an AI agent who offers accurate content support could enhance the confidence of the student experts. (3) **enriched brain-storming:** Students need help to avoid duplicating proposals.

3 AI Partner Design for Jigsaw Classrooms

Building AI agents based on the above lessons requires designs at different levels of complexity. To offer both group-level and individual-level interactions, we design a dual-partner solution: (1) Answer Assistant (*AA*): a per-student partner to help students with answer completion; (2) Jigsaw AI (*JA*): a per-group partner to observe the group conversation by jointly interacting with *AA* and the students. Fig. 2 shows the overarching design of our AI agent. Ovals and rectangles show the states and primitive actions that can be widely used in general jigsaw AI state machine. Polygons are content-based components that require more fine-grained interactions between *JA* and*AA* (Table 2). To minimize the interventions, agents will only handle a few exceptional states actively, mostly quietly track the dialog state, and passively wait for explicit commands to intervene. We leave the detailed design of dialog management and reinforcement learning framework in Table 2 in the Appendix. In the following, we focus on language-based interventions in our AI agents.

Flow Control and Encouragement. Flow control is useful in three kinds of states: (1) When the students are off-task [4], our agent could **resume the dialog state** [3] by saying "How is the wiring going?" (2) When the typical steps *sharing*, *discussion* and *filling* steps (Fig. 1 for each question that dies down, our agent may **check the status** by asking "That was good sharing, did you note this as done and move on?" (3) at the end of a question discussion: e.g, our agent could optionally **summarize the dialog** for this question [9, 28] and **issue a completion reminder** and **advance** the next question. Encouragement is another key function of our AI agent. When there is *not enough discussion*, our agent could say "Keep going" to **encourage new sharing**, or say "Can you elaborate your idea?" to **evoke deeper explanation**. When a student is *dominating the discussion*, our agent may **encourage other students to join** by saying "What do you think, Student A?" Besides the above, flow-control and encouragement will also help in the following content-based components.

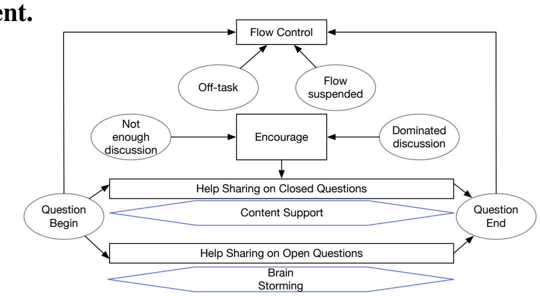

Fig. 2: The Design Overview of Jigsaw AI.

Content Support. The goal of content support is to help students to obtain a clear consensus understanding, thus faciltating sharing and answer completion, and reducing informal, vague and incorrect answers in the sheets. The key challenge for content support is knowledge-enhanced dialog [19, 26, 23]. For example, when students got stuck, they will hesitate to share or ask their partners in their previous expert group for help, which not only slows the current group but also disrupts other groups from their discussions. Here, *JA* can help the expert share with the group by **retrieving the relevant content** generated in the previous expert groups such as course materials, wiring, codes, and experiment notes. To check the accuracy and give hints for completion, we do a **semantic match** between students' answers with structured knowledge representations for answers and previous course content. For example, by using Abstract Meaning Representations [6], *AA* can highlight the differences and suggest answers. Also, the per-group *JA* can aggregate the answer-checking reports from each *AA*, and help build a consensus understanding when bringing students together to discuss their answers. Please see more details in Appendix, Table 2 demonstrates example interactions between *AA* and *JA*.
Brainstorming. The goal of brainstorming is mainly to help students plan novel collaborative project ideas. The key challenge for brainstorming is how to evoke diverse ideas via dialog. When answering open questions for project planning, students struggle to get past well-known ideas already mentioned in class, such as air pollution and noise detection. Instead, we can narrow down open questions by **providing a specific scenario** outside the classroom, thus guiding their imaginations. For example, augmenting **question generation** with world knowledge and common sense reasoning [11, 12, 1, 27], a question like "What kind of sensors can be useful in Wimbledon games?" could prompt students to creatively use the sensors for novel cases. Furthermore, the per-group *JA* can do the following tasks to enrich brainstorming: (1) **ask clarification questions**, e.g, "I like that idea, but how do you combine the results of the two sensors?", (2) **offer reasoning**, e.g., "In a wildfire, the temperature and pressure will change, which sensor can help here? What other changes can be detected with our sensors?" (3) **feasibility check**, e.g., "We don't have an emotion sensor in this project". (4) **diversity check**, e.g., "Is this idea similar to that previous one? What are the differences here?" (5) **personalized questions**, e.g., "Instead of Wimbledon games, you like sailing, how can we help there?"

