JrnlID 11412_ArtID 9321_Proof# 1 - 17/06/2020

International Journal of Computer-Supported Collaborative Learning
https://doi.org/10.1007/s11412-020-09321-7

Talking about group (but not individual) ProcessX aids group performance

Deanna Kuhn[1] · Noel Capon[1] · Hueiyi Lai[1]

Received: 30 December 2019 / Accepted: 15 June 2020

Abstract
The discourse of small groups of 3–4 adults enrolled in a graduate business course was audio-recorded as they participated in a computer-supported simulation in which the group represented a firm and worked over a series of eight sessions in making a series of decisions. Discourse transcripts were analyzed using a coding scheme that classified utterances expressed during group interaction as types of topic-talk (constituting a part of the activity itself) vs. meta-talk (reflecting on the activity). Supporting our hypothesis regarding the importance of meta-level discourse about group process in a group's achieving coordinated action and a successful outcome, analysis suggested that discourse about the group's process, but not discourse about individuals' actions, was associated with superior group outcomes.

Keywords Group process · Collaboration · Discourse · Metacognition

New and challenging real-world problems are most often confronted collaboratively. They are assumed too difficult for one individual to solve optimally. Yet we are only beginning to make progress in determining what makes collaborative problem solving effective (Graesser et al. 2018), and what are the boundary conditions for its producing better outcomes than an individual working alone. Similarly, strategies for effective collaborative work are not taught as a part of standard curriculum at any age level (Kuhn 2015). Arguably we must await more research on group problem solving before sound, evidence-based curricula can be designed. Meanwhile, today's young people enter into adult work careers and social lives in which collaboration increasingly is both expected and essential, as the complexities of modern life escalate.

In their review of research related to collaborative problem solving, Graesser et al. (2018) attribute the scarcity of research to the labor-intensive nature of observing, recording, coding, and analyzing the interactions of a group of individuals as they work together to address a complex problem. Some studies have responded to the challenge by employing technology to

✉ Deanna Kuhn
dk100@tc.columbia.edu

[1] Columbia University Teachers College Official Bookstore, New York, NY, USA

automate one or more aspects of the research task, including tracking and coding participants' behaviors, and even to the extent of substituting an automated rather than human other with whom a participant interacts (Graesser et al. 2018) or providing other kinds of automated supports (Vogel, Wecker, Kollar, & Fischer, 2017), leaving generalizability to a naturalistic interpersonal context a resulting concern. Q1

The study of collaborative cognition in naturalistic human settings is especially complex, even when we confine the domain of study to individuals collaborating to solve a new problem (rather than working together to achieve mastery of a predetermined body of knowledge). Potential factors influencing group success in achieving an acceptable solution to the problem fall into two broad categories, attributes of the individual members of the group and attributes of the group as a whole. Individual factors can be subdivided into cognitive and social-personality ones. Group factors are also of two types. One is patterns of relationship among individual factors, such as whether group members are of similar or mixed cognitive abilities or personality types. The other, and most complex to examine, is the interaction that characterizes the group in action. Those who have undertaken the task have found painstaking analysis necessary to observe how patterns of collaboration emerge, develop and characterize a group's functioning at a level beyond that of the behavior of individual participants (Graesser et al. 2018; Jacobson et al. 2016; Järvelä et al. 2016; Kapur 2008; Siqin et al. 2015).

What does a group need to do in order to execute a problem-solving task more successfully than would one of them working alone? Sloman and Rabb (2019) put the matter succinctly in noting that "… humans operate within a division of cognitive labor: Each individual brings a fairly narrow expertise to bear on issues, and communities combine these narrow areas of expertise to create a much broader and richer database of skills and knowledge …" Yet the coordination of these individual resources remains to be achieved. How do group members put their respective capabilities to work jointly in an effective manner?

It is such patterns of interaction we undertake to examine in the present work, with a particular focus on participants' metacognitive regulation. In an earlier initial investigation of individual factors in the same sample of small groups examined here, Kuhn and Modrek (2018) investigated a particular individual cognitive characteristic that participants brought to the activity – the extent to which a participant was assessed as possessing mental models that entailed multivariable causality, i.e., ones recognizing multiple contributory factors in order to account for a phenomenon (Kuhn et al. 2015; Kuhn 2019). If it is a complex problem being addressed, multiple factors are almost always necessary to consider. It was therefore predicted that this cognitive competency was a necessary individual prerequisite to effective group outcomes (Osiurak and Reynaud 2019), a prediction the results supported. In the absence of at least two group members sharing this individual cognitive competency, a group did less well.

In the more labor-intensive exploratory study presented here, we examine collaborative problem solving at its most complex level, the collaborative process itself, following Shea et al. (2014) in proposing metacognitive processes to play a key role. These can be defined as the representation, monitoring, and management of cognitive processes. Metacognitive skills emerge early in life and continue to develop (Flavell 1979; Kuhn 2000), with some individuals achieving greater proficiency than others. To aid at a group level, meta-level understandings must not only exist but must be conveyed to others in the group. Shea et al. emphasize individuals' potential to share their metacognitive understandings with others, in what these authors refer to as "supra-personal cognitive control." In doing so, these understandings potentially influence others involved in a shared task and thereby enhance complex forms of coordinated action (Järvelä et al. 2016).

