

Unpacking Collaborative Sense-Making: The Role of Reflective Accuracy in Collaborative Process Quality

Tugce Aldemir and Marcela Borge
tfa5065@psu.edu, mbs15@psu.edu
Pennsylvania State University

Abstract: In this paper, we build upon a series of studies that aimed to explore the dynamics between socio-metacognition (knowledge about and ability to regulate team cognition) and the ability of teams to improve the quality of collaborative discussions. In these studies, discussion teams in an online course are provided with open-ended discussion questions about course content to discuss. Then, individuals score the quality of the team's collaborative processes on six items, followed by entire team reviewing their scores and devising a plan for improvement. In this paper, we examine the relationship between team reflective accuracy, collaborative sense-making, and collaborative process quality for 12 teams over five time points. Findings show a significant correlation between the reflective accuracy and collaborative process quality. A further comparative case analysis shows the underlying patterns of collaborative sense-making and reflective accuracy that impact team process quality.

Introduction

Collaborative competence has become one of the most demanded skills in work and educational contexts. However, collaborative activities in Computer-Supported Collaborative Learning (CSCL) contexts do not always yield positive outcomes. One of the reasons behind the discrepancy between ideal and observed outcomes in CSCL is that individuals do not have opportunities to develop competencies necessary to engage in high quality collaborative processes (Kozlowski & Ilgen, 2006). Collaborative competence is defined as the ability to carry out complex forms of intersubjective knowledge construction and collective sense making (Borge, Ong, & Rosé, 2018; Stahl, 2006). Providing a digital environment to bring learners together is not enough to prompt these processes (Fischer et al., 2002). Such a complex and nested forms of cognition require individuals to develop a clear understanding of what a high-quality collaborative activity looks like, and to monitor and regulate individual and collective thinking processes during collaborative activities (Järvelä & Hadwin, 2013). We refer this form of metacognition at the team level as socio-metacognition. A growing number of researchers are examining how to leverage technological interventions to increase socio-metacognition to prompt individual and shared regulatory behaviors in collaborative activities (e.g. Phielix et al., 2010; Järvelä, & Hadwin, 2013). There is a good number of studies supporting the effectiveness of these tools (e.g. Panadero, Tapia, & Huertas, 2012). However, these findings are limited in their ability to explain whether and to what extent such tools improve quality of collaborative process, and collective regulation (Panadero et al., 2015; Phielix et al., 2010).

In our previous works, we developed a theoretically informed technological intervention to help teams improve their socio-metacognitive expertise so they could regulate and improve their own collaborative processes. Our preliminary findings suggested that teams were able to improve the quality of their collaborative activity over time, but they needed additional support at self-assessing the quality of their collaborative processes as they tended to overrate their process quality. Given the critical role self-assessment plays in regulatory behaviors (Harris & Brown, 2013), these findings call for further investigation at a micro-analytic level to understand the dynamics between self-assessment processes and team process quality over time. In-depth analysis of teams' socio-metacognitive sense-making patterns as they collectively make sense of and reflect on their collaborative discussions can help the field to better understand why teams may not improve their team reflective accuracy, but may still improve the quality of their collaborative processes over time. Thus, in this study, we explore the dynamics between teams' process quality (PQ) during content-based discussion and the team's ability to accurately assess the quality of their team's collaborative processes after their discussion (reflective accuracy). We ask, (RQ1) What is the relationship between team reflective accuracy and Team PQ score for low, medium, and high PQ teams? and (RQ2) what differences exist in the communication patterns of teams with high and low PQ levels?

Methods

Participants and procedure

The data were collected from a 16-week online undergraduate course on information sciences. Thirty-four online students participated in the study (33.3% female, 66.7% male). Students were divided into teams (10 triads, 2 dyads) based on availability for online synchronous meetings, and required to participate in online synchronous meetings five times over the semester. Before each discussion, they were provided with readings and questions about the readings, and asked to write individual reflections addressing those questions. During 60-minute discussions, they shared their responses to the reflection questions. After the discussion, students used the system to engage in individual sense-making about the team's process for 15 minutes by scoring the team's process quality (PQ) on six items and justifying their scores with written text. Then team members participated in a 15-minute collective sense-making discussion where they shared their PQ scores, diagnosed their process problems, and selected strategies provided by the system, to address problems in future discussions.

Collaborative process quality (PQ)

Collaborative process quality (PQ) was scored using a previously established assessment (Borge et al., 2018). This assessment operationalized six communication behaviors associated with high-quality information synthesis and knowledge negotiation. Categories of behaviors associated with Information synthesis included: Verbal Equity (to what extent team members contributed equally), Developing Joint Understanding (to what extent team members make an effort to ensure that they fully understand the ideas), and Joint Idea Building (to what extent team members elaborate on others' contributions). Categories of behaviors associated with knowledge negotiation included: Exploration of Different Perspectives (to what extent teams present and discuss alternative perspectives), Quality of Claims (to what extent teams provide logical and fact-based evidence to their claim), and Norms of Evaluation (to what extent teams adhere to social norms). Each item was scored on a range from 1 to 5, where 5 included evidence of desired patterns of communication and 1 included evidence of highly dysfunctional patterns. Using the same rubric that students used, a trained graduate student (expert scorer) analyzed the 60-minute content-based discussions for each of the six items across the five time-points. Another trained student analyzed the 20% of the data to ensure inter-reliability, $r = .86, p < .001$; Kappa = .64, $p < .001$. All six category scores were averaged to determine mean process quality score for each team. These scores were then used to categorize the teams by process quality level, i.e. low PQ, medium PQ, and high PQ teams.

