Running head: Correct contributions

Flowing toward correct contributions during group problem solving: A statistical discourse analysis

Ming Ming Chiu
SUNY - Buffalo

May 4, 2008

Abstract

This study tested whether groups that created more correct ideas (correct contributions or CCs) or more clusters of CCs were more likely to solve a problem. Also, it tested whether students' recent actions (micro-time context) aided CC creation. Specifically, did new ideas, argumentation (evaluations, questions, and justifications), politeness, or status differences affect the likelihood of a CC?

Eighty high school students worked in groups of 4 on an algebra problem. Groups with higher mathematics grades or more CCs were more likely to solve the problem, but the number of clusters of CCs was unrelated to the solution outcome.

A tool for analyzing individual or group processes, dynamic multilevel analyses (DMA) modeled the groups' 2,951 conversation turns. It statistically identified watersheds (breakpoints) that divided each group's conversation into distinct time periods with many CCs vs. few CCs. Wrong contributions, correct evaluations of one another's ideas, justifications, and polite disagreements increased the likelihood of a CC, while questions, rude disagreements, and agreements reduced it. Status differences were not linked to CCs. Students recognized flaws in wrong contributions and used them to build CCs. Compared to incorrect evaluations, correct evaluations increased the likelihoods of CCs, justifications, and subsequent correct evaluations. A justification often yielded a subsequent justification and reduced the likelihood of a rude disagreement. Justifications had the largest effects, while correct evaluations' effects lasted 3 speaker turns. Some effects differed across groups or across time periods. In groups that solved the problem, justifications were more likely to yield CCs, and questions were more likely to elicit explanations. Meanwhile, agreements' and correct evaluations' effects on correct contributions differed across time periods. Applied to practice, teachers can encourage students to explain their answers to group members' questions, express and justify their own ideas, and evaluate others' ideas carefully and politely.

Keywords: Cooperative learning, social interactions, hierarchical linear modeling, time series analysis, quantitative discourse analysis

This research began at the Chinese University of Hong Kong and was partially funded by a Spencer Foundation research grant, a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CUHK 4117/04H), and a Chinese University of Hong Kong direct grant. I appreciate the research assistance of Choi Yik Ting and the helpful comments by Noreen Webb, Bonnie Chow Wing-yin, Sung Wook Joh, and six anonymous reviewers.

This is an accepted manuscript of an article published by Taylor & Francis in Journal of the Learning Sciences, Volume 17, Issue 3, 2008, available online: http://www.tandfonline.com/doi/abs/10.1080/10508400802224830

Flowing toward correct contributions during group problem solving: A statistical discourse analysis

Students who work together often show many positive outcomes, including more learning, greater motivation, less racial tension, etc. (Johnson & Johnson, 2002; Slavin, 1990; Webb & Palincsar, 1996). However, these positive results are not universal (e.g., Barron, 2003). Why are some groups more successful than others?

Earlier researchers used group structures and group member traits to explain different outcomes, while recently researchers have focused on group processes. Theoretical models of group problem solving highlight the importance of group members' new ideas (contributions), especially correct ones (correct contributions or CCs), as the building blocks of successful outcomes (e.g., Chiu, 2000a, 2001; Hinsz, Tinsdale & Vollrath, 1997). This raises the issue of how group member properties and actions affect the process of creating CCs and whether these effects differ across groups or across time periods within each group.

Case studies of group problem solving showed that diverse views and argumentation affect the creation of CCs (Cobb, 1995), and past studies showed that status and public self-image concerns are linked to group problem solving actions (Dembo & McAuliffe, 1987). Extending this line of research, I *statistically modeled the temporal development of group processes*, specifically *how sequences of micro-processes (micro-time context) helped or hindered creation of CCs* during group problem solving in high school algebra classrooms (cf. Mercer, 2008).

This study contributes to the research literature in four ways. First, I *statistically identify distinct time periods* and analyze whether CCs occur uniformly throughout a problem-solving session or cluster within specific time periods. Second, I examine how prior speakers' actions (e.g., evaluations, ideas, justifications, and rudeness) and interactions create a micro-time context that affects the likelihood of creating a CC. Third, I test whether the above effects differ for groups with correct solutions vs. those with incorrect solutions. Lastly, I apply a new statistical method to address the above research questions, dynamic multilevel analysis (DMA, Chiu & Khoo, 2005). As CCs are central to correct solutions, understanding when they occur and the factors that affect their creation can help educators improve students' group problem solving.

Group Problem Solving Processes

Past research suggests that groups with more CCs are more likely to solve a problem correctly than other groups (*functional theory of group decision-making*, e.g., Orlitzky & Hirokawa, 2001). Let us call this hypothesis H-1. (Hypotheses are numbered according to their level of analysis [group: 1; time period: 2; speaker turn: 3]).

Hypothesis H-1. Groups with more CCs are more likely to solve the problem correctly.

Successful group problem solving might yield more clusters of CCs via diverse ideas and argumentation (Amason, 1996; Cobb, 1995). However, rude arguments might hinder CCs and group problem solving, especially arguments centered on status struggles. This study investigated CCs by examining whether they clustered together and by identifying group problem solving processes that helped or hindered the creation of CCs.

Clusters of CCs in time periods

Several researchers claim that group problem solving tends to be more successful if a group steps through each formal problem solving phase in order (*phase model*) such as (a) clarify the problem, (b) discuss criteria, (c) propose solutions, and (d) evaluate proposals (Ellis & Fisher, 1994; Pavitt, 1993). In such a model, components of a correct solution are split across the different phases. Thus, group members likely voice correct and incorrect ideas at each stage (barring an algorithmic march to a correct solution). Hence, successful groups with multiple phases might tend to have clusters with many CCs alternating with clusters with few CCs.

However, groups often do not step through each phase (Hirokawa, 1983; Pavitt & Johnson, 2001). Instead, many groups prefer to discuss one solution proposal in full, then another, and so on (*reach-testers*; Pavitt & Johnson, 2001). These reach-tester groups might have all CCs clustered around the correct proposal and incorrect ideas elsewhere, yielding only one or two distinct time periods. As phase model groups are more likely to be successful than reach-tester groups, CCs might occur in more clusters in successful groups rather than in unsuccessful ones, especially for difficult problems.

Hypothesis H-2. CCs occur in many clusters in successful groups, but in fewer clusters in unsuccessful groups.

Group problem solving actions that help create CCs

Compared to individuals, group members' diverse perspectives and argumentation might create more CCs (Cobb, 1995; Paulus & Brown, 2003). Diverse points of view can help a group create more ideas and judge them more accurately compared to individuals (Paulus & Brown, 2003).

New ideas. Group members often have diverse perspectives and sources of knowledge (Stasser, 1992). Capitalizing on this diversity, groups (especially heterogeneous ones) often create many ideas, representations, and solution proposals, thereby raising the likelihood that at least one of their ideas is correct/optimal (Paulus & Brown, 2003).

Group members can express idiosyncratic ideas and build on them to create new alternatives through processes such as sparked ideas, jigsaw pieces, and creative misinterpretations (Paulus & Brown, 2003; Chiu, 1997). Comments by one person (e.g., a key word) might spark another person to activate related concepts in his or her semantic network and propose a CC (Nijstad, Diehl, & Stroebe, 2003). Or, two or more members can put together different pieces to construct a CC, like fitting jigsaw pieces together (Chiu, 1997). Finally, a person might misinterpret a group member's incorrect idea to create a new, correct one (Chiu, 1997). Thus, even wrong contributions can lead to CCs.

Group members' diverse views also help them recognize flaws, refine these incorrect ideas, and create CCs (Cobb, 1995; Piaget, 1985). Groups with diverse views might create more wrong ideas, but their diverse views also improve their judgment of their validity. Hence, they can detect and correct these flaws to create CCs. This contrasts with the view that people primarily build on correct ideas and that wrong ideas often lead the group astray (see the long middle column of Figure 1, which summarizes the hypotheses and their relationships).

Hypothesis H-3a. Contributions, including wrong contributions, help create CCs.

Argumentation. Successful group problem solving often involves argumentation in the cognitive/problem content space (Roschelle, 1992), a social process by which people explain and justify their own views to convince both themselves and others (Amason, 1996; Cobb, 1995). During this process, group members evaluate one another's ideas, detect flaws, and justify their ideas (Cobb, 1995; Kuhn, Shaw & Felton, 1997). These argumentation processes can help students develop their understanding of the specific content (e.g., algebra's structural relationships among equality, arithmetic operations, and properties of number, Kieran, 1992).