There exists some research suggestive of the possibility that more frequent or more effective metacognition is associated with better group outcomes (Duhigg 2016; Hogan et al. 2016; Morris et al. 2019; Pifarre and Cobos 2010; Slof et al. 2012; Yoon et al. 2018). Here we investigate this possibility directly, not by employing individual metacognitive skills as predictors but rather by observing group discourse directly, with a particular focus on its metacognitive dimensions. Meta-level utterances are defined as those that reflect on the activity, rather than constituting a part of the activity itself and addressing the task subject matter. We further divide meta-level utterances into Meta-self and Meta-group categories. In addition to our core hypothesis that the prevalence and nature of meta-level talk will be predictive of group productivity, we advance the further hypothesis that it is Meta-group talk that will most benefit coordinated action and hence group performance.

Method

Participants

Participants were 35 students (16 female) in one section of an Executive MBA program at a major US graduate school of business. All were enrollees in a capstone market strategy course taught by one of the authors. EMBA students all have prior experience in positions in the business or non-profit world, and the large majority were continuing their professional employment while attending the program part-time. All held at least bachelor's degrees and many had earned other post-graduate degrees prior to entering the EMBA program. Their ages ranged from mid-20s to early 40s.

Design

Small groups of three or four participants worked over a series of eight sessions. We chose for analysis both an early and a late session, since early tasks, such as establishing shared understandings, may entail different processes and patterns of interaction than later ones, such as reaching joint conclusions. The groups were composed of students enrolled in a graduate business course that featured a simulation in which groups were required to work together as a firm in making a series of decisions. Participants had varying amounts of previous job experience that required team work, but they had not worked together previously. Within their degree program all participants had substantial experience doing collaborative work.

Motivation for groups to perform well in the simulation was high, as this performance contributed heavily to course grade.

Procedure

The major component of the course was the *Markstrat* simulation (https://web.stratxsimulations.com/simulation/strategic-marketing-simulation). In the simulation, each randomly-assigned student team of three or four represents a firm that competes against four other firms in its industry (represented by other student teams). The present group operated on only one industry, but the simulation has the capacity to allow course groups to operate on multiple industries. The computer simulation differs from many computer supported learning aids in not being designed to scaffold students' activities but rather to provide the learning environment. Some analytic tools, however,

JrnlID 11412_ArtID 9321_Proof#1 - 17/06/2020

were embedded in the simulation, e.g., a tool to perform regression analyses on consumer responses that the simulation generated from groups' input.

The starting positions of firms (groups) are different, but all firms are roughly equivalent in terms of strengths, weaknesses, opportunities, and threats. Each firm makes decisions over eight periods (equivalent to business years). Each firm secures marketing research data, assesses likely competitor reactions to its potential moves, and makes marketing decisions involving both strategy and implementation. Because the simulation extends across eight periods, teams can measure results – sales, market share, profit contribution, share price – that follow from their decisions, and accordingly evolve objectives, strategies, and implementation plans from period to period.

At the beginning of the course, each student gave signed consent to having their small-group discussion sessions audio-recorded for research purposes. During eight periods of decision-making, groups accomplished one or two decisions each day across the five full days that constituted the duration of the course. The times groups took to examine the last period's results, collaborate, and make new decisions varied across the eight periods. On average, groups spent 1.5 h of discussion per period. Once the simulation concluded, each group prepared a class presentation intended to address objectives, strategies, implementation programs, and performance and to identify key lessons learned. Presentations were graded and contributed to the final course grade.

Questionnaires were distributed to each participant to complete individually at three points in time: early in the simulation (at the end of the second decision period, period 2), late in the simulation (at the end of the second from last decision period, period 7), and immediately after the final decision (period 8). The questionnaire designed by the authors contained six questions pertaining to how well the group worked together and the relative contributions of individual members. Completion rates were 100%, 100% and 94% for the first, second, and final questionnaires respectively.

Coding of group interaction

Coding categories appear in Table 1. A key distinction the coding scheme makes is that between statements addressing the subject matter and meta-level statements referring to the discourse itself. Meta-talk categories are further divided into meta-talk about the speaker him/herself ("Meta-Self") and meta-talk referring to one or more members of the group or the group as a whole ("Meta-Group"). Topic-talk utterances were analyzed using a coding scheme established by one of the authors and colleagues for classifying dialogic moves in argumentative discourse. This scheme has been used in numerous previous studies on argumentation (see Rapanta et al. 2013, for review). Classification is based on the function of an utterance in relation to the utterance immediately preceding it. A rationale for employing this scheme is the anticipation that it is this relational function that is key to the coordinated action central to group process (Kuhn et al. 2013; Macagno 2016).

