Connecting the Dots Towards Collaborative AIED: Linking Group Makeup to Process to Learning

Angela Stewart[0000-0002- 6004-9266] and Sidney K. D'Mello[0000-0003-0347- 2807]

[1] University of Colorado Boulder, Boulder CO 80309, USA
{angela.stewart, sidney.dmello}@colorado.edu

Abstract. We model collaborative problem solving outcomes using data from 37 triads who completed a challenging computer programming task. Participants individually rated their group's performance, communication, cooperation, and agreeableness after the session, which were aggregated to produce group-level measures of subjective outcomes. We scored teams on objective task outcomes and measured individual students' learning outcomes with a posttest. Groups with similar personalities performed better on the task and had higher ratings of communication, cooperation, and agreeableness. Importantly, greater deviation in teammates' perception of group performance and higher ratings of communication, cooperation, and agreeableness negatively predicted individual learning. We discuss findings from the perspective of group work norms and consider applications to intelligent systems that support collaborative problem solving.

Keywords: Collaborative problem solving, collaborative learning, group composition, teamwork, small group work.

1 Introduction

Collaborative problem solving (CPS) is considered a core competency for the 21st century workforce [1]. Accordingly, modern curricula increasingly include tasks that require students to work together to achieve some goal [1]. But it is not enough for students to simply work together – they should also learn relevant content/skills [1, 2] as well as learn to collaborate effectively. Despite its importance to educators, small groups rarely achieve their full potential. They are often plagued by *process loss*, where the group does not achieve a level of performance compared to what is theoretically achievable [3]. Additionally, process gain, where group performance is better than any combination of individual members' performance, is difficult to achieve [3].

Can next-generation artificial intelligence in education (AIED) systems facilitate development of CPS skills similar to their success in promoting learning in more traditional domains [4]? An AIED-CPS system could monitor group CPS processes to create a *group model* and intervene when students are not collaborating effectively. Such systems could consider a variety of static factors related to group composition (e.g., gender, personality, motivation, prior knowledge) and dynamically monitor aspects of the collaboration process (e.g., level of frustration, individual member participation, number of ideas generated). The system could provide motivating feedback or offer hints when problems are uncovered with the end goal of promoting successful task performance, satisfaction with the collaborative process, and individual learning.

What principles should guide the design of an AIED-CPS system? There has been considerable research in the fields of small group problem solving [5, 6], collaborative learning (CL) [2, 3] and computer-supported collaborative learning (CSCL) [7, 8]. However, these areas tend to be disjoint, either focusing on group makeup factors [9–11], the collaborative process itself [3], optimal group outcomes [12, 13], or learning as a result of the collaboration [2]. Our work aims to connect the dots –from group makeup to collaborative process to individual learning (Fig. 1) – by addressing the following questions: (1) How does group makeup affect subjective and objective collaboration outcomes? (2) How does subjective and objective collaborative outcomes affect individual learning?

We consider group makeup from the perspective of gender and personality as they are related to CPS outcomes (see Section 1.1). Further, both of these factors are fairly static and can be quickly measured through self-report, making them ideal candidates to incorporate into a group model (group equivalent of a student model in AIED).

We emphasize three types of collaborative outcomes: group-level objective task performance, group-level subjective perception of the collaboration, and individual-level learning. It is important to consider both group- and individual-level outcomes because they may depend on different factors and might even be negatively related. Furthermore, an AIED-CPS system cannot solely produce positive group or individual outcomes, but must balance the two. We also consider objective outcomes in the form of task performance and learning, as well as subjective perceptions of the CPS process as both are valued by individuals [3, 12, 14].

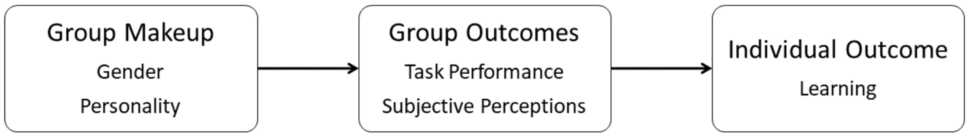

Fig. 1. In this work, we connect the dots of group makeup to group outcomes to individual learning outcomes.