4 Conclusion

Our case study review shows how Jigsaw activities while promoting collaboration, have the potential to better the achievement of their instructional goals if augmented with additional scaffolding. We have presented exciting opportunities to provide that scaffolding with various designs for AI agents in jigsaw classrooms. By outlining a dual-partner design, we demonstrated the potential language-based interventions and highlighted the challenges and future research directions.

Acknowledgements This research was supported by the NSF National AI Institute for Student-AI Teaming (iSAT) under grant DRL 2019805. The opinions expressed are those of the authors and do not represent views of the NSF.

Appendix

Detailed Analysis on Jigsaw sheets. Consider question (1) of the closed questions in Table 1. For the environmental sensor, the expected answer is that it can collect 6 different things (Temperature in degrees Celsius, Temperature in degrees Fahrenheit, CO2, Humidity, Pressure, Total Volatile Organic Compound (TVOC)). However, among 27 sampled worksheets, only five are correct (including two of them with the answer "6 things" without further detail). Other answers include: (1) two incorrect answers related to "wind", which is related to pressure but not directly measurable; (2) six vague answers ("measure the outside world"); (3) answers that only mention the four things learned in the previous tutorial lessons. In summary, the Jigsaw classroom implementation results that answer for closed questions are generally **vague, informal, and inaccurate understanding** of sensor usage.

For open questions, consider question (9). The goal is for students to brainstorm how to combine multiple sensors for a new problem. However, over 27 sample sheets, six are not answered, some answered with **broad and vague** answers, e.g, "any questions in our daily life", while most answers use only one sensor for a well-known the problem, such as using environment sensors for "air pollution", sound sensors for "noise detection", and soil sensors for the "garden". In summary, the answers to open questions are vague, duplicated, and obsolete. It is possible that a well-designed

Options	Initial Set: I_o	Option's Policy: π_o	Termination Condition: β_o
Content Support	1. Command CONTENT-SUPPORT issued. 2. Students blocked when sharing in a closed question	1. Encourage. 2. Flow control. 3. Offer support (e.g., wiring, code, and other materials). 4. Record students feedback if cannot support it. 5. Check if any further help is needed?	1. Commands STOP-CONTENT-SUPPORT or CHECK-ACCURACY issued. 2. Students need no more help. 3. No reply from students.
Check Accuracy	1. Command CHECK-ACCURACY issued. 2. Student finished the answer filling in a closed question.	1. Aggregate reports from AA with dialogue analysis and show them in the group. 2. Explaining the accuracy report, evoking discussion, and refilling. 3. Check any further help needed?	1. Command STOP-CHECK-ACCURACY or CONTENT-SUPPORT issued. 2. No more questions. 3. No reply from students.
Brain Storming	1. Command BRAIN-STORMING issued. 2. Students blocked or no sharing when during the open question	1. Encourage. 2. Flow control. 3. Question generation to evoke creative thinking. 4. Remark students' feedback if cannot support it. 5. Check if any further help needed?	1. Command STOP-BRAIN-STORMING or CHECK-DIVERSITY issued. 2. Students need no more help 3. No reply from students.
Check Diversity	1. Command CHECK-DIVERSITY issued. 2. Students finished answering an open question.	1. Aggregate reports from AA with dialogue analysis, and show it to the group. 2. Explaining the diversity report, evoking discussion and refilling. 3. Check if any further help needed?	1. Commands STOP-CHECK-DIVERSITY or BRAIN-STORMING issued. 2. Students need no more help. 3. No reply from students.