Transcripts were segmented into individual utterances by the first and second author with the few disagreements resolved by discussion, with each utterance coded blind to group, individual speaking, and period. To establish coding reliability, a total of 1783 units (16% of the entire database) were coded independently by an author and another blind coder. Independent coding by the two trained coders was above 90% for segmenting (Cohen's kappa $\kappa = .947$).and above 80% for assignment to category (Cohen's kappa = .805). Differences were resolved by discussion, and remaining coding was performed by one of the coders.

JrnlID 11412_ArtID 9321_Proof# 1 - 17/06/2020

t1.1 **Table 1** Coding categories for group interactions

t1.2 Meta-Talk	Definition	Sample utterances
t1.3 Meta-Self	An utterance that relates to self, rather than the subject matter of the discussion	"I am very concerned what the R&D portfolio is." "It just does not make sense to me." "Okay so basically I come to compare our brand awareness by consumer segment."
t1.4 Meta-Group	An utterance that relates to the group's discussion itself, rather than the subject matter of the discussion	"So then if we go back to R&D." "And we see, the feasibility and the R&D." "Because the question we need to answer for the research part."
t1.5 Topic-Talk	Definition	Sample utterance
t1.6 Add	An addition to preceding utterance	"So, here's another thing."
t1.7 Agree?	A question asking whether other will accept or agree with a claim	"Do you want to keep the base cost the same?"
t1.8 Agree	A statement of agreement with the other's preceding assertion	"Yes, that is what I thought."
t1.9 Claim	An utterance that asserts something	"It is showing this is from the production design."
t1.10 Clarify?	A request for the other to clarify a proximal utterance	"Okay you are not doing R&D and stuff right?"
t1.11 Clarify	A clarification of the speaker's position or argument in response to the other's immediately preceding utterance	"Yeah, we need to focus on branding."
t1.12 Counter	A disagreement with the other's immediately preceding utterance	"No, last time was 1500."
t1.13 Cut Off	An utterance to interrupt	"But wait."
t1.14 Strategy-Suggestion	An utterance to suggest a strategy	"We can also introduce Sonite; we have high earners and professionals."
t1.15 Question	An informational query that does not refer to a proximal utterance	"How do you define perceptual message?"

Results

The course instructor made available to the researchers each group's score on the key performance variable, the final stock price index (SPI) that the group's firm had attained by the end of the simulation. An index of 2000 or higher reflects strong performance, while scores of 1500 or below reflect weak performance. Other measures of a firm's final performance exist, but because these were highly correlated with SPI, only SPI is used in the present analysis.

Given the exploratory nature of this research, we sought first of all and primarily to identify differences in group process associated with the strongest and weakest group performance outcomes. For this purpose, we chose for initial comparison the highest performing of the ten small groups (defined by SPI) and the lowest performing group. For each of these two groups, two early group sessions (Periods 1 and 2) and two later group sessions (Periods 7 and 8) were examined.

Comparison of highest and lowest performing groups

As seen in Tables 2 and 3, the highest performing and lowest performing groups differed markedly at all four periods examined with respect to the duration of the discussion for that

t2.1 **Table 2** Number of minutes of discourse time of the highest and lowest performing groups

t2.2		PERIOD 1	PERIOD 2	*PERIOD 7*	*PERIOD 8*
t2.3	Highest-Performing	120	163	99	73
t2.4	Lowest-Performing	96	64	58	40

period and the number of utterance units it contained, with the longer durations and larger number of units shown by the highest performing group in all cases. As a result, category usage is subsequently examined by proportion rather than only frequency.

The proportion usage of each of the discourse types by the highest and lowest performing groups are presented in the following tables, for each of the four decision periods examined (Tables 4 through 7) and in summary form across all four periods (Table 8). As summarized in Table 8, the two Meta categories (Meta-Group and Meta-Self) are the only categories to distinguish the highest and lowest performing groups consistently, with the association in opposite directions for the two categories. The Agree and Counter categories distinguish the groups at more than one period, but neither does so consistently (Table 8).

Repeated-measures analyses of variance (ANOVAs) were conducted to investigate the possibility of change over time. These were carried out separately for the proportions of use of "Agree," "Agree Question," "Counter," "Meta-Group," and "Meta-Self" within the high-performing and low-performing groups separately. For the low-performing group, tests of normality revealed these four categories to be approximately normally distributed. Mauchly's Test of Sphericity reveals that the assumption of sphericity holds. The analysis showed no significant effect of decision period for any of the five utterance categories, $F(4, 10) = .149$, $p = .719$. Similarly, for the high-performing group, tests of normality and sphericity were normal as expected and no effect of decision period appeared.

A qualitative examination of all Meta-group statements suggested they were of four major types. The first consists of a speaker's effort to understand another group member or members. A second involves taking stock and defining the group's position. A third consists of planning a particular next step. A fourth consists of evaluating the group's work. Table 9 contains examples of each of the four types, taken verbatim from group transcripts.