1.1 Related Work

The CPS literature is vast (see review by [3]). For example, researchers have investigated various factors related to CPS outcomes, such as social loafing [14], awareness and shared intentionality [15], information distribution [11], and problem solving approach [16]. CPS research in the learning sciences has mainly focused on online forums or massively open online courses [17, 18], CPS skill assessment [12, 13], automated support of collaborative discussions [19], and collaborative discourse [20]. To keep scope manageable, we focus on related work germane to our research questions, specifically group composition and cohesion, subjective perceptions of the CPS process, and group to individual transfer of learning [21].

There is a comprehensive literature on group composition and its relationship to collaborative processes and outcomes. Researchers have investigated group composition from the perspective of team familiarity [11], gender [9, 10, 22], ethnicity [22], team member ability [22], and group diversity [23].

Related research on group cohesion [3, 24–26] – the extent to which group members feel like they are part of the group and wish to remain in the group [24] – has indicated that cohesiveness can be affected by demographic factors (e.g. sex, race, age), attitudes or beliefs, and even task performance [3, 27]. Though research has generally found that more cohesive groups perform better [3, 24], Langfred [24] found that the relationship between cohesiveness and performance is moderated by the expectations of group member behavior, referred to as group work norms. When group work norms focused on behavior related to the task rather than social behavior (such as agreeableness or conflict minimization), high cohesiveness was predictive of group effectiveness.

Beyond performance on the collaborative task, it is also important to consider how teammates perceive the collaboration [28–30]. Attitudes and perceptions can be affected by process conflict [31], number of people collaborating (pairs versus multiple individuals) [28], or group affect [29]. Similarly, in the context of collaborative learning, students must also engage in transfer learning, where they apply knowledge gained in one context to a different context. Transfer in small group learning is achievable as indicated in a meta-analysis [32] of 38 studies where students learned independently or in a small groups The weighted average of the effect size (Hedges' *g*) across studies was .30 sigma. The authors concluded that small group learning promotes a deeper understanding of the content as students must coordinate differing viewpoints, an opportunity not present in individual learning [32].

In an effort to understand the factors that contribute to successful transfer, Olivera and Straus [21] compared transfer learning outcomes for individuals learning alone, students in a group, and students observing a group. They found that students learning either in a group or by observing a group outperformed students learning individually. They concluded that cognitive factors associated with the collaborative process were important for transfer learning. However, social factors related to the collaboration might not be as important for transfer learning because students who merely observed groups obtained the same benefits as working in a group.

1.2 Contribution

Our work contributes to basic research on collaborative problem solving with an eye towards developing AIED systems to facilitate CPS. We provide insights into the connection between group composition, group outcomes, and individual learning as a result of the collaboration. We focus on a combination of factors and outcomes because processes that lead to success on one type of outcome might not lead to success in the others (subjective vs. objective outcomes, group vs. individual outcomes).

Additionally, our context of inquiry is novel. We consider CPS in triads, which afford seven units of analysis (3 individuals, 3 dyads, and the triad), thereby providing a range of interesting interactions to explore. We focus on triad-level analysis to understand the group as a whole and individual-level analysis to understand how students situated within a group learn. We investigate CPS in the domain of computer programming where students used an interactive visual programming environment (see Section 2.3) to learn computer programming skills. Finally, we study CPS during virtual collaborations as an increasing amount of collaboration in the 21st century workplace is performed by teams communicating virtually [1].

2 Methods

2.1 Participants

Participants were 111 (63.1% female, average age = 19.4) undergraduate students from a medium-sized private Midwestern university, who were compensated with course credit. Students were 74.8% Caucasian, 9.9% Hispanic/Latino, 8.1% Asian, 2.7% Other, 0.9% Black, 0.9% American Indian/Native Alaskan, and 2.7% did not report ethnicity. Students were assigned groups of three based on scheduling constraints, with a total of 37 groups. Nineteen students from ten teams indicated they knew at least one person from their team prior to participation. Students had no prior computer programming experience.

2.2 Learning Environment

Students learned basic computer programming principles as a group using code.org's Minecraft-themed Hour of Code (Fig. 2) [33]. Hour of Code is an online resource for students of all ages to learn basic programming principles in an hour. It uses Blockly [34], a visual programming language that represents lines of code (such as if statements) as interlocking blocks. Blocks only interlock in a syntactically correct manner, allowing students to focus on the coding logic and programming principles, without considering syntax errors.

2.3 Procedure

Students were each randomly assigned to one of three computer-enabled rooms with video-conferencing capabilities and screen sharing through Zoom (https://zoom.us). Each computer had a webcam with a microphone so students could see and hear each other. During the collaboration, one student's screen was shared so that everyone viewed the same content.