Evaluations characterize how a person assesses the previous speaker's action and problem solving approach (*functional theory of group decision-making*, e.g., Chiu, 2000a, 2001; Orlitzky & Hirokawa, 2001). For example, Sean says "three times four is seven." Maya can agree ("uh-huh"), use a neutral action ("what did you say?"), disagree ("nope, you're wrong"), or change the topic ("when is class over?"). While agreements continue the current problem-solving trajectory, disagreements and changes of topic try to change the trajectory (Chiu, 2001). Hence, these evaluations reflect the accountability of a person's ideas and actions to his or her learning community, to its collective knowledge, and to its local standards of reasoning (e.g., validity of transformations among equivalent algebraic expressions, Michaels, O'Connor, & Resnick, in press).

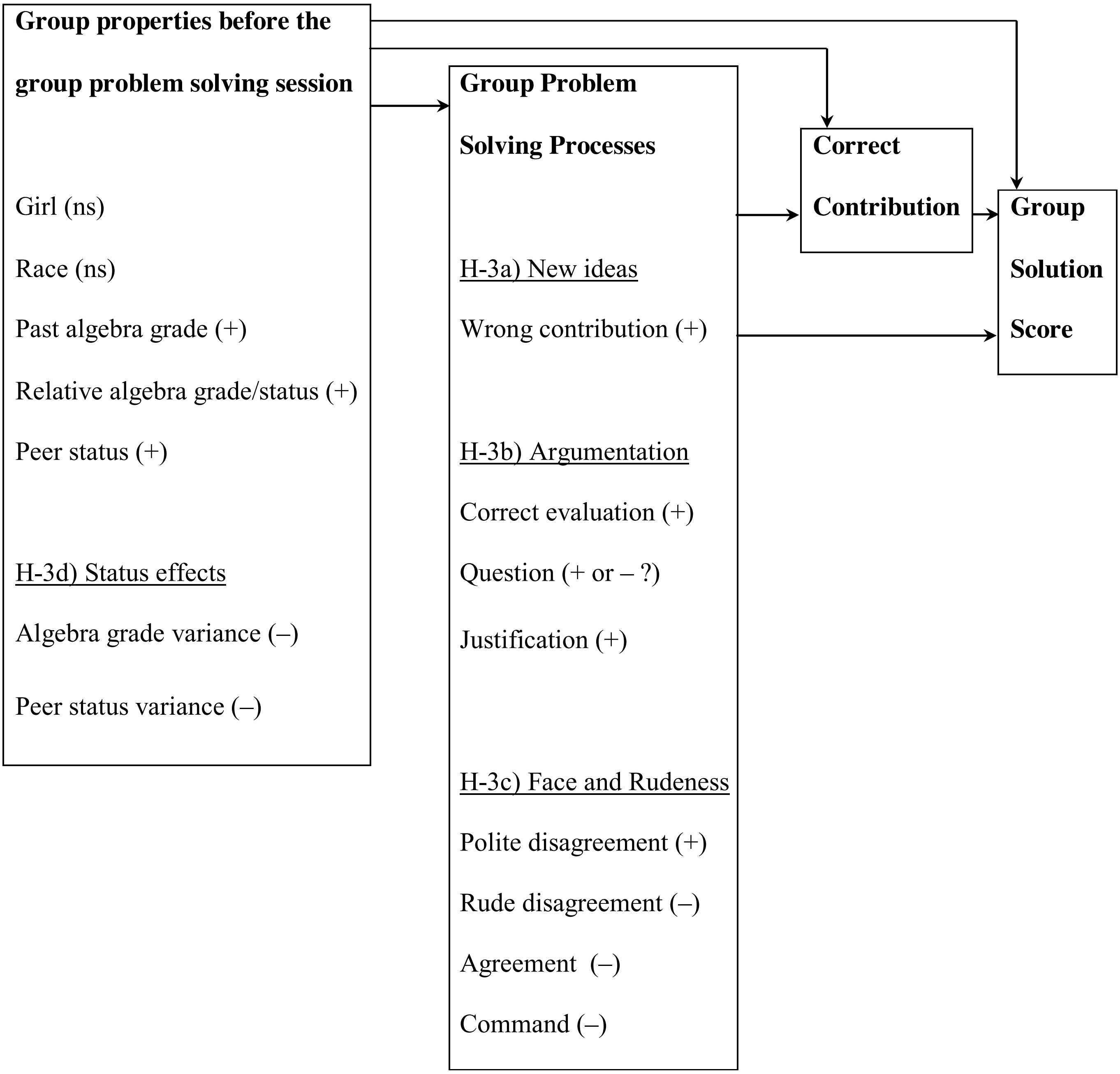

Figure 1. Model of the effects of student and group properties before and during the group problem solving process on the outcome variables correct contributions and on group solution score (symbols in parentheses indicate expected direction of relationship with the outcome variables: positive [+], negative [–], or not significant [ns]).

Evaluations can also be right or wrong in many contexts (e.g., high school algebra).

Correct evaluations support correct ideas (“three times four is twelve, right”) or identify flawed ideas (“no, three times four is not seven,”), thereby creating a foundation of partially shared understandings of correct ideas that group members can use to build new CCs. In contrast, incorrect evaluations reject CCs (“nope, five times two isn’t ten,”) or accept flawed ideas (“three times four is seven, yeah”), embedding flaws in their partially shared understandings. Group members using these partially shared understandings can carry these flaws into their new ideas, resulting in more wrong contributions and fewer CCs. A group's collective attention and diverse perspectives can help it evaluate ideas correctly and create a partially-shared foundation of understanding to aid creation of CCs (see Figure 1, middle column; Hinsz, 1990; Cobb, 1995).

According to *socio-cognitive conflict* theory, group members can recognize problems or difficulties (perturbations), express them through disagreements or questions, and address them to improve their understanding (Doise, Mugny and Perret-Clermont, 1975; Piaget, 1985). Piaget (1985) defines two types of perturbations: (a) obstacles, which give negative feedback and (b) lacunae, gaps in understanding. Thus, disagreements can identify obstacles (e.g., "no, that’s wrong, three times four isn’t seven") and motivate the need to create CCs.

Meanwhile, a question (e.g., Juan asks, "how do we find the speed?") can indicate an individual gap or a group gap. For an individual gap, other group members who know the answer can explain it. Thus individual gap questions invite explanations that often review previous ideas rather than create new CCs. In contrast, no one knows the answer to a group gap question, which motivates the need for a CC and points to a new direction for creating it. By expressing their ideas and explanations, students open up their reasoning for group members to analyze and discuss (Franke, Carpenter, & Battey, 2007). De Lisi and Goldbeck (1999) argue that group members' diverse perspectives and levels of knowledge facilitate both perturbations and responses to them. In short, these perturbations can motivate and inform creation of more CCs (see Figure 1, middle column).

Both disagreements and questions invite justifications that establish an idea's validity within their local classroom community's negotiated norms for a specific content area (e.g., algebraic relationships and mathematics proofs, Balacheff, 1988). Chiu and Khoo (2003) showed that members of successful groups often anticipated criticisms and justified their new ideas. Likewise, after a person disagrees with a proposal (e.g. Maya), the original proposer (Sean) might justify it by linking it to data, using a warrant, or supporting a warrant with backing (Toulmin, 2003). Along with appeals to external authorities (e.g., teacher, textbook), mathematics also allows students to build on example-based justifications by generalizing them to create principled, internal justifications of structural relationships within a closed system (Sowder & Harel, 1998). In response, other members can give different views and justifications (Piaget's, 1985, genuine argument). Similarly, when Juan shows a gap in understanding by asking a question, other members can respond with explanations and justifications (Coleman, 1998). As justifications support an idea’s validity, they can help create CCs (see Figure 1, middle column; e.g., Goldbeck, 1998).

Hypothesis H-3b. Correct evaluations, group knowledge gap questions, and justifications facilitate the creation of CCs.

Group problem solving actions that hinder CCs

Disagreements can help create CCs according to socio-cognitive conflict theory, but the effects might differ for polite and rude disagreements according to politeness theory (Brown & Levinson, 1987; Chiu, 2001). Polite disagreements likely facilitate CCs and group problem solving, but rude disagreements can hinder them, especially during status struggles. When arguments spill over from the problem content space into the social relational space (Barron, 2003), groups members might sacrifice further problem solving progress in favor of protecting their public self-images (*face*, Brown & Levinson, 1987; Chiu & Khoo, 2003). Status differences can further aggravate these face concerns.