Full-sample analysis

Given the converging evidence, in particular regarding the differential use of meta-level discourse categories between highest- and lowest-performing groups, in a further analysis we examined the extent to which this pattern would hold among the full sample that included middle-level-performing groups. We thus randomly selected 10-min segments of each of the ten groups' discussion for Periods 2 and 7 (to represent early and late stages of decision cycle), to investigate the association between a groups' category usage during the discussion and

t3.1 **Table 3** Number of utterances produced by the highest and lowest performing groups

t3.2		PERIOD 1	PERIOD 2	*PERIOD 7*	*PERIOD 8*
t3.3	Highest-Performing	1068	1137	955	781
t3.4	Lowest-Performing	773	646	305	341

group performance. For period 2, mean discussion time was 116.1 min (range 64–163). For period 7, mean discussion time was 62 min (range 14–113). Both of these means were intermediate between the times for the lowest and the highest performing groups for these respective periods. With the duration of randomly selected segments of session 2 and of session 7 held constant at 10 min, there was not a great range in number of utterances within the 10 min segments, with a range across groups from 99 to 204 for period 2 and a similar range of 95–198 for Period 7.

Given there are no longer length differences across groups, as in the two-group analysis, we can avoid conversion to percentages and instead directly examine frequencies across the two periods 2 and 7 for different types of units. We did this for each of the discourse categories in Table 1 that distinguished the highest from the lowest performing groups in the earlier analysis, but with a particular focus on the Meta-talk categories hypothesized to play a role in group performance. Results showed associations with performance outcomes for both of these categories – most prominently for the Meta-group and to a lesser extent the Meta-self categories. Qualitative inspection reveals the highest three performing groups (based on final SPI) all showed a combined (periods 2 and 7) frequency of Meta-group utterances of greater than 40. By comparison, of the remaining seven groups, less than half (3 of 7) reached a frequency of 40. Quantitative treatment of these data shows significant correlations of .76 for Period 2 and .70 for Period 7 between proportions of Meta-group usage and groups' final SPI performance score. In contrast, the correlation between Meta-self proportion and performance is negative, specifically with a significant −.76 for Period 7 and nonsignificant −.45 for Period 2. The small number of cases, however, means that these correlations must be treated only as suggestive. No other correlations between discourse categories and performance achieved significance.

The patterns across groups are shown in Fig. 1 for the Meta-group category and Fig. 2 for the Meta-self category. Worthy of note in these figures is the fact that incidence is fairly stable across the two time periods, at least for nine of the ten groups, suggesting that these are relatively stable characteristics of a group across time. Also, finally, as seen in Fig. 1, Meta-group incidence appear to increase over time among the majority of groups, suggesting they

t4.1 **Table 4** Period 1 proportion usage of discourse types by highest and lowest performing groups

t4.2	Meta-Talk	Highest-Performing	Lowest-Performing	$\chi^2(1)$	*p value*	*Bonferroni adjusted alpha*
t4.3	Meta-Self	8.1%	15.8%	34.21	< .00001*	.004
t4.4	Meta-Group	16.2%	8.2%	33.69	< .00001*	.004
t4.5	Topic-Talk	Highest-Performing	Lowest-Performing	$\chi^2(1)$	*p value*	*Bonferroni adjusted alpha*
t4.6	Add	7.7%	6.9%	0.51	0.4768	.004
t4.7	Agree?	2.7%	1.8%	1.30	0.2538	.004
t4.8	Agree	25%	15.9%	37.18	< .00001*	.004
t4.9	Claim	15.5%	17.6%	1.87	0.1712	.004
t4.10	Clarify?	6.6%	6.6%	< .01	1	.004
t4.11	Clarify	4.8%	6.0%	1.15	0.2831	.004
t4.12	Counter	1.4%	5.3%	23.43	0.00001*	.004
t4.13	Cut Off	0.9%	3.6%	15.43	0.00009*	.004
t4.14	Strategy-Suggestion	7.9%	7.2%	0.20	0.6535	.004
t4.15	Question	3.1%	5.2%	5.02	0.0251	.004

Ns are shown in Table 3. During Period 1, the highest performing group produced 1068 utterances while the lowest performing group produced 773

t5.1 **Table 5** Decision Period 2 proportion usage of discourse types by highest and lowest performing groups

t5.2	Meta-Talk	Highest-Performing	Lowest-Performing	χ^2(df = 1)	*p value*	*Bonferroni adjusted alpha*
t5.3	Meta-Self	2.5%	10.7%	60.56	< .00001*	.004
t5.4	Meta-Group	18.6%	7.1%	58.35	< .00001*	.004
t5.5	Topic-Talk	Highest-Performing	Lowest-Performing	χ^2(df = 1)	*p value*	*Bonferroni adjusted alpha*
t5.6	Add	5.8%	6.7%	0.45	0.5042	.004
t5.7	Agree?	0.4%	2.0%	10.6	0.0011*	.004
t5.8	Agree	15.6%	12.1%	5.35	0.0207	.004
t5.9	Claim	20.8%	18.3%	2.44	0.1184	.004
t5.10	Clarify?	4.7%	8.5%	11.65	0.0006*	.004
t5.11	Clarify	3.8%	5.7%	3.56	0.0593	.004
t5.12	Counter	4.1%	3.7%	0.11	0.7440	.004
t5.13	Cut Off	0.4%	1.9%	7.52	0.0061	.004
t5.14	Strategy-Suggestion	12.5%	15.8%	4.96	0.0260	.004
t5.15	Question	11.0%	7.6%	6.26	0.0123	.004

Ns are shown in Table 3. During Period 2, the highest performing group produced 1137 utterances while the lowest group produced 646

are increasing their intra-group coordination over time, although again, as in the earlier two-sample analysis, differences in category use across time periods did not reach significance.