Surveys. Each student individually filled out demographic data including gender, age, major, and standardized test score (ACT and/or SAT). Students also completed the short version of Big Five Inventory (BFI) [35] to assess the following personality traits:

extraversion, agreeableness, conscientiousness, emotional stability (previously neuroticism), and openness to experience.

Introductory levels. After individually completing the surveys, students were asked to complete five levels and view three accompanying videos that taught basic computer programming principles, such as loops and if statements. Students were required to build structures within the game and navigate around obstacles. One randomly assigned student was tasked with controlling the group's actions in the environment. The other two students were tasked with contributing to the collaboration. Students were specifically instructed to collaborate as a team to complete the levels within 20 minutes.

Students then *individually* rated their satisfaction with their team's performance, communication, cooperation, and agreeableness using six point scales. Specifically, students rated the statements "I am satisfied with my team's performance at completing the lessons," "I am satisfied with how we communicated with each other," "I am satisfied with how we mutually cooperated to complete the lessons," and "I am satisfied with how agreeable my teammates are." They indicated whether they were very dissatisfied, somewhat dissatisfied, slightly dissatisfied, slightly satisfied, somewhat satisfied, or very satisfied for all items.

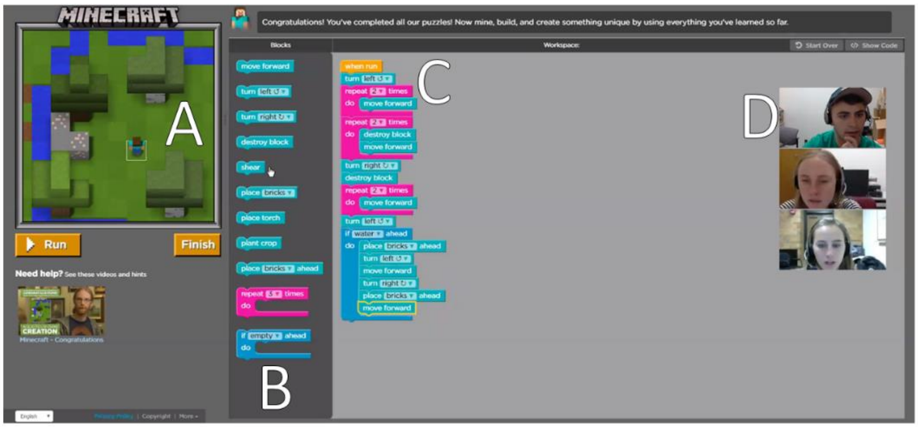

Fig. 2. Minecraft-themed Hour of Code from Code.org. Students could visualize the results of running their code (A), a code bank of possible blocks to use (B), the code they generated (C) and their team's faces (D).

Challenge Levels. Teams were then tasked with working together to complete a challenging programming task in the same Hour of Code environment. The same team member who controlled interaction with the environment during the initial lessons also controlled the interaction during the coding challenge. In the challenge, teams were given 20 minutes to build a 4×4 brick building with the following constraints: use at least one if statement, use at least one repeat loop, build at least three bricks over water, and use 15 blocks of code or less. Students *individually* completed the same subjective measures of their team's performance, communication, cooperation, and agreeableness with the wording adapted for the challenge level.

Posttest. Students *individually* completed a ten-item researcher-created multiple choice test to assess their conceptual knowledge of coding concepts (such as repeat loops and if statements). Each posttest item had one correct answer out of four possible answers, and the possible range of scores was 0% to 100%. Fig. 3 shows an example posttest question.

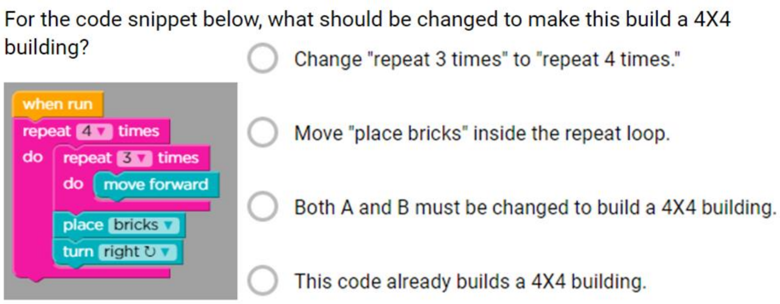

Fig. 3. An example posttest question.