Face and rudeness. As problem solving occurs in the dual space of problem content and social relations, each type of evaluation can affect both the problem solving (as noted above) and the previous speaker's face (Chiu, 2000b, 2001). Evaluations range from polite to rude: agreement, neutral, change of topic, and disagreements (Brown & Levinson, 1987; Chiu, 2000b, 2001). Consider Sean's utterance again, "three times four is seven." If Maya agrees with Sean ("uh-huh"), she supports him, promotes his face, and enhances their social relationship (Brown & Levinson, 1987). Thus, members often repeat shared information to create common ground and solidarity (Clark & Brennan, 1991). Moreover, people spontaneously reciprocate positive affective displays, such as eye contact, to suggest agreement with one another (Burgoon, Dillman & Stern, 1993).

In contrast, other actions do not support face. Neutral actions include discourse management or meta-discourse actions (e.g., "what did you say?"). Although changes of topic ("when is class over?") can be neutral, they can be rude if the previous speaker (Sean) expects a response (e.g., if Sean asks the question, "three times four is seven?"). If Maya says "when is class over?" after Sean's question, she either ignores him or does not listen to him, both of which are rude. Lastly, disagreements (e.g., "no, you're wrong") can threaten face by lowering public perception of the previous speaker's (Sean's) competence (Brown & Levinson, 1987).

When a person disagrees (e.g., Maya says, "nope, you're wrong"), the target person (Sean) ideally tries to understand the criticism and use the information to create a CC. Instead, the threat to Sean's face may encourage him to retaliate emotionally (*face attack*, "no, I'm not! You are. You're always making mistakes in math …." Chiu & Khoo, 2003; Tracy & Tracy, 1998). Thus, rude disagreements threaten face, escalate interpersonal conflict, and often hinder creation and recognition of CCs (see Figure 1, middle column). In the worst case, a spiral of rude disagreements can kill the collaboration. Even if the collaboration survives after a rude disagreement (or some other rude action, e.g., insult), group members might withhold CCs or correct evaluations rather than risk losing face (Chiu & Khoo, 2003).

To avoid threatening Sean's face, Maya might go to the opposite extreme and publicly agree. By doing so, Maya enhances her social relationship with Sean at the expense of their problem solving. Such false agreements allow errors to persist and potential CCs to remain unspoken (see Figure 1, middle column). For example, teenage girls often avoid disagreeing with one another (Tudge, 1989). Even authority does not eliminate this effect, as tutors often do not point out their students' errors (Person, Kreuz, Zwaan, & Graesser, 1995).

Avoiding the extremes of rude disagreement and false agreement, Maya can disagree politely (with redress) to reduce the threat to Sean's face and maintain problem solving integrity (Brown & Levinson, 1987; Chiu & Khoo, 2003). Instead of "no, you're wrong," Maya can disagree politely, "If three is multiplied by four, we don't get seven." The polite disagreement both reduces blame and creates common ground. First, Maya uses the hypothetical "if," thereby distancing the error away from them. Second, she does not refer to Sean, (no "you") thereby avoiding assignment of blame. Third, Maya uses the passive voice, "is multiplied," not the active voice, to hide causal agency and responsibility. Lastly, she uses the passive circumstantial verb "get," thereby implicating agency in external conditions.

Maya's polite disagreement creates common ground by repetition and shared positioning. By repeating Sean's computation, "three is multiplied by four . . . seven," Maya suggests that she shares his understanding. Maya also uses shared positioning, "we," to claim common cause with Sean.

Maya's polite disagreement supports her relationship with Sean, so he is less likely to retaliate. Instead, Sean is more likely to try to understand Maya's criticism, recognize the flaw, and correct it with a CC (Chiu & Khoo, 2003). Indeed, the benefits of polite disagreements are so strong that it is the accepted norm among peers, as lack of redress during a disagreement is noticeably rude and unacceptable (Holtgraves, 1997). In short, polite disagreements might support social relationships, CC construction, and CC recognition, thereby enhancing both the problem content and social relational spaces (see Figure 1, middle column).

Other rude actions include commands and insults. As commands demand action from the target listener(s), they impinge on the target listener's freedom and are less polite than questions or statements (see Figure 1, middle column). Likewise, insults attack the target listener's face (Tracy & Tracy, 1998).

Hypothesis H-3c. Polite disagreements help creation of CCs, but rude disagreements, false agreements, and commands hinder them.

Status. According to *status characteristics* theory, status differences can reduce CCs and distort evaluations through the pursuit of high status via status struggles (Bales, 2001; Gersick, 1988) or through the greater influence of high status members (Cohen, 1994). Cohen (1994) defined status as "an agreed-on rank order where it is generally felt to be better to be high than low rank" (p. 23).

As a higher status person often receives more group resources and attention, people often compete for higher status (status struggles), especially if no clear status hierarchy exists (Bales, 2001; Gersick, 1988). During status struggles, intentional rude disagreements can hinder creation of CCs, but they can also enhance one's face by forcing a competitor to lose face (face attacks; e.g., "five times two is *obviously* ten, not seven," Tracy & Tracy, 1998).

After a status hierarchy has been established, group members expect higher status members to have greater task competencies and to contribute more toward their desired outcome(s) (Dembo & McAuliffe, 1987). As a result, higher status members have more opportunities to perform and receive rewards, as others selectively invite and defer to their opinions while discouraging, undervaluing, or outright ignoring lower status members' ideas. Thus, excessive attention to status can distort evaluations toward excessive agreement with higher status members. By doing so, group members enact their expectations of high status members dominating the interaction and might increase the ratio of flaws to correct ideas in their partially shared understandings.

High status member's influence can also increase over time. High status people tend to speak early and often (Hackman & Johnson, 2000). As group members value and prefer supporting previously-discussed, shared information rather than introducing new, unshared information (Stasser & Titus, 1985), high status member's domination increases in severity over time (Stasser & Taylor, 1991).

Greater status differences might increase the incentives for status struggles and yield greater status effects, both of which might reduce CCs (see Figure 1, left column). For group problem solving among students, the primary status characteristic is often past achievement, but group members might also use diffuse status characteristics (e.g., race, gender) to make assumptions about one another's competence (Webb, 1984; Cohen, 1982).

Hypothesis H-3d. Greater differences among group members' statuses (achievement, peer status, gender, or race) reduce CCs.

Successful vs. unsuccessful groups

Considering the above group processes, groups with four types of properties might be more likely than groups without these properties to solve a problem successfully. First, groups with more CCs are more likely to succeed than other groups. Second, phase model groups are more likely to have their CCs occur in clusters and to succeed compared to reach-tester groups. Third, groups that more often engage in processes that help create CCs are more likely to succeed (see Figure 1). Fourth, groups that show more rude behaviors or have larger status differences (processes that hinder CCs) are less likely to succeed.

In sum, this study examines the process of CC creation during group problem solving by testing the following hypotheses. First, groups with more CCs are more likely to solve the problem correctly. Second, CCs occur in more clusters in successful groups than in unsuccessful ones. Third, correct and wrong contributions help create CCs. Fourth, correct evaluations, group knowledge gap questions, and justifications aid CC creation. Fifth, polite disagreements help creation of CCs, but rude disagreements, false agreements, and commands hinder their creation. Lastly, greater status differences hinder CC creation.

Method

Using videotapes and transcripts described in Chiu and Khoo (2003, 2005, and in press), this study addresses a different research question with a different outcome variable (correct contribution). In Chiu and Khoo's (2003) study, the variation in student evaluations of one another's ideas during group problem solving occurred mostly at the speaker turn level rather than at higher levels (e.g., group or classroom). Thus, this study focuses on simpler, proximal analyses of speaker turns, time periods, and groups (leaving more complex, distal analyses involving classroom and school differences for future studies).

I analyzed the data at the group, time period, and speaker turn levels to model problem solving outcomes and processes. Representative transcript segments illustrate the relationships among variables.

Participants

The participants attended four ninth grade algebra classes in an urban US high school, which scored at the 40th percentile (maximum = 100; California Department of Education, 2005). Eighty-seven students were asked to answer a peer status survey and to be videotaped. Of the 87 students, 7 (or 8%) declined to participate. (Of these 7 students, 4 were girls and 3 were boys. Their average grade was 77 / 100.) There were 40 girls and 40 boys. Their races were 12 Asian, 27 Black, 28 Hispanic, and 13 White.

These students worked in groups of four. There were no same gender groups and no same race groups. These students attended the same algebra class for seven months and were likely aware of one another's mathematics abilities through conversations inside and/or outside of class. However, these students had not received any group work training and had not previously worked together in groups. Thus, group members' relative mathematics abilities were more likely to have a primary status effect. Likewise, diffuse status characteristics such as gender and race were likely to have smaller effects compared to that of strangers (Sharan & Shachar, 1988; Webb, 1991).