Types of meta-level talk

A qualitative examination of all Meta-group statements suggested they were of four major types. The first consists of a speaker's effort to understand another group member or members. A second involves taking stock and defining the group's position. A third consists of planning a particular next step. A fourth consists of evaluating the group's work. Table 9 contains examples of each of the four types, taken verbatim from group transcripts.

t6.1 **Table 6** Decision Period 7 proportion usage of discourse types by highest and lowest performing groups

t6.2	Meta-Talk	Highest-Performing	Lowest-Performing	χ^2(df = 1)	*p value*	*Bonferroni adjusted alpha*
t6.3	Meta-Self	4.9%	15.1%	41.82	< .00001*	.004
t6.4	Meta-Group	30.0%	10.8%	63.33	< .00001*	.004
t6.5	Topic-Talk	Highest-Performing	Lowest-Performing	χ^2(df = 1)	*p value*	*Bonferroni adjusted alpha*
t6.6	Add	9.0%	13.8%	6.83	0.0090	.004
t6.7	Agree?	0.4%	1.3%	1.71	0.1907	.004
t6.8	Agree	10.8%	8.2%	1.80	0.1802	.004
t6.9	Claim	16.5%	19.0%	1.32	0.2502	.004
t6.10	Clarify?	5.7%	6.9%	0.50	0.4803	.004
t6.11	Clarify	4.1%	5.6%	0.99	0.3193	.004
t6.12	Counter	4.4%	3.3%	0.53	0.4681	.004
t6.13	Cut Off	0.6%	1.3%	0.66	0.4174	.004
t6.14	Strategy-Suggestion	7.5%	8.5%	0.24	0.6264	.004
t6.15	Question	8.1%	6.2%	1.02	0.3117	.004

Ns are shown in Table 3. During Period 7, the highest performing group produced 955 utterances while the lowest group produced 305

JrnlID 11412_ArtID 9321_Proof# 1 - 17/06/2020

t7.1 **Table 7** Decision Period 8 proportion usage of discourse types by highest and lowest performing groups

t7.2 Meta-Talk	Highest-Performing	Lowest-Performing	χ^2(df = 1)	*p value*	*Bonferroni adjusted alpha*
t7.3 Meta-Self	7.0%	11.4%	6.69	0.0097	.004
t7.4 Meta-Group	19.3%	9.1%	25.08	< .00001*	.004
t7.5 Topic-Talk	Highest-Performing	Lowest-Performing	χ^2(df = 1)	*p value*	*Bonferroni adjusted alpha*
t7.6 Add	10.2%	9.4%	0.15	0.7010	.004
t7.7 Agree?	5.0%	1.2%	8.99	0.0027*	.004
t7.8 Agree	23.7%	13.8%	22.38	< .00001*	.004
t7.9 Claim	12.5%	18.5%	9.20	0.0024*	.004
t7.10 Clarify?	6.4%	8.8%	2.04	0.1528	.004
t7.11 Clarify	5.2%	4.4%	0.23	0.6281	.004
t7.12 Counter	0.9%	5.9%	24.6	< .00001*	.004
t7.13 Cut Off	0.5%	2.3%	6.10	0.0135	.004
t7.14 Strategy-Suggestion	4.9%	12.0%	21.12	< .00001*	.004
t7.15 Question	4.2%	3.2%	0.43	0.5107	.004

Ns are shown in Table 3. During Period 8, the highest performing group produced 781 utterances while the lowest group produced 341

As illustrated in Table 10, Meta-self statements are less varied and can be classified into two main categories, seeking and expressing understanding. In both cases, these pertain to the speaker's position and do little if anything to address or move forward the group's agenda.

These qualitative categorizations are thus consistent with the finding that frequent Meta-group discourse is associated with good group performance. Such statements appear to move the group's work along in a number of ways. The fact that a majority refer to "we" indicates formation of a group identity. Meta-self statements, in contrast, do not serve this function and are not associated with good outcomes. A further qualitative finding consistent with these conclusions comes from an analysis of who within a group is responsible for making Meta-group statements. A comparison of the highest performing and lowest performing groups at early (Periods 1 and 2) and late (Periods 7 and 8) sessions shows contrasting pictures in the two groups. In the low group, Meta-group statements are less frequent but more evenly distributed. All four of the group members are the highest scoring in producing Meta-group statements during one of the sessions. In contrast, in the high group, of the four members the

t8.1 **Table 8** Discourse categories distinguishing highest and lowest performing groups for Decision Periods 1, 2, 7 & 8

t8.2 Meta-Talk	PERIOD 1	PERIOD 2	PERIOD 7	*PERIOD 8*
t8.3 Meta-Self	*	*	*	
t8.4 Meta-Group	*	*	*	*
t8.5 Topic-Talk	PERIOD 1	PERIOD 2	PERIOD 7	*PERIOD 8*
t8.6 Agree?		*		*
t8.7 Agree	*			*
t8.8 Clarify?		*		
t8.9 Claim				*
t8.10 Counter	*			*
t8.11 Cut Off	*			
t8.12 Strategy-Suggestion				*