2.4 Measures

We focus on team's performance during the challenge phase as this was the critical phase of the study; the introductory phase merely familiarized the teammates with each other and with the programming environment.

Subjective Perceptions. For each team, we computed the mean and standard deviation (SD) of the three students' performance ratings. The mean measured the overall assessment of the team's performance, and the SD measured divergence of assessments. Across all teams, students tended to rate the team's performance quite highly (performance mean, *M* = 4.52, *SD* = .89) with relatively little disparity (performance SD, *M* = 1.03, *SD* = .67).

We combined measures of communication, cooperation, and agreeableness (CCA) because ratings were highly correlated (Cronbach's alpha = .89). We first computed the mean and of the communication, cooperation, and agreeableness ratings for each team member, yielding three aggregated measures of CCA (one per team member). We took the mean of the CCA measure across the triad, to yield a single measure of CCA mean. Similarly, we took the SD of the CCA measure across the triad, to yield a single measure of CCA SD. The CCA SD was strongly correlated with the CCA mean (r = -.76), so we focused on the mean. Overall, students were quite satisfied with their team's CCA (*M* = 5.45, *SD* =.56).

Objective task score. Each team's final solution was scored by two independent raters on the five challenge criteria (see Section 2.3). The two raters reconciled any disagreements via discussion. Each criterion was worth a single point, so scores ranged from 0 to 5 with a mean of 2.86 (*SD* = 1.06).

Individual learning (Posttest). We computed the average of the ten posttest items for each student (*M* = 43.0%; *SD* = 15.7).

Group Makeup. We considered group makeup based on gender using two binary variables: whether all members of the group were the same gender (23%) and whether the group was female-dominated (71%). Group makeup based on personality was measured using dissimilarity scores of the BFI. Specifically, Euclidean distances between the 5-dimensional BFI vectors of each pair of participants in a team (three pairs in total) was computed and then averaged to obtain a single measure of personality dissimilarity. A larger value indicated that participants had more disparate personalities (M = 7.19, SD = 1.58).

2.5 Data Treatment

Data was collected over two semesters, with a minor experimental change made between semesters, where we added a five minute warning before the end of the challenges. In order to prevent this change from influencing outcomes, we z-scored each of the following measures by semester: performance mean, performance SD, CCA mean, task score, and individual posttest score. Two teams were removed from the group-level analyses because students did not to report ACT scores (which was used as a covariate in several analyses – see Section 3.2).

3 Results

3.1 Relationship between Group-Level Outcomes

At the group-level, task score was positively correlated with subjective perceptions of performance (r = .56, p < .001) as well as perceptions of the collaborative process (i.e., mean of communication, coordination, and agreeableness) (r = .44, p = .01), which were also correlated (r = .57, p < .001). Deviation in each individual's perception of group performance was not significantly correlated with task score (r = -.26, p = .13).

3.2 Predicting Group-Level Outcomes from Group Makeup

We used linear regression to predict subjective (performance mean, performance SD, and CCA mean) and objective (task score) group-level outcomes from group makeup (mixed gender, female dominated, and personality dissimilarity). We included the mean ACT score of the group and binary team member familiarity as covariates. Neither gender-based measure of group makeup was predictive of task performance or perception (see Table 1). However, teams with more disparate personalities performed significantly worse (task score) and were less satisfied with their team's communication, cooperation, and agreeableness, but not their team's performance. A follow-up analysis on individual dimensions of personality (obtained by including the standard deviation across the three team members as a predictor in five separate regression models) indicated that deviation in emotional stability predicted CCA mean (B = -.40, p = .02), but no individual dimensions of personality predicted task score.

Table 1. Results of predicting group-level outcomes from gender and personality variables. The standardized beta coefficients (with p-values in parenthesis) are shown.

	Outcome Variables			
	Task Score	**Perf. Mean**	**Perf. SD**	**CCA Mean**
Mixed Gender	.01 (.94)	.04 (.83)	.10 (.59)	-.10 (.59)
Female Dominated	.27 (.13)	.01 (.97)	-.23 (.20)	.14 (.47)
Personality Dissimilarity	**-.36 (.03)**	-.02 (.91)	-.04 (.83)	**-.33 (.07)**

Note, Perf. Mean and Perf. SD indicate the mean and SD of the subjective performance measure. CCA Mean indicates the mean of the subjective communication, cooperation, and agreeableness measure.