Procedure

All 80 algebra students who agreed to participate answered two questions regarding peer status, "Who are 3 classmates you would most like to hang out with? Name 3 classmates who are the easiest for you to talk with outside of school work." Later, their teacher presented the following problem in their algebra classes:

> "You won a cruise from New York to London, but you arrive 5 hours late. So, the ship left without you. To catch the ship, you rent a helicopter. The ship travels at 22 miles an hour. The helicopter moves at 90 miles an hour. How long will it take you to catch the ship?"

As advocated by cooperative learning researchers (e.g., Cohen, 1994; Johnson & Johnson, 1994), this problem was challenging for these groups of students and had multiple solution methods (see Appendix A). The classes had studied equations with single variables, and the teacher used the above problem to introduce them to a new unit on algebraic equations with multiple variables. Hence, the students had not yet learned, in class, any procedures for solving this problem. Furthermore, the problem involved complicated mathematics relationships, non-trivial combinations of algebraic operations, and a non-integer solution. One solution equates the distance computations for each vehicle (cruise ship and helicopter; 22 mph x [Time + 5 hours] = 90 mph x Time), to obtain 1.618 hours or 1 hour 37 minutes.

The students worked in groups for 30 minutes. (If students finished early and chatted off-task, this off-task talk was not included in the analysis. Only groups that successfully solved the problem finished early.) They had pens, paper, and calculators available for their use. There were six to seven videocameras in a classroom, one following the teacher and one for each group of students. Likewise, the teacher and each group of students had their own microphone and audiotape recorder to backup the video recordings. The videotape data were transcribed, coded, and analyzed.

Variables

See Table 1 for summary statistics and descriptions of variables. Using a similar set of data from a pilot study, I trained two research assistants (RAs) to transcribe and code the videotapes. Each transcript was divided into sequences of words or actions (e.g., writing "3 + 40") by the same person (*speaker turns*). Blind to the study's hypotheses, the RAs coded each speaker turn from the videotape on to a transcript, maintaining a log of each videotape to aid their coding. To compute the inter-rater reliability, I used Krippendorff's α (2004). Unlike other reliability measures, Krippendorff's α applies to any number of coders, any number of categories or scale values, any level of measurement, any sample size, and incomplete data. Its values range from -1 (maximum disagreement) to 1 (perfect agreement). A value near 0 indicates chance agreement, and a value of 0.7 or higher indicates satisfactory agreement.

The RAs tried to settle coding disagreements by consensus. They could not agree in 19 cases, so I made the final coding decision. Due to poor sound quality, 49 speaker turns could not be coded. These turns were coded as *missing* and inspected with adjacent outcomes and predictors for significant correlations. As they did not correlate significantly with other variables, omitting them was not likely to affect the results.

Speaker turn variables. Unlike flat classification schemes that only allow one or two codes for each speaker utterance (e.g., Bales, 2001), the RAs coded each speaker turn along five dimensions: evaluation of the previous action (EPA), knowledge content (KC), validity, justification, and invitational form (IF), (Chiu, 2000a; Chiu & Khoo, 2003). EPA, KC, and IF captured interactions and relationships across speaker turns (*relational* measures). See Table 1, Table 2, and Appendix B for coding examples, coding decision trees and further details. As the data only had two insults, the statistical power was too low to test their effects.

Data Analysis

A group-level analysis tests the relationship between CCs and a correct solution, followed by a discussion of some difficulties with lower level analyses and strategies for addressing them. Afterwards, I identified watersheds (breakpoints) and time periods, followed by speaker turn-level analyses of CCs. See Table 3 for a summary of the hypotheses, data, models of variables, and theoretical rationales.

Predicting solution score at the group level. First, I ran an analysis of the outcome variable *solution score* to test hypothesis H-1, that groups with more CCs were more likely to solve the problem. Hierarchical regressions and path analyses were used to test for total, direct, and indirect effects on solution score. Solution score was an ordered variable, not a continuous one, so using a least squares regression would have biased the estimation of the standard errors (Kennedy, 2004). Ordered Logit addressed this problem (Kennedy, 2004).

Table 1
Summary table of group level variables

Group Variable	Mean	SD	Min	Max	Description
Outcome variables					
Solution score	1.90	1.25	1	3	Score of group's final solution. See appendix A.
Correct contribution (CC)	0.28	0.18	0.02	0.59	A correct idea that has not been mentioned earlier during the group problem solving session
Before problem solving					
Classroom_1	0.25	0.44	0	1	Binary variable for students in classroom 1. Baseline classroom is classroom 4.
Classroom_2	0.30	0.46	0	1	Binary variable for students in classroom 2.
Classroom_3	0.25	0.44	0	1	Binary variable for students in classroom 3.
Girl	2.00	0.65	1	3	Number of girls in each group (0 indicates all boys).
Asian	0.60	0.50	0	2	Number of Asians in each group.
Latino	1.40	0.75	0	4	Number of Latinos in each group.
White	0.65	0.49	0	2	Number of Whites in each group.
Mathematics grade	82	7	71	92	Mean of all students' last semester's mathematics grades within a group.
Highest mathematics grade	92	8	77	99	Highest mathematics grade of any student within a group
Peer status	23	8	9	37	Mean number of times a student's name appeared in classmates' answers to the following questions. Who are 3 classmates you would most like to hang out with? Name 3 classmates who are the easiest for you to talk with outside of school work. This measure is the mean for the group and serves as a proxy for the group's social skills.
<u>Measures of status effects</u>					
Math grade Variance	101	70	12	300	Variance of students' mathematics grade within each group
Peer status Variance	37	29	0.25	108	Variance of peer status within each group
During problem solving					
Total no. of words	1363	1174	371	3,885	Total number of words during a group's problem solving sessions. The large standard deviation is partially due to five groups that spoke over 2,400 words each.
Total on-task words	1338	1277	342	2,841	Total number of words spoken on on-task turns during a group's problem solving sessions. See total words note.
<u>New ideas</u>					
Wrong contribution	0.12	0.05	0.04	0.20	A flawed idea that has not been mentioned earlier during the group problem solving session
<u>Argumentation</u>					

Correct evaluation	0.37	0.19	0.14	0.79	Agree with the previous speaker's correct idea or disagree with the previous speaker's incorrect idea
Unresponsive	0.16	0.09	0.02	0.31	Ignore the previous speaker; initiate new topic
Polite disagreement	0.16	0.06	0.06	0.27	Disagreement with the previous speaker with at least one form of redress
Question	0.23	0.07	0.15	0.45	An elicitation that expects a verbal response or a non-verbal substitute (Tsui, 1992; Sinclair & Coulthard, 1992)
Justification	0.12	0.09	0.04	0.30	An action that supports an answer or claim by at least one of the following: linking it to *data*, using a *warrant*, or supported by *backing* (Toulmin, 2003).
<u>Face and Rudeness</u>					
Rude disagreement	0.09	0.05	0.02	0.19	Disagree with the previous speaker without redress
Agreement	0.58	0.10	0.39	0.86	Agree with the previous speaker
Command	0.06	0.07	0.00	0.20	A directive that invites a non-verbal response (Sinclair & Coulthard, 1992)

Table 2.
Coding of a classroom discourse segment along five dimensions: (1) evaluation of the previous action (EPA: agreement [+], polite disagreement [–], rude disagreement [---], ignore/new topic[*]), (2) knowledge content (KC: contribution [C], repetition [R], null academic content [N]), (3) validity (right [√], wrong [X], null academic content [N]), (4) justification (justification [J], no justification [], null academic content [N]), and (5) form of invitation to participate (IF: (command [!], question [?], statement [_.]).

Person	Action	EPA	KC	Validity	Justify	IF
Ana	Do three times four hours.	*	C	√	[]	!
Ben	Three times four is–	+	R	√	[]	_.
Eva	–three times four is seven hours.	+	C	X	[]	_.
Jay	Wrong, three times four is eight hours.	---	C	X	[]	_.
Ben	If we do three times four, don't we get twelve hours because four plus four plus four is twelve?	–	C	√	J	?
Ana	Yep.	+	N	√	N	_.

To predict solution score, the following independent variables were added to the regression. First, classroom identification binary variables were entered to control for classroom effects. Then, the group's mean mathematics grade and its members' highest mathematics grade were entered into the regression both separately and together. If both were significant predictors alone but neither was significant together, the one that explained more solution variance (McFadden's, 1974, R^2) was kept. The variables *total number of words* and *total number of on-task words* controlled for the total talk in each group. Time constrains the direction of causality, so group processes cannot affect characteristics prior to the group problem solving. Hence, I entered characteristics of group members into the regression before entering group processes. The order was: mathematics grade (mean and/or highest), peer status, mathematics grade variance, peer status variance, words, on-task words, and percentage of CCs over total group turns. (Unlike percentage of CCs over total turns, a simple CC total favors groups that generate lots of ideas, both correct and incorrect. Meanwhile, the ratio of CCs over new ideas is a measure of accuracy that might overrate groups that produced few ideas. Regressions of solution scores with these other variables tested the results' robustness.)