Socio-metacognitive sense-making

Socio-metacognitive sense-making (SMS) talk can occur during the 15-minute collective sense-making discussion when teams share their PQ scores and figure out how to improve future discussions. To analyze socio-metacognitive sense-making, we coded these discussions for all 5 timepoints at the turn level. Each turn was coded as either productive SMS talk or "other talk" and then calculated percentages of SMS talks to normalize data. SMS talk includes patterns described by Borge et al. (2018): Reporting (sharing PQ scores without justification), Process Monitoring (pointing out a specific activity/requirement that the team has done or needs to do), Process Reflection (pointing out a reason for why an incident happened), Process Planning (unpacking collaborative problem and proposing new goal), and Process Revising (reconsidering their activity based on new information). Inter-rater reliability for 20% initial coding was Kappa = .79, $p < .001$. Other talk includes off-task talk not related to sensemaking activity (e.g. talking about an upcoming exam).

Team reflective accuracy

Reflective accuracy is defined as team PQ scores that align with expert PQ scores. To calculate Team reflective accuracy, we first averaged individual PQ scores to create a team PQ score. We subtracted the team PQ score from the expert score: the closer it is to zero, the better the team reflective accuracy is. Negative scores denote over estimation and positive scores denote under-estimation of PQ score.

Results

(RQ1) What is the relationship between team reflective accuracy and Team PQ score for low, medium, and high PQ teams?

We found a strong negative correlation between team reflective accuracy and PQ, $r = -0.734, n=60, p<0.01$, meaning that teams with higher quality content-based discussions were also more accurate assessors of their process quality. When we compared the team reflective accuracy average scores for all five discussions with PQ levels (high, med, and low) and we found a significant difference between levels, Welch's $F(2, 35.38) = 9.46, p=0.01$. Further Games-Howell post hoc comparisons indicated that the mean accuracy of high PQ teams ($M=1.71, SD=1.58$) was significantly higher than the mean accuracy of the low PQ teams ($M=5.18, SD=3.34$),

$p=0.01$, while there were no significant differences between the medium PQ teams ($M=3.06$, $SD=1.58$) and the other two PQ categories.

To ensure that these differences were not due to the changes in expert scores over time, or a ceiling effect for high PQ teams, we further analyzed the correlation between expert and team scores for different PQ levels, and found a positive, moderate correlation between mean team and expert scores for high PQ teams, $r = -0.595$, $n=20$, $p=0.06$. In contrast, no significant relationship was found for low PQ teams, $p=.200$. These results suggest high PQ teams better calibrated their Team PQ scores to expert scores over time.

(RQ2) What differences exist in the patterns of team sense-making during process-reflection in teams with high and low PQ?

Even though all teams engaged in productive sense-making in their collective sense-making discussions, high PQ teams had significantly more productive SMS talk ($M=52.13$, $SD=13.12$) than low PQ teams ($M=31.38$, $SD=16.50$), $F(1,30)=15.494$, $p=.00$. When we examined what types of SMS talk differed between PQ levels, we found Planning talk was significantly higher in high PQ teams ($M=21.00$, $SD=13.31$) than low PQ teams ($M=6.67$, $SD=11.02$), $F(1,30)=10.990$, $p=.00$. These quantitative differences in the reflective discussions presented themselves qualitatively as more sophisticated and collectivist gap analysis and strategy identification patterns between the members of high PQ teams. For example, accurately diagnosing the collaborative problems would require collective efforts to correct such misconceptions; however, only 10% of the problems recognized by individual members of low PQ teams were discussed as a joint attempt to conceptualize the underlying factors of the problems, while 39.39 % of the problems recognized by the individual members of the high PQ teams were discussed.

Table 1 shows a short excerpt from two teams who discussed alternative perspectives as their area of weakness. The low PQ team did not take time to analyze their existing patterns, nor did they discuss concrete strategies they could use to improve. Even though they accurately recognized their weakness, they struggled at accurate appraisal of it as they listed external factors as the underlying causes of the problem. The high PQ team, on the other hand, takes time to justify why it is important to address the problem they correctly identified, and suggest concrete strategies that they can use to improve.