Table 2: Detailed setting for Options 'Content Support', 'Check Accuracy', 'Brain Storming', 'Check Diversity'

digital worksheet with format constraints could help somewhat with this, but the occurrence of inaccurate, duplicated, and obsolete answers indicate the need for other solutions to encourage collaborative learning activities in the classroom.

Reinforcement Learning and Interaction Examples. The overall **reward** is aimed for encouraging the average accuracy for closed questions r_{closed}, average diversity measure for open questions r_{open}, conversation time [18] among students $r_{stu} = \frac{t_{stu}}{t_{total}}$ and balanced multi-party conversation the ratio is calculated by using the coefficient of variation of student talk time $r_{balance} = 1 - \frac{\sigma(t_{stu})}{\mu(t_{stu})}$. Hierarchical reinforcement learning, such as Options Framework [24, 5] can potentially offer a scalable design for structured exploration on those content-based components(polygons) as shown in Fig. 1. A typical Markov option o in the Options Framework consists of three parts: an initial set of states I_o, a local policy π_o for that option o, and the termination condition β_o. Table 2 list the details of our design for options including content support, accuracy check, brain-storm, and diversity check. To carefully mitigate the interventions using AI in students' dialog, we only use AI to handle a few exceptional states, but let students decide when to use the AI agent with active commands.

References

1. Abdelghani, R., Wang, Y.H., Yuan, X., Wang, T., Sauzéon, H., Oudeyer, P.Y.: Gpt-3-driven pedagogical agents for training children's curious question-asking skills. arXiv preprint arXiv:2211.14228 (2022)
2. Anderson, F.J., Palmer, J.: The jigsaw approach: Students motivating students. Education **109**(1) (1988)
3. Andreas, J., Bufe, J., Burkett, D., Chen, C., Clausman, J., Crawford, J., Crim, K., DeLoach, J., Dorner, L., Eisner, J., et al.: Task-oriented dialogue as dataflow synthesis. Transactions of the Association for Computational Linguistics **8**, 556–571 (2020)
4. Anonymous: Navigating wanderland: Highlighting off-task discussions in classrooms (2023)
5. Bacon, P.L., Harb, J., Precup, D.: The option-critic architecture. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 31 (2017)
6. Banarescu, L., Bonial, C., Cai, S., Georgescu, M., Griffitt, K., Hermjakob, U., Knight, K., Koehn, P., Palmer, M., Schneider, N.: Abstract meaning representation for sembanking. In: Proceedings of the 7th linguistic annotation workshop and interoperability with discourse, pp. 178–186 (2013)
7. Biddy, Q., Chakarov, A.G., Bush, J., Elliott, C.H., Jacobs, J., Recker, M., Sumner, T., Penuel, W.: A professional development model to integrate computational thinking into middle school science through codesigned storylines. Contemporary issues in technology and teacher education **21**(1), 53–96 (2021)
8. Bransford, J.D., Brown, A.L., Cocking, R.R., et al.: How people learn, vol. 11. Washington, DC: National academy press (2000)
9. Feng, X., Feng, X., Qin, B.: A survey on dialogue summarization: Recent advances and new frontiers. arXiv preprint arXiv:2107.03175 (2021)
10. Gallardo, T., Guerrero, L.A., Collazos, C., Pino, J.A., Ochoa, S.: Supporting jigsaw-type collaborative learning. In: 36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings of the, pp. 8–pp. IEEE (2003)
11. Hwang, J.D., Bhagavatula, C., Le Bras, R., Da, J., Sakaguchi, K., Bosselut, A., Choi, Y.: (comet-) atomic 2020: On symbolic and neural commonsense knowledge graphs. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, pp. 6384–6392 (2021)