A nested hypothesis test (χ^2 log likelihood) checked whether each set of added variables was significant (Kennedy, 2004). Only significant variables were retained in subsequent regressions.

A path analysis tested for direct and indirect effects. As time constrains the direction of causality, the predictors were entered in temporal order into the path analysis. These computations were performed with the statistical software, E-views (Lilien, Startz, Ellsworth, Noh & Engle, 1995). As the underlying distribution was not known, I repeated the above analyses with ordered Probit to ensure that the results did not depend on the Logit distribution. Note that the small sample size (N = 20) limits the statistical power of this analysis to identify non-significant results at the group level (power = 0.25 for an effect size of 0.3).

Addressing difficulties of group process analyses. Statistical analyses of group processes at the speaker turn level must overcome three difficulties. First, group members' behaviors and effects differ across groups and across time (nested data). Second, the outcome variable is discrete, not continuous. Third, events are often similar to recent events in time-series data (serial correlation).

Ordinary least squares (OLS) regressions do not address these difficulties. First, OLS often underestimates the standard errors of regression coefficients when applied to nested data (Goldstein, 1995). Second, OLS is inefficient for discrete variables and yields biased results (Kennedy, 2004). Lastly, if the time-series relationships are not modeled properly, the model residuals can be serially correlated, resulting in inefficient parameter estimates and biased estimates of the parameters' standard errors (Kennedy, 2004).

Thus, I address these difficulties by using a statistical discourse analysis tool, dynamic multilevel analysis (DMA, Chiu & Khoo, 2005). DMA identifies distinct time periods, tests for group and time period differences, builds an explanatory model for CCs, tests for serial correlation, and models direct and indirect effects. See appendix C for the underlying mathematics equations.

Watersheds separate distinct time periods of many vs. few CCs. Within a problem solving session, there might be fewer CCs at the start when people are trying to understand the problem than at the end when they are close to a solution. Hence, dividing the time series data into time periods with significantly more vs. fewer CCs allows testing of hypothesis H-2 (groups with more clusters of CCs are more likely to solve the problem correctly) and testing of predictors' different effects across time.

For each group, I used a modified version of the method outlined in Maddala and Kim (1998) based on information criteria to identify the watersheds in time (breakpoints) that divided each group's problem solving activity into distinct time periods. Conceptually, information criteria measure whether a model strikes a good balance between parsimony and goodness of fit. Unlike other information criteria, the Schwarz or Bayesian information criterion (BIC) provides a consistent estimator for the number of lagged variables in the true model (Grasa, 1989). Predicting the outcome variable, CC, I added locations of possible breakpoints as independent variables and computed the BIC for a simple univariate time-series model (an auto-regressive order 1 model). Assuming a given number of breakpoints (first 0 breaks, then 1 break, then 2 breaks, etc.), and using the model above, I calculated the BIC for all possible locations of those breakpoints in the time series. (For example, for one break, calculate the BIC if the break is between turn 1 and turn 2, then if it is between turn 2 and turn 3, etc.) This was done for all possible numbers of break points from 0 to 5. (Current microcomputers lack the computational speed to test more than five break points [six time periods]). The optimal model has the lowest BIC. Applying this method to each group yielded the number and locations of breakpoints (and hence time periods) for each group.

Then, I used a t-test to determine whether successful groups have more CC clusters/time periods than unsuccessful groups (hypothesis H-2). The small sample size limits the statistical power of this analysis to identify non-significant results (number of time periods = 72; power = 0.75 for an effect size of 0.3). Transcript segments to illustrate representative breakpoints.

Predicting CCs at the speaker turn level. I used a multi-level Logit variance components model (Goldstein, 1995; Bryk & Raudenbush, 1992) to test if the outcome variable, CC, significantly varied across groups or across time periods. Multi-level models separated unexplained error into speaker turn (level 1), time period (level 2), and group (level 3) variance components, thereby removing the correlation among error terms resulting from speaker turns nested within time periods within groups. If the variance components model showed significant variation at both the group and time period levels, then both the groups and time periods were heterogeneous. In that case, a 3-level model was needed.

Next, I entered the following independent variables. First, I added a vector of *s* classroom identification variables as control variables (**S**). As the likelihood ratio test for significance of additional explanatory variables was not reliable for this estimation method, Wald tests were used (Goldstein, 1995). Non-significant variables were removed from the specification.

Then, I entered *t* variables at the group level: correct group solution, mean of group members' mathematics grades, mean of group members' peer statuses, variance of mathematics grades, and variance of peer statuses (**T**). The last two variables tested the status effects hypothesis (H-3d). As with **S**, a Wald test was done on **T.** Then, I tested for interaction effects among pairs of significant variables in **T**. Non-significant variables and interactions were removed from the specification.

Next, I added *u* current speaker variables at the speaker turn level: gender, race, mathematics grade, peer status, correct evaluation, agree, politely disagree, rudely disagree, justify, question and command (**U**). Likewise, I applied the procedure for **T** to **U**. Then, I tested if the speaker turn level regression coefficients differed significantly at the time period or group levels (Goldstein, 1995). If yes, I kept these parameters in the model. Otherwise, I removed them.

Using a vector autoregression (VAR, Kennedy, 2004), I entered lag variables for the previous speakers, first lag 1 (indicating the previous turn and denoted –1), then at lag 2 (denoted –2), then at lag 3, and so on until none of the variables in the last lag were significant (lag 4 in this case). First, I added *v* previous speaker variables: gender (–1), race (–1), mathematics grade (–1), peer status (–1), correct evaluation (–1), agree (–1), politely disagree (–1), rudely disagree (–1), CC (–1), wrong contribution (–1), correct old idea (–1), justify (–1), question (–1), and command (–1) (**V**). As shown in Figure 1, these variables test the new ideas, argumentation, and rudeness hypotheses (3a, 3b, and 3c). I applied the procedure for **U** to **V**. Then, I repeated the procedure for lags –2, –3, and –4 of the variables in **V**. The parameters were estimated first with marginal quasi-likelihood, and these results served as starting values for predictive quasi-likelihood estimation (Goldstein, 1995).

All statistical tests used an alpha level of .05. Benjamini, Krieger, and Yekutieli's (2006) two-stage linear step-up procedure controlled the false discovery rate, as computer simulations showed that their procedure addressed this issue better than 13 other methods.

I used Ljung-Box (1979) Q-statistics to test for serial correlation (up to order 4) in the residuals for all 20 groups. If the residuals are serially correlated, then the parameter estimates are likely inefficient, and standard error estimates are likely biased (Kennedy, 2004). Then, the explanatory model must be modified with extra lagged outcome variables (lags of CC) or direct modeling of the serial correlation (see Goldstein, 1995, for details).

Based on the multilevel analysis results, the path analysis estimated the direct and indirect effects of the significant predictors separately to compute their total effects (Kennedy, 2004). As time constrains the direction of causality, I entered the explanatory variables in temporal order into the path analysis.

To aid interpretation of these results, I converted each predictor's total effect (direct plus indirect) to odds ratios, reported as the percentage increase or decrease (+X% or –X%) in a CC's likelihood (Kennedy, 2004). I repeated the above analyses with multi-level Probit to test if the results depended on the Logit distribution. I also estimated the predictive accuracy of the final model by comparing the final model's prediction of whether a CC occurred at each speaker turn in each group (y_{ijk}*) with the CC's actual presence or absence (y_{ijk}).

A multilevel analysis has multiple units of analysis, so the statistical power for each one (group, time period, speaker turn) must be computed separately. As noted earlier, the statistical power for groups and time periods are fairly low, so non-significant results at these levels must be interpreted cautiously. At the speaker turn level however, the sample size is 2,951, so the statistical power is over 0.99 even for a small effect size of 0.1. None of these units of analyses (speaker turn, time period, group, classroom, school, country) are necessarily representative, so results might differ in other contexts. As students can change behaviors during another student's speaker turn, modeling students as a level of analysis requires multivariate outcome, multilevel, cross-classification Logit/Probit, but no implementation of such a method has been shown at the publication time of this journal article.

Results

After reporting the preliminary results, I showed that groups with a larger percentage of CCs had higher solution scores. Then, I examined the differences across time periods, followed by the predictors of CCs at the speaker turn level. Due to space considerations, I include only the main results here; all results are available upon request.