3.3 Predicting Posttest Scores from Group Outcomes and Group Makeup

We used linear mixed effects models for individual posttest scores because participants were nested within groups. The fixed effects included group-level task scores, performance mean, performance SD, CCA mean, personality dissimilarity, mixed gender, and female dominated groups (each in separate models). Individual ACT scores, gender, teammate familiarity, and whether the student was controlling the interaction with the environment (see Section 2.3) were included as covariates. Group identity was a random factor (intercept only).

None of the group makeup variables were predictive of posttest score (B = .03, p = .70 for personality dissimilarity, B = -.26, p = .33 for mixed gender, and B = .32, p = .36 for female dominated).

Surprisingly, neither task score (B = .08, p = .69) nor performance mean (B = .07, p = .72) was predictive of individual posttest scores. However, participants' perceptions of the group's communication, cooperation, and agreeableness (CCA mean, B = -.34, p = .05), and variability in their perceptions of the group's performance (performance SD, B = -.45, p = .06) were negative predictors of posttest scores. These measures remained significant when included in a joint model (B = -.75, p = .03 for performance SD and B = -.56, p < .001 for CCA mean), demonstrating that they explain unique variance in posttest scores.

We also considered moderation of group-level outcomes by group makeup variables when predicting individual posttest score, but there were no significant interactions.

4 Discussion

4.1 Main Findings

It has long been known that composition of a group effects outcomes [9, 22, 23]. However, what group makeup variables should be measured is not clear. We found that personality dissimilarity negatively predicted how well a team did on the task and also negatively correlated with the team's perceptions of the collaborative process. This suggests that AIED systems that support collaboration should consider personality differences among teammates as part of the *group* model.

It is important that we do not imply causality between personality differences, task score, and perceptions of the collaboration. Teams that had similar personalities might have performed well on the task, and thus perceived their collaboration more positively. Alternatively, they could have perceived themselves as working well together, which might have resulted in enhanced task performance. There might even be a more tightly coupled relationship between task performance and subjective perceptions, where each boosts the other as teams sustain positive interactions and make progress towards the goal.

Gender-based group makeup measures were not predictive of any outcomes, despite some previous work suggesting a link [9]. It is possible that specific factors related to gender, such as expectations of social behavior [36], confidence in computing ability [37, 38], and perceptions of stereotype threat [39] might be more informative measures of group makeup than gender itself.

Prior research suggests teams should maintain common ground by constructing solutions jointly and working towards a shared goal, referred to as shared task alignment [40]. Thus, it is unsurprising that variability in students' perceptions of team performance negatively predicted individual posttest scores.

We additionally found that positive perceptions of the group's communication, cooperation, and agreeableness was negatively related to individual posttest scores. Although this finding might seem counterintuitive, it is possible that it could reflect group work norms, which should be task-oriented rather than social [24]. Groups with more socially-oriented group work norms, such as minimizing conflict and agreeability, might have achieved favorable subjective outcomes at the expense of learning.

Because objective task scores were unrelated to individual learning outcomes, it might not be enough to simply support successful task performance and subjective collaboration outcomes. Instead, a range of CPS goals (e.g. positive group perceptions, task performance, group to individual transfer of learning) should be considered when designing intelligent systems that support CPS.

4.2 Applications

This work can inform future AIED systems that aim to support CPS, with particular emphasis on virtual collaborations. For example, measurements of group makeup (specifically personality differences) could inform sensitivity to unproductive group interactions. Perhaps teams with highly dissimilar personalities should encounter interventions sooner and more often than teams with similar personalities.

Given the differences between group and individual learning outcomes, intelligent systems that support collaboration could be customizable to the outcome pertinent to the learning goals. For example, if successful task completion is the goal, an intelligent system might emphasize subjective outcomes by providing real-time feedback on how to improve collaborative processes if students are not communicating effectively or are being uncooperative and disagreeable. Alternatively, if individual learning is the goal, the intelligent system might encourage coordination amongst team members and task-oriented group work norms by prompting students to build on a particular idea or voice disagreements and discuss tradeoffs between competing solutions.

4.3 Limitations and Future Work

Like all studies, ours has limitations. First our sample size was modest and consisted of students from a private university with little age or ethnic diversity. Further, the data was collected in a laboratory context, thereby limiting broader claims of generalizability. This should be addressed using more diverse and larger samples collected in more authentic context. Indeed, we are in the process of collecting data from more diverse samples both in the lab and in classrooms and with multiple CPS environments.