* indicates significant difference observed between highest and lowest performing groups with Bonferroni adjusted alpha = .004

t9.1 **Table 9** Types of Meta-group discourse

t9.2	**Understanding other**:
t9.3	I guess what you are trying to say is….
t9.4	What do you think?
t9.5	Are you saying TOPS is targeting the same segment?
t9.6	You are saying we have to lower the price in order to get the high earners.
t9.7	You want to keep the price?
t9.8	I don't know if you saw this.
t9.9	**Taking stock/Defining position:**
t9.10	What do we think?
t9.11	That's what we are saying.
t9.12	Which is what we are concerned about.
t9.13	What do we care about?
t9.14	We have to think about it.
t9.15	I don't think we have talked about it.
t9.16	That's what we have to do to get to there.
t9.17	The question we need to answer…
t9.18	We just want to hit the ideal point.
t9.19	We have been talking about this before.
t9.20	Okay last time we said we just leave those.
t9.21	Okay, I think we all agree.
t9.22	It's okay we can do it.
t9.23	**Planning:**
t9.24	I think all we need is probably like 30 to 40 min.
t9.25	Let's start with that.
t9.26	Let's do it now.
t9.27	Let's try that.
t9.28	So let's pick a price, what do you think.
t9.29	One thing we need to focus on is…
t9.30	Let's go on to the next thing.
t9.31	So let's just compare one.
t9.32	We can then go to our review report.
t9.33	Now let's talk about launching a new product.
t9.34	It's a thousand – should we go for it?
t9.35	**Evaluating:**
t9.36	We didn't do something right.
t9.37	We still haven't figured out how many units.
t9.38	We have an error – can you check that?
t9.39	Wait, we are working on MOST.
t9.40	We already did that, okay.
t9.41	We are close, yeah?
t9.42	We are pretty close.
t9.43	Okay so we are good.
t9.44	It's something really good for us.
t9.45	Yeah that's probably the biggest win for us.
t9.46	So we increased the market share.
t9.47	So I think this will help us for the next round.

same individual is the highest in producing Meta-group statements at all of the sessions. One of these members (not the high Meta-group member) happened to be available subsequent to the course's completion and was interviewed by one of the authors. In the interview, she confirmed that the high Meta-group member had assumed a role as group leader, for example by organizing the early group-level planning based on what skills and knowledge each group member brought to the task, and his role was recognized by all of the group members as playing an important role in their group's success.

JrnlID 11412_ArtID 9321_Proof# 1 - 17/06/2020

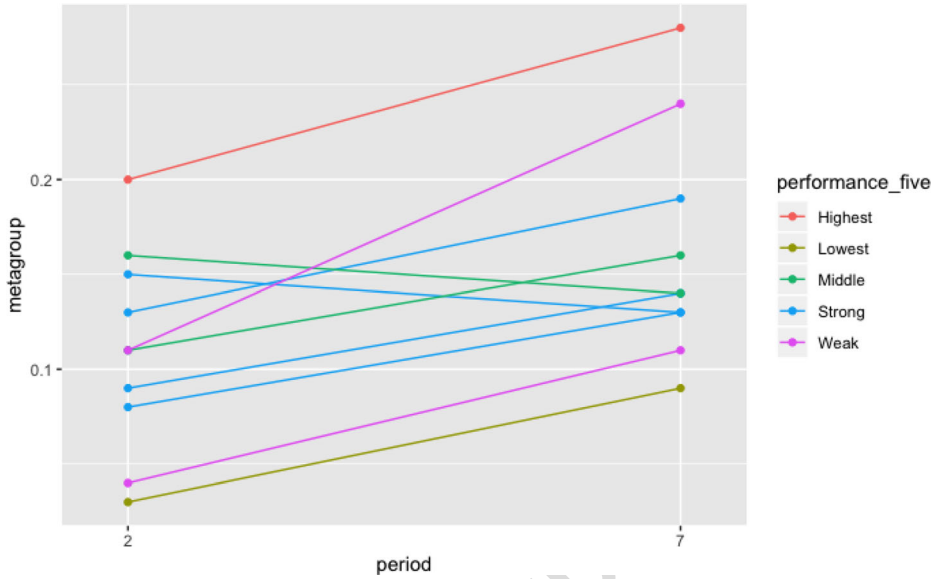

Fig. 1 Meta-group discourse proportions by group and performance outcome for Periods 2 and 7

Participant evaluations of group process

None of the distinctions across groups that have been reported were reflected in participants' questionnaire responses. Overall, group members did not show substantial variation in their responses, within or across groups, nor did any trends appear over time. At the end of their collaboration, members overall reported that their group worked very well together, with all but one group showing a group mean between 7.33 and 9.00 on a scale of 1–10. With the