Preliminary results

Of the 3,153 total speaker turns, 49 turns were not coded because of poor sound quality (see Appendix D, Table D1). The omitted turns did not significantly correlate with other variables, so they likely did not affect the results. As lag variables required data from preceding turns, 153 turns could not be used. Coding of each dimension showed high inter-rater reliability (see Appendix D, Table D2).

The summary statistics showed that correct contributions occurred only 20% of the time (see Table 1), more often in successful groups that solved the problem (26%) than in unsuccessful ones (16%; see Table 4). Moreover, successful groups often had higher mathematics grades than unsuccessful groups (91 vs. 82). Compared to unsuccessful groups, successful groups were more likely to evaluate ideas correctly and justify their ideas, and less likely to disagree rudely (38% vs. 24%; 18% vs. 13%; 6% vs. 12%)

Predicting solution score at the group level

As expected, the students found the problem difficult. Only 10 of the 20 groups solved it correctly, and every group made at least three mistakes. See Table 1 for overall summary statistics, Table 3 for summary statistics of successful groups that solved the problem vs. unsuccessful groups, and Appendix D Table D3 for the correlation matrix. All groups were reach-testers; no group used linear problem phases. This result suggests that the natural inclination of these students is to reach-test, similar to many adults (Pavitt & Johnson, 2001; Poole, 1981).

Table 3
Summary table of hypotheses, data, model and theories regarding correct contributions (CCs)

Hypothesis	Data	Model of predictors	Theories
H-1. Groups with more CCs are more likely to solve the problem correctly.	Final group answers; Student characteristics; Computed group characteristics; Summary statistics of variables coded from 2,951 turns of transcripts	Math grade (mean vs. highest), ratio of CCs over total group turns. Other variables: peer status, math grade variance, peer status variance, words, on-task words	Functional group decision-making (Orlitzky & Hirokawa)
H-2. CCs occur in more clusters in successful groups.	Final group answers; Each group's time periods of high vs. low CCs	T-test of differences in time periods in successful vs. unsuccessful groups	Phases vs. Reach-tester (Ellis & Fisher vs. Pavitt & Johnson)
H-3a. Correct and wrong contributions aid CC creation	Final answers to algebra problem; Student characteristics; Group characteristics; Variables coded from 2,951 turns of transcripts;	CC (–i), wrong contribution (–i), for i = 1..4 [a]	Functional group decision-making (Orlitzky & Hirokawa)
H-3b. Correct evaluations, questions, and justifications aid CC creation		Correct evaluation (–i), question (–i), justification (–i) for i = 0..4 [a]	Socio-cognitive conflict (Doise, Mugny & Perret-Clermont; Piaget)
H-3c. Polite disagreements aid CC creation but rude disagreements, false agreements, and commands hinder them.		Politely disagree (–i), rudely disagree (–i), agree (–i), command (–i) for i = 1..4 [a]	Politeness (Brown & Levinson)
H-3d. Greater status differences hinder CC creation		Variances of Math grade, peer status, gender, and race [a]	Status characteristics (Cohen)

[a] Variables in full model. Classroom identification variables: Group level variables: correct solution, mean math grade, mean peer status, math grade variance, peer status variance, gender variance, race variance. Current speaker (0) variables: gender, race, math grade, peer status, correct evaluation, agree, politely disagree, rudely disagree, justify, question and command. Previous speakers' lag variables (i = 1..4): gender (–i), race (–i), math grade (–i), peer status (–i), correct evaluation (–i), agree (–i), politely disagree (–i), rudely disagree (–i), CC (–i), wrong contribution (–i), correct old idea (–i), justify (–i), question (–i), command (–i)

Groups with higher mean mathematics grades or a greater percentage of CCs had higher solution scores, supporting hypothesis H-1 (see Table 5). When mean mathematics grade and highest mathematics grade were both entered, only mean mathematics grade was significant ($\beta = 0.16$, $SE = .05$, $p < .05$, McFadden's $R^2 = .24$). None of the other predictors were significantly related to solution score. Mean mathematics grade also predicted percentage of CCs (β = .012, *SE* = .005, $p < .05$, $R^2 = .20$). Replacing percentage of CC with CC frequency or CC over new ideas ratio yielded similar results but explained less variance.

Table 4
Summary table of speaker turn variables for successful and unsuccessful groups

	Overall		Successful		Unsuccessful			
Speaker turn level variable	Mean	SD	Mean	SD	Mean	SD	Min.	Max.
Unsolved [a]	0.50	0.50	0	0	1	0	0	1
Correct contribution	0.20	0.40	0.26	0.44	0.16	0.37	0	1
Before problem solving								
Girl	0.47	0.50	0.42	0.49	0.50	0.50	0	1
Asian	0.15	0.35	0.13	0.34	0.16	0.36	0	1
Latino	0.25	0.43	0.31	0.46	0.21	0.41	0	1
White	0.26	0.44	0.20	0.40	0.30	0.46	0	1
Mathematics grade	86	10	91	8	82	10	64	99
Peer status	6	6	8	7	6	5	2	9
During problem solving								
New ideas								
Wrong contribution	0.10	0.30	0.10	0.31	0.09	0.29	0	1
Argumentation								
Correct evaluation	0.30	0.46	0.38	0.49	0.24	0.43	0	1
Polite disagreement	0.18	0.39	0.19	0.39	0.18	0.39	0	1
Ignore / Unresponsive	0.17	0.37	0.12	0.33	0.20	0.40	0	1
Question	0.24	0.43	0.27	0.44	0.22	0.42	0	1
Justification	0.15	0.36	0.18	0.38	0.13	0.33	0	1
Face and Rudeness								
Rude disagreement	0.10	0.29	0.06	0.23	0.12	0.33	0	1
Agreement	0.56	0.50	0.63	0.48	0.50	0.50	0	1
Command	0.07	0.26	0.10	0.30	0.05	0.22	0	1

[a] Separate analyses for groups with each solution score showed substantial differences between groups that did and did not solve the problem correctly, and similar results across the latter unsuccessful groups. Thus, *unsolved* was coded as a binary variable (0 or 1) in the turn-level analysis to facilitate interpretation of the results.

Table 5.
Significant, unstandardized parameter coefficients of hierarchical set ordered Logit regressions predicting solution score (with standard errors in parentheses)

	2 Ordered Logit Regressions predicting solution scores	
Predictor	Model 1	Model 2
Mathematics Grade	0.256 **	0.217 *
	(0.089)	(0.095)
% Correct Contributions		11.377 *
		(5.033)
McFadden's R2	0.228	0.377

Note. *Significant constant term is omitted. *p < .05, **p < .01, ***p< .001*

Watersheds identifying distinct time periods of many vs. few CCs

CCs did not vary across classrooms or across groups, but CCs varied across time periods and across speaker turns. The classroom identification variables did not significantly predict CCs, so the prevalence of CCs did not differ across classrooms. The variance components model showed that the likelihood of CCs did not vary significantly across groups (M = 0.000, SE = 0.001), but CCs did vary significantly across time-periods (M = 3.457, SE = 0.682) and speaker turns (M = 0.908, SE = 0.024). On average, successful groups produced more CCs than unsuccessful groups. However, they did not do so consistently, as a CC's likelihood differed substantially across time periods within a group. Some time periods had many CCs while other time periods had few CCs. Hence the likelihood of CCs differed mostly across time periods, not across groups. Of the total variance, less than 0.1% of the variance occurred at the group or classroom level, 79% was across time periods while 21% was within time periods. (The high variance of CCs across time periods [79%] also showed that the breakpoint method effectively identified distinct time periods.) As the variance of CCs across groups was not significant, a 2-level model (time periods and turns) with group interaction terms was used.

CCs occurred in similar numbers of clusters in both successful and unsuccessful groups, so the results did not support hypothesis H-2. See figures 2 and 3, which show each group's time periods of high CCs vs. low CCs. The time periods for each group ranged from one to five. The number of time periods did not differ significantly in successful vs. unsuccessful groups (successful: M = 3.3, SD = 2.1; unsuccessful: M = 3.9, SD = 1.7; t-test = 0.702; $p > .05$).

An exploratory analysis of the time periods showed that only three groups had consistent CC production rates. Two groups that consistently produced CCs at a *high rate* exceeding 50% successfully solved the problem. Meanwhile one group that consistently produced CCs at a *low rate* below 20% failed to solve the problem. Otherwise, CCs clustered in 17 of the 20 groups. Seven of eight groups that started with a high CC rate successfully solved the problem, and all eight groups that ended with a high CC rate successfully solved the problem. Eight of ten groups that ended with low CC rates failed to solve the problem.