Table 1: Sample problem diagnosis and strategy identification patterns

Team 11, Session 1, Low PQ team	Team 2, Session 1, High PQ team
11b: <i>One weakness I thought we have is alternative ideas and that's not really a fault of ours but more that there is only 2 of us so not as many thoughts and ideas are brought to the table versus a group of say 4.</i>	2a: <i>We just need to be more critical on each other's judgement and be holistic// try and think of all perspectives</i>
11a: <i>I agree we need to challenge each other more with different ways of thinking.</i>	2b: <i>Yes, I agree even though it may be hard</i>
11b: <i>with 2 people it's not really a "group" discussion and more of a "chat" I guess you could say // but I agree we should challenge more</i>	2c: <i>Yes. I wrote that as one our potential areas to improve as well // that was we all grow as students // being able to play devil's advocate is sometimes a necessary thing</i>
11a: <i>Yes, but we can still bring a different perspective to the table.</i>	2a: <i>Yes, in the case of the topics we are covering I really think it is applicable.</i>
11b: <i>the most important thing of all that is killing us is not having a good constant contact through the week. (Team started to discuss an external problem not directly related to their discussion)</i>	2b: <i>I feel we should try to constructively challenge each other in future chats to bring out more views and opinions.</i>
	2a: <i>It doesn't seem like always the morally right thing to do at least for me, but it must be done in order to touch al bases</i>
	...
	2c: <i>yes. that way it'll force us to consider all options (even if we don't necessarily agree) (shortened as the conversation goes on about why it is important to address that problem and how the team can address it)</i>

Discussion

In this paper, we examined the relationship between the quality of teams' collaborative course content-based discussions (PQ) and the team's ability to accurately assess their own process quality (reflective accuracy). We

identified differences in individual and group sense-making patterns during process-reflection in teams with high and low PQ. Our results showed a significant relationship between team reflective accuracy and PQ, with a significant difference between high and low PQ groups. We also found that teams who had higher content-based process quality (high PQ teams) engaged in higher quality socio-metacognitive sense-making (SMS) talk than teams that displayed lower content-based process quality (low PQ teams). Low PQ teams spent most of their reflection time discussing non-related topics. One reason why low PQ teams did not allocate sufficient time for productive SMS talk might be that they had unrealistic confidence in their ability to engage in collaborative activity (Alicke & Govorun, 2015). Because while high PQ teams mostly engaged in planning to identify and agree on strategies to correct their process problems, low PQ teams mostly engaged in reporting their individual scores and sharing their opinions (mainly positive) about their performance without any concrete justification or discussion. However, the difference in the number of SMS patterns displayed by different PQ levels does not necessarily suggest that high PQ teams were better at group regulation even though they displayed relatively better self-regulatory behaviors (i.e. better calibration of self-assessment scores (reflective accuracy)). Further research is needed to examine how self-regulatory behaviors are associated with group regulatory behaviors.

The study has some limitations. One is the number of the teams participated in the study. Further research is needed to see whether similar patterns can be observed with different samples. Another limitation is that we did not collect any individual data even though the discrepancies observed between high and low PQ levels might be due to the individual differences (Kruger & Dunning, 1999). Thus, further research examining the individual factors in the group regulatory behaviors and individual accuracy scores is needed.

References

- Alicke, M.D., & Govorun, O. (2005). The better-than-average effect. In Alicke, M. D., Dunning, D., & Krueger, J. (Eds.), *The self in social judgment* (pp. 85-106). New York, NY: Psychology Press.
- Borge, M., Ong, Y. S., & Rosé, C. P. (2018). Learning to monitor and regulate collective thinking processes. *International Journal of Computer-Supported Collaborative Learning*, 13(1), 61-92.
- Fischer, F., Bruhn, J., Gräsel, C., & Mandl, H. (2002). Fostering collaborative knowledge construction with visualization tools. *Learning and Instruction*, 12, 213-232.
- Harris, L. R., & Brown, G. T. (2013). Opportunities and obstacles to consider when using peer-and self-assessment to improve student learning: Case studies into teachers' implementation. *Teaching and Teacher Education*, 36, 101-111.
- Järvelä, S., & Hadwin, A. F. (2013). New frontiers: Regulating learning in CSCL. *Educational Psychologist*, 48(1), 25-39.
- Kozlowski, S., & Ilgen, D. (2006). Enhancing the effectiveness of work groups and teams. *Psychological Science in the Public Interest*, 7(3), 77-124.
- Krueger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121.
- Panadero, E., Kirschner, P. A., Järvelä, S., Malmberg, J., & Järvenoja, H. (2015). How individual self-regulation affects group regulation and performance: A shared regulation intervention. *Small Group Research*, 46(4), 431-454.
- Panadero, E., Tapia, J. A., & Huertas, J. A. (2012). Rubrics and self-assessment scripts effects on self-regulation, learning and self-efficacy in secondary education. *Learning and Individual Differences*, 22(6), 806-813.
- Phielix, C., Prins, F. J., Kirschner, P. A., Erkens, G., & Jaspers, J. (2010). Group awareness of social and cognitive performance in a CSCL environment: Effects of a peer feedback and reflection tool. *Computers in Human Behavior*, 27(3), 1087-1102.
- Stahl, G. (2006). *Group cognition: Computer support for building collaborative knowledge*. Cambridge, MA: MIT Press.