Our work considers a holistic view of the collaboration, and not how it unfolds over time. Analysis of the collaboration at more fine-grained intervals would provide insight into how perceptions, task performance, and learning change as students interact over time. We are also conducting expert coding of the collaborative process as a complement to the subjective assessments.

Finally, interventions based on our findings should be explored. We do not precisely know the most effective way to respond to unproductive teams. Research is needed to explore how to support aspects of the collaborative processes such as communication cooperation, and agreeableness, as well as instilling task-related group work norms.

4.4 Conclusion

Our work is an initial step in building AIED systems that intelligently support productive collaborative problem solving in groups. We found that group makeup with regards to personality dissimilarity influenced group-level task performance and subjective perception outcomes of the CPS process. Further, outcomes related to successful task performance were different from those related to individual learning, suggesting there might be different processes at play. Thus AIED systems that support CPS should consider both types of outcomes as well as the influence of group makeup in supporting effective collaborations.

5 Acknowledgements

We would like to thank David Blair, Eugene Choi, Mae Raab, and Samantha Scaglione for their invaluable help. This research was supported by the National Science Foundation (NSF DUE-1745442) and the Institute of Educational Sciences (IES R305A170432). Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies.

6 References

1. PISA 2015 Collaborative Problem Solving Framework. (2015).
2. Dillenbourg, P.: What do you mean by collaborative learning? Collaborative-learning Cognitive and Computational Approaches. 1, 1–15 (1999).
3. Kerr, N.L., Tindale, R.S.: Group performance and decision making. Annual

Review of Psychology. 55, 623–655 (2004).
4. Du Boulay, B.: Artificial Intelligence as an Effective Classroom Assistant. IEEE Intelligent Systems. 31, 76–81 (2016).
5. Lou, Y., Abrami, P.C., D'Apollonia, S.: Small group and individual learning with technology: A meta-analysis. Review of Educational Research. 71, 449–521 (2001).
6. Vijayaratnam, P.: Developing higher order thinking skills and team commitment via group problem solving: A bridge to the real world. Procedia-Social and Behavioral Sciences. 66, 53–63 (2012).
7. Alavi, M.: Computer-mediated collaborative learning: An empirical evaluation. MIS Quarterly. 18, 159–174 (1994).
8. Borghoff, U.M., Schlichter, J.H.: Computer-supported cooperative work. In: Computer-Supported Cooperative Work. pp. 87–141. Springer (2000).
9. Fenwick, G.D., Neal, D.J.: Effect of gender composition on group performance. Gender, Work & Organization. 8, 205–225 (2001).
10. Savicki, V., Kelley, M., Lingenfelter, D.: Gender and group composition in small task groups using computer-mediated communication. Computers in Human Behavior. 12, 209–224 (1996).
11. Gruenfeld, D.H., Mannix, E.A., Williams, K.Y., Neale, M.A.: Group composition and decision making: How member familiarity and information distribution affect process and performance. Organizational Behavior and Human Decision Processes. 67, 1–15 (1996).
12. McManus, M.M., Aiken, R.M.: Supporting effective collaboration: Using a rearview mirror to look forward. International Journal of Artificial Intelligence in Education. 26, 365–377 (2016).
13. Rosen, Y.: Computer-based assessment of collaborative problem solving: Exploring the feasibility of human-to-agent approach. International Journal of Artificial Intelligence in Education. 25, 380–406 (2015).
14. Linnenbrink-Garcia, L., Rogat, T.K., Koskey, K.L.K.K.: Affect and engagement during small group instruction. Contemporary Educational Psychology. 36, 13–24 (2011).
15. Tenenberg, J., Roth, W.-M., Socha, D.: From I-awareness to we-awareness in CSCW. Computer Supported Cooperative Work (CSCW). 25, 235–278 (2016).
16. Chang, C.-J., Chang, M.-H., Chiu, B.-C., Liu, C.-C., Fan Chiang, S.-H., Wen, C.-T., Hwang, F.-K., Wu, Y.-T., Chao, P.-Y., Lai, C.-H., Wu, S.-W., Chang, C.-K., Chen, W.: An analysis of student collaborative problem solving activities mediated by collaborative simulations. Computers & Education. 114, 222–235 (2017).
17. Crossley, S., Dascalu, M., McNamara, D.S., Baker, R., Trausan-Matu, S.: Predicting success in massive open online courses (MOOCs) using cohesion network analysis. In: International Conference on Computer-Supported Collaborative Learning. Philadelphia, PA: International Society of the Learning Sciences. (2017).
18. Yoo, J., Kim, J.: Can online discussion participation predict group project performance? Investigating the roles of linguistic features and participation