As shown in the examples below, an exploratory classification of the 52 breakpoints suggests 3 broad categories: off-task ↔ on-task transitions, insights, and critical errors. At 26 of the breakpoints (8 in successful groups, 18 in unsuccessful groups), groups transitioned from off-task to on-task or vice-versa. At 14 of the breakpoints (8, 6), a group member had an insight and their CCs increased sharply. At the remaining 12 breakpoints (7, 5), group members made a critical error, and their CCs fell sharply.

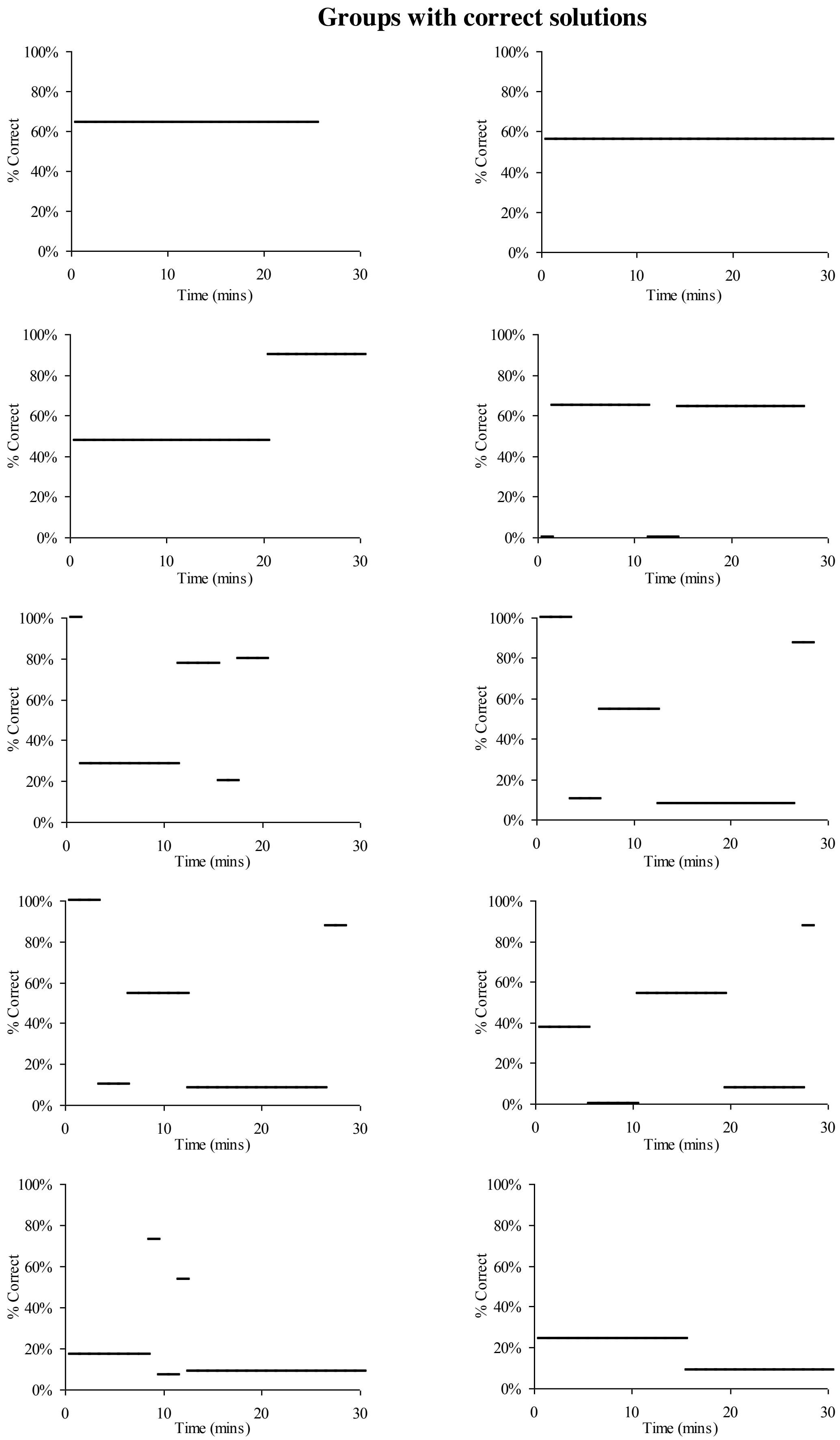

Figure 2. Line graphs of the percentage of new ideas that were correct in each time period for groups with correct solutions. Each line segment indicates a distinct time period.

Groups with incorrect solutions (Score)

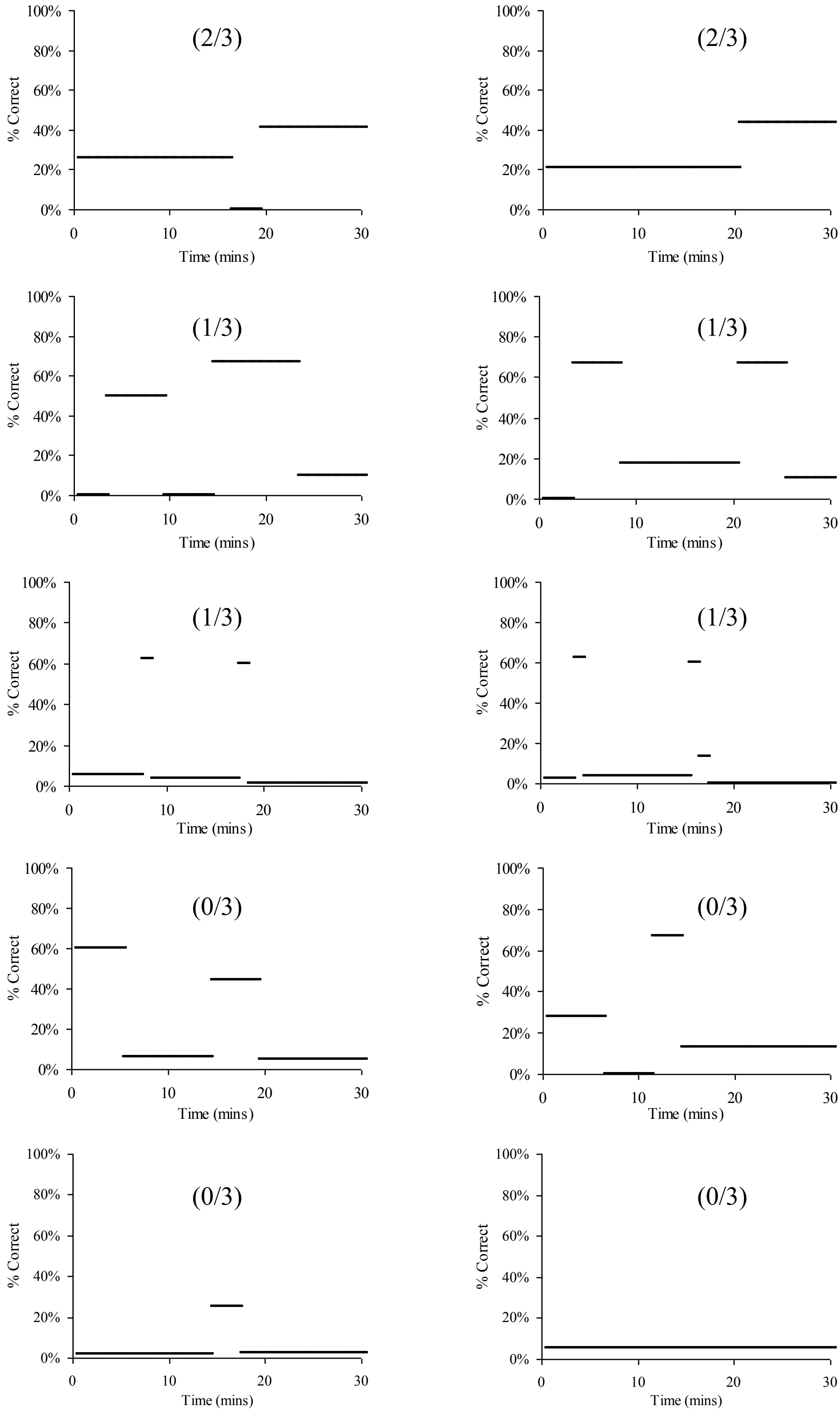

Figure 3. Line graphs of the percentage of new ideas that were correct in each time period for groups with incorrect solutions (with scores in parentheses). Each line segment indicates a distinct time period.