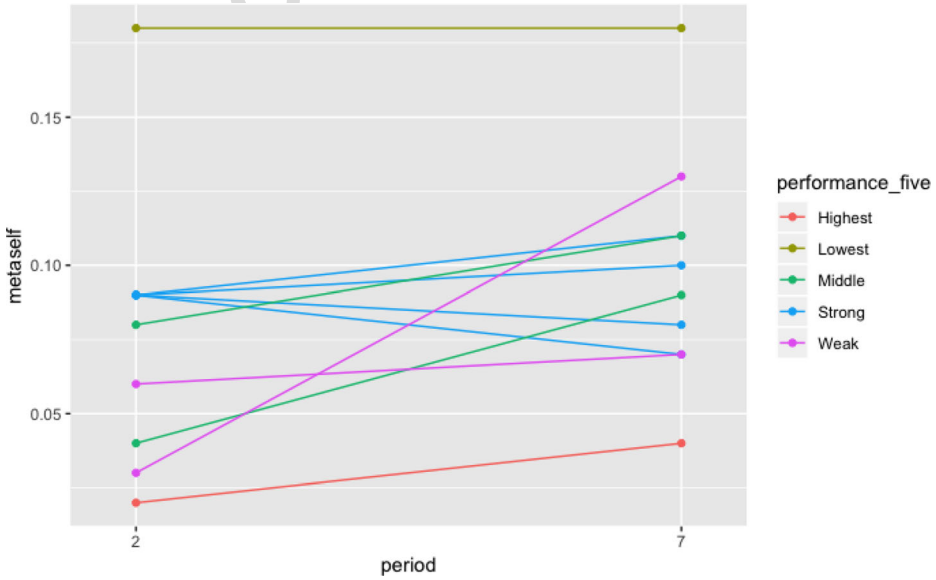

Fig. 2 Meta-self discourse proportions by group and performance outcome for Periods 2 and 7

t10.1 **Table 10** Types of Meta-self Discourse

t10.2	**Seeking understanding:**
t10.3	I don't know.
t10.4	No I haven't read it yet.
t10.5	I was confused, sorry.
t10.6	I still don't understand why initiate allocation.
t10.7	That's what doesn't make sense to me.
t10.8	I don't know what's our price point, comparatively?
t10.9	I haven't figured out the scale, the number yet.
t10.10	I don't know what you guys think about that.
t10.11	Cool, so I also want to read this.
t10.12	**Expressing understanding:**
t10.13	This is what I am saying.
t10.14	Yeah, that's in my mind.
t10.15	It's a good area in my opinion, but I don't know.
t10.16	I am hoping this will tell us.
t10.17	So I guess I am trying to figure out is there market research.

exception of this group, the large majority of members of all groups reported that their group made better decisions than they would have made alone. This outlying group was also one of only two groups in which one member reported playing a larger than proportional role in their group's decision making.

Discussion

The results of the present exploratory study support our hypothesis that meta-level talk about the group's activity stands to benefit the group's performance. Talk referring only to individual performance, in contrast, did not show this effect. Incidence of Meta-group statements and incidence of Meta-self statements in fact showed opposing relations to group performance, with only the Meta-group category relating positively to performance. Our results thus support the value of close observational analysis, employing both qualitative and quantitative indicators, of how groups undertaking a joint task talk to one another during the process and how differences in such talk relate to group success. The importance of attention to the thinking of others has been suggested in previous literature involving group process (Duhigg 2016; Hogan et al. 2016; Goldstone et al. 2008; Graesser et al. 2018; Järvelä et al. 2016). Much of the existing empirical literature, however, relies on group members' self-report. The present study thus makes the important contribution of supporting such findings when group process is studied directly.

Effective group functioning entails a process of coordination in which patterns of collaboration emerge and develop, as theorists studying group processes have emphasized (Graesser et al. 2018; Järvelä et al. 2016; Kapur 2008; Siqin et al. 2015). Meta-level processes are central to such coordination. Järvelä et al. (2016) identified three kinds of self-regulation within group interaction – forethought, performance and reflection (paralleling the planning and evaluating categories we report in Table 9). Järvelä et al. also reported systematic changes in their distribution across time. Although our results did not show significant trends in meta-level categories of talk across time, there was variation, and it is certainly the case that a group must work toward achieving the interpersonal coordination that allows group work to be productive. It is not automatic.

JrnlID 11412_ArtID 9321_Proof# 1 - 17/06/2020

Group members not only engage in coordinated ways to execute their task; they may reflect on this coordinated engagement, as a step toward enhancing it. To engage in such reflection, one must be attuned to learn the contents of others' minds, remain aware of them, and adjust one's own thinking in light of them (Kuhn 2015). Members of a group must also be cognizant of how other members perceive them. Failure of individual group members to fully employ these reflective tools stands to limit the productivity of the group as a whole (Siqin et al. 2015). When the task is one in which achievement is assessed at an individual level and does not depend on sensitivity to other minds, group work may not produce learning outcomes superior to that of individuals working on the problem alone. For example, the group nature of problem-based learning regarded as central to its definition appears not in fact to be an essential component of its success in the concept learning outcomes PBL achieves (Kuhn 2015; Pease and Kuhn 2011).