patterns. International Journal of Artificial Intelligence in Education. 24, 8–32 (2014).
19. Adamson, D., Dyke, G., Jang, H., Rosé, C.P.: Towards an agile approach to adapting dynamic collaboration support to student needs. International Journal of Artificial Intelligence in Education. 24, 92–124 (2014).
20. Hmelo-Silver, C.E., Barrows, H.S.: Facilitating Collaborative Knowledge Building. Cognition and Instruction. 26, 48–94 (2008).
21. Olivera, F., Straus, S.G.: Group-to-individual transfer of learning: Cognitive and social factors. Small Group Research. 35, 440–465 (2004).
22. Wilkinson, I.A.G.G., Fung, I.Y.Y.Y.: Small-group composition and peer effects. International Journal of Educational Research. 37, 425–447 (2002).
23. Cummings, J.N.: Work groups, structural diversity, and knowledge sharing in a global organization. Management Science. 50, 352–364 (2004).
24. Langfred, C.W.: Is group cohesiveness a double-edged sword? An investigation of the effects of cohesiveness on performance. Small Group Research. 29, 124–143 (1998).
25. Gully, S.M., Devine, D.J., Whitney, D.J.: A meta-analysis of cohesion and performance: Effects of level of analysis and task interdependence. Small Group Research. 26, 497–520 (1995).
26. Evans, C.R., Dion, K.L.: Group cohesion and performance: A meta-analysis. Small Group Research. 22, 175–186 (1991).
27. Harrison, D.A., Price, K.H., Bell, M.P.: Beyond relational demography: Time and the effects of surface-and deep-level diversity on work group cohesion. Academy of Management Journal. 41, 96–107 (1998).
28. Fernández Dobao, A., Blum, A., Dobao, A.F., Blum, A.: Collaborative writing in pairs and small groups: Learners' attitudes and perceptions. System. 41, 365–378 (2013).
29. Barsade, S.G., Gibson, D.E.: Group affect: Its influence on individual and group outcomes. Current Directions in Psychological Science. 21, 119–123 (2012).
30. Ku, H.-Y.Y., Tseng, H.W., Akarasriworn, C.: Collaboration factors, teamwork satisfaction, and student attitudes toward online collaborative learning. Computers in Human Behavior. 29, 922–929 (2013).
31. Behfar, K.J., Mannix, E.A., Peterson, R.S., Trochim, W.M.: Conflict in small groups: The meaning and consequences of process conflict. Small Group Research. 42, 127–176 (2011).
32. Pai, H.-H.H., Sears, D.A., Maeda, Y.: Effects of small-group learning on transfer: A meta-analysis. Educational Psychology Review. 27, 79–102 (2015).
33. Code Studio, https://studio.code.org/s/mc/stage/1/puzzle/1.
34. Fraser, N.: Ten things we've learned from Blockly. In: Proceedings of the 2015 IEEE Blocks and Beyond Workshop. pp. 49–50 (2015).
35. Gosling, S.D., Rentfrow, P.J., Swann, W.B.: A very brief measure of the Big-Five personality domains. Journal of Research in Personality. 37, 504–528 (2003).
36. Stake, J.E., Eisele, H.: Gender and personality. In: Handbook of Gender

Research in Psychology. pp. 19–40. Springer (2010).
37. Shashaani, L.: Gender differences in computer attitudes and use among college students. Journal of Educational Computing Research. 16, 37–51 (1997).
38. Beyer, S., Rynes, K., Perrault, J., Hay, K., Haller, S.: Gender differences in computer science students. ACM SIGCSE Bulletin. 35, 49–53 (2003).
39. Spencer, S.J., Steele, C.M., Quinn, D.M.: Stereotype threat and women's math performance. Journal of Experimental Social Psychology. 35, 4–28 (1999).
40. Barron, B.: Achieving coordination in collaborative problem-solving groups. Journal of the Learning Sciences. 9, 403–436 (2000).