12. Kulshreshtha, D., Shayan, M., Belfer, R., Reddy, S., Serban, I.V., Kochmar, E.: Few-shot question generation for personalized feedback in intelligent tutoring systems. arXiv preprint arXiv:2206.04187 (2022)
13. Learn II, H.P.: How people learn ii: Learners, contexts, and cultures (2018)
14. Moskowitz, J.M., Malvin, J.H., Schaeffer, G.A., Schaps, E.: Evaluation of jigsaw, a cooperative learning technique. Contemporary educational psychology **10**(2), 104–112 (1985)
15. NFS-iSAT: NSF National AI Institute for Student-AI Teaming. `https://www.colorado.edu/research/ai-institute/` (2023). [Online; accessed 02-13-2023]
16. Penuel, W.R., Allen, A.R., Henson, K., Campanella, M., Patton, R., Rademaker, K., Reed, W., Watkins, D., Wingert, K., Reiser, B., et al.: Learning practical design knowledge through co-designing storyline science curriculum units. Cognition and Instruction **40**(1), 148–170 (2022)
17. Penuel, W.R., Gallagher, D.J.: Creating Research Practice Partnerships in Education. ERIC (2017)
18. Ravari, P.B., Lee, K.J., Law, E., Kulić, D.: Effects of an adaptive robot encouraging teamwork on students' learning. In: 2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN), pp. 250–257. IEEE (2021)
19. Reddy, S., Chen, D., Manning, C.D.: Coqa: A conversational question answering challenge. Transactions of the Association for Computational Linguistics **7**, 249–266 (2019)
20. Reiser, B.J., Novak, M., McGill, T.A., Penuel, W.R.: Storyline units: An instructional model to support coherence from the students' perspective. Journal of Science Teacher Education **32**(7), 805–829 (2021)
21. Roschelle, J., Rafanan, K., Bhanot, R., Estrella, G., Penuel, B., Nussbaum, M., Claro, S.: Scaffolding group explanation and feedback with handheld technology: Impact on students' mathematics learning. Educational Technology Research and Development **58**, 399–419 (2010)
22. Roschelle, J., Teasley, S.D.: The construction of shared knowledge in collaborative problem solving. In: Computer supported collaborative learning, pp. 69–97. Springer (1995)
23. Shaier, S., Hunter, L., Kann, K.: Mind the knowledge gap: A survey of knowledge-enhanced dialogue systems. arXiv preprint arXiv:2212.09252 (2022)
24. Sutton, R.S., Precup, D., Singh, S.: Between mdps and semi-mdps: A framework for temporal abstraction in reinforcement learning. Artificial intelligence **112**(1-2), 181–211 (1999)
25. Van Es, E.A., Sherin, M.G.: Learning to notice: Scaffolding new teachers' interpretations of classroom interactions. Journal of technology and teacher education **10**(4), 571–596 (2002)
26. Wang, J., Liu, J., Bi, W., Liu, X., He, K., Xu, R., Yang, M.: Improving knowledge-aware dialogue generation via knowledge base question answering. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 9169–9176 (2020)
27. Wei, J., Wang, X., Schuurmans, D., Bosma, M., Chi, E., Le, Q., Zhou, D.: Chain of thought prompting elicits reasoning in large language models. arXiv preprint arXiv:2201.11903 (2022)
28. Zhang, Y., Ni, A., Yu, T., Zhang, R., Zhu, C., Deb, B., Celikyilmaz, A., Awadallah, A.H., Radev, D.: An exploratory study on long dialogue summarization: What works and what's next. arXiv preprint arXiv:2109.04609 (2021)