Our present findings support the view that shared reflection aids the group in achieving the coordinated action that is required if a group outcome is to be achieved superior to that of individuals working independently. The present findings should only be generalized with caution beyond the highly educated, highly motivated sample we studied, a group already well experienced in group problem solving. Our results nonetheless suggest that extension of this form of observational analysis to broader populations, while labor intensive (Graesser et al. 2018) is nonetheless worthwhile.

References

Duhigg, C. (2016). *Smarter faster better: The transformative power of real productivity*. New York: Random House.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist, 34*(10), 906–911.

Goldstone, R. L., Roberts, M. E., & Gureckis, T. M. (2008). Emergent processes in group behavior. *Current Directions in Psychological Science, 17*, 10–15.

Graesser, A., Fiore, S., Greiff, S., et al. (2018). Advancing the science of collaborative problem solving. *Psychological Science in the Public Interest*. https://doi.org/10.1177/1529100618808244.

Hogan, M., Dwyer, C., Harney, O., Noone, C., & Conway, R. (2016). Metacognitive skill development and applied systems science: A framework of metacognitive skills, self-regulatory functions and real-world applications. In A. Pena-Ayala (Ed.), *Metacognition: Fundamentals, applications, and trends* (pp. 75–106). New York: Springer.

Jacobson, M., Kapur, M., & Reimann, P. (2016). Conceptualizing debates in learning and educational research: Toward a complex systems conceptual framework of learning. *Educational Psychologist, 51*, 210–218.

Järvelä, S., Järvenoja, H., Malmberg, J., Isohätälä, J., Sobocinski, M., & Arta. (2016). How do types of interaction and phases of self-regulated learning set a stage for collaborative engagement? *Learning and Instruction, 43*, 39–51.

Kapur, M. (2008). Productive failure. *Cognition and Instruction, 26*, 379–424.

Kuhn, D. (2000). Metacognitive development. *Current Directions in Psychological Science, 9*, 178–181.

Kuhn, D. (2015). Thinking together and alone. *Educational Researcher, 44*, 46–53.

Kuhn, D. (2019). Why is reconciling divergent views a challenge? *Current Directions in Psychological Science, 29*, 27–32.

Kuhn, D., & Modrek, A. (2018). Do reasoning limitations undermine discourse? *Thinking and Reasoning, 24*, 97–116.

Kuhn, D., Ramsey, S., & Arvidsson, T. S. (2015). Developing multivariable thinkers. *Cognitive Development, 35*, 92–110.

Kuhn, D., Zillmer, N., Crowell, A., & Zavala, J. (2013). Developing norms of argumentation: Metacognitive, epistemological, and social dimensions of developing argumentive competence. *Cognition & Instruction, 31*, 456–496.

Macagno, F. (2016). Argument relevance and structure: Assessing and developing students' uses of evidence. *International Journal of Educational Research, 79*, 180–194.
Morris, M., Savani, K., & Fincher, K. (2019). Metacognition fosters cultural learning: Evidence from individual differences and situational prompts. *Journal of Personality and Social Psychology, 116*, 48–68.
Osiurak, F., & Reynaud, S. (2019). The elephant in the room: What matters cognitively in cumulative technological culture. *Behavioral and Brain Sciences*. https://doi.org/10.1017/SO140525X19003236.
Pease, M., & Kuhn, D. (2011). Experimental analysis of the effective components of problem-based learning. *Science Education, 95*, 57–86.
Pifarre, M., & Cobos, R. (2010). Promoting metacognitive skills through peer scaffolding in a CSCL environment. *International Journal of Computer-Supported Collaborative Learning, 5(2),* 237–253.
Rapanta, C., Garcia-Mila, M., & Gilabert, S. (2013). What is meant by argumentative competence? An integrative review of methods of analysis and assessment in education. *Review of Educational Research, 83*, 483–520.
Shea, N., Boldt, A., Bang, D., Yeung, N., Heyes, C., & Frith, C. D. (2014). Supra-personal cognitive control and metacognition. *Trends in Cognitive Sciences, 18*(4), 186–193.
Siqin, T., van Aalst, J., & Chu, S. K. W. (2015). Fixed group and opportunistic collaboration in *a CSCL environment*. *International Journal of Computer-Supported Collaborative Learning, 10*, 161–181.
Slof, B., Erkens, G., Kirschner, P., Janssen, J., & Jaspers, J. (2012). Successfully carrying out complex learning-tasks through guiding teams' qualitative and quantitative reasoning. *Instructional Science, 40*, 623–643.
Sloman, S., & Rabb, N. (2019). Thought as a determinant of political opinion. *Cognition, 188*, 1–7.
Yoon, S., Park, M., & Anderson, E. (2018). *Identifying reflective and non-reflective group consensus strategies for evidence-based scientific argumentation*. London: Proceedings of the International Conference of the Learning Sciences.