The following transcript examples illustrate the different types of breakpoints that separate time periods of many vs. few CCs. At the following breakpoint at turn seven, the group moved from an off-task conversation to work on the problem. (All names are pseudonyms.)

Turn	Person	Talk and/or Action
1	Jim	I'm going to Idaho with my family.
2	Bob	We went to Maryland last summer.
3	Jim	I don't want to go there.
4	Pat	What's the distance between New York and London?
5	Bob	Oh, thirty five hundred. We're going; we're going helicopter.
6	Pat	Something like three thousand five hundred.
7	Jim	I really think I just don't want to go. Okay, so the boat, how fast is the cruise going? It's twenty two miles per hour? That's over five miles, so five hours, so five times twenty two would be, wait.
8	Tim	[raises hand] Ms. T______ [teacher's name]
9	Jim	They're only, they're only really like, luckily, a hundred and ten miles out.
10	Bob	If they go ninety miles an hour, they can get there in less than two hours.

Jim and Bob discussed their travels before Pat asks a question about the problem ("What's the distance between New York and London?"). Although not relevant to the solution, this question drew the students' attention to the problem. Bob projected himself into the problem situation ("we're going, we're going helicopter") after answering Pat's question ("thirty-five hundred"). After Pat acknowledged Bob's answer, Jim concluded his off-task thoughts ("I really think I just don't want to go") and began working on the problem ("Okay, so the boat, how fast …"). Jim's proposed multiplication in the breakpoint turn seven ("five times twenty-two) was the first CC in their group. After Tim asked for the teacher, Jim and Bob began a series of CCs ("a hundred and ten miles out" and "they can get there in less than two hours"). So, this breakpoint indicated a change from off-task to on-task behavior at the first CC. (To distinguish on-task time periods from off-task time periods, the breakpoint method can be used on an on-task [vs. off-task] outcome variable to yield an on-task breakpoint at turn four.)

Breakpoints also occurred at major insights. In the following example, the students did not make much progress until a student drew a diagram.

61	Rex	One-ten and ninety is [hits calculator keys, 110 + 90] two hundred
62	Amy	What's two hundred?
63	Rex	Two hundred miles?
64	Liz	It's five hours, so [hits calculator keys, 90 x 5] Four hundred and fifty?
65	Amy	[draws] Ok, this is like. Okay, so like, ok, this [points to drawing] is New York, right? And that's London [points to drawing].
66	Liz	Right.
67	Amy	[pointing to drawing] Okay, okay um, that's the cruise ship. Ok. And like the cruise ship is ahead of the helicopter, right?
68	Liz	Yeah. At a hundred and ten.
69	Amy	Okay, [writes 110 near the cruise ship symbol], the helicopter's moving up.
70	Max	That's a helicopter?
71	Amy	Well, [raises open hands] I can't draw [laughs]
72	Max	[laughs] Alright.
73	Amy	Okay? You've got to think about the time. We have ninety–

74 Liz –We have this um to deal with this [points to cruise ship on drawing] too because it's not gonna stop.
75 Amy Oh, the cruise ship's not gonna stop.

Rex and Liz had been adding, multiplying, and dividing several numbers from the problem (5, 22, 90) without making much progress. After Amy's questioning of Rex ("what's two hundred?") yielded an uncertain answer ("two hundred miles?"), Amy drew a diagram of the problem situation, correctly modeling the ship's location, the helicopter's location and its movement (a CC in the breakpoint turn 67). Liz elaborated the diagram with the distance of the cruise ship from shore ("at a hundred and ten"). After some friendly teasing about the quality of the drawing, Amy highlighted the time traveled by the helicopter. Then, Liz interrupted with a CC about the cruise ship's continuing motion ("it's not gonna stop"). Amy validated Liz's idea, and the group then computed each vehicle's movements and marked their new locations on the diagram to solve the problem. In short, the diagram was a breakpoint that ignited several CCs by helping these students model changes in the problem situation rather than simply trying different computations.

Breakpoints also occurred at critical errors. In this example, the group recognized that the helicopter and the ship both moved and tried to compute their movements.

91 Bob [hits calculator keys 15 x 50 = 750]
92 Ben What is --what is seventy thousand – seven hundred-fifty mean?
93 Lex I think the wrong things got multiplied; try ninety and one point five.
94 Bob [laughs, hits calculator keys 1.5 x 90 = 135] It's one thirty-five.
95 Jim I don't know. I'm telling you, you gotta, what you gotta do is divide thirty-three miles into that {helicopter speed?}. Cause twenty-two plus eleven. Thirty-three.
96 Bob Thirty-three added onto a hundred and ten [writes 33 + 110 =].
97 Jim Thirty-three.
98 Bob Then add.
99 Jim Why are you adding? No. Just listen. Okay. Ninety. Nine, okay, no, okay, one-thirty-three, ninety into one thirty-three. Goes in, um, once. Gives me forty-seven?
100 Ben [laughs]
101 Bob [laughs]
102 Lex [laughs]
103 Jim Forty-seven? Huh?
104 Bob Forty-seven?
105 Jim Yeah. Forty-seven, so. Um, goes in, whad da da. {"whad da da" shows confusion}
106 Bob Try five hours?
107 Jim Five? Maybe one more.

After Lex noted that Bob hit the wrong calculator keys, Bob laughed and corrected his computation of the helicopter's distance from shore after an hour and a half. At the breakpoint turn 95, Jim incorrectly disagreed and mistakenly suggested dividing the helicopter speed into the distance traveled by the ship in an hour and a half (33) rather than into the remaining distance between the two vehicles (= 9 = 143 –135; 143 = 33 + 110). When Bob suggested adding 33 and 110, Jim re-asserted his position ("thirty-three"), challenged Bob ("why are you adding?"), ignored Bob ("no, just listen"), changed his mind, and followed Bob's suggestion but added incorrectly ("no, okay, one-thirty-three"). Jim compounded this arithmetic error with further division errors to compute 133 ÷ 90 = 47 rather than 1.47 (actually, 1.4777). Surprised by the result, Jim did not know how to proceed ("Forty-seven, so. Um, goes in, whad da da"). Bob suggested using the five hours from the problem, and Jim decided to add one more hour. Neither these ideas nor those for the next few minutes were productive. So, this breakpoint indicated a critical error, changing a productive time period with many CCs to an unproductive time period with few CCs.

Predicting CCs at the speaker turn level

The explanatory model at the speaker turn level showed that wrong, new ideas, argumentation, and politeness affected the likelihood of CCs (see Figure 4 for the path analysis; Table D4 for the correlation-covariance matrix; Table D5 for the multilevel logit regression results). However, status differences were not linked to CCs, and CCs did not predict subsequent CCs.

New ideas. These problem-solving sessions had few chain reactions of CCs, as a CC did not help create a subsequent CC. However, a wrong contribution (–1, in the previous turn) was 4% more likely to yield a CC (see Figure 4; +4%: 19% →23%; after a turn *without* a wrong, new idea, a CC occurred 19% of the time; after a turn *with* a wrong, new idea, a CC occurred 23% of the time; see Appendix C for computation details.) Thus, these results only partially supported hypothesis H-3a. In response to wrong contributions, group members were more likely to rudely disagree (+7%: 9% →16%) and less likely to agree (–17%: 57% →40%), suggesting that they often recognized the flaws of wrong, new ideas. In the following segment, a student incorrectly multiplied the helicopter speed by five hours.

Amy: In five hours, multiply [enters 90 x 5 on calculator]
Four hundred and fifty in five hours.

Rex: Four fifty? That can't be right 'cause the cruise ship is only at one-ten. Oh! Oh!
The helicopter leaves later! Multiply by two hours! Multiply by two hours!

Rex recognized that the outcome was wrong ("that can't be right") because the helicopter would have passed the target cruise ship ("is only at one-ten"). This error helped Rex detect and correct the flaw in the number of hours from five to two ("Oh! The helicopter leaves later! Multiply by two hours!"). Building on Amy's partially correct idea to multiply the helicopter speed by five hours, Rex created a CC.

Argumentation. Correct evaluations and justifications helped create CCs, but questions did not (partially supporting hypothesis H-3b). Correct evaluations from any of the three previous speakers (–1, –2, –3) were more likely to yield a CC (+2%, +5%, and +3%, respectively; 19% → 21%; 19% →24%; 19% → 22%). Each correct evaluation also yielded more subsequent correct evaluations, both in the next turn and in the following turn (+12%: 27% →39% for both cases). These correct evaluations made a subsequent CC more likely in part through more justifications (+3%: 15% →18%). Correct evaluations also made wrong contributions less likely (–3%: 10% → 7%) and agreements more likely (+10%: 51% → 61%).

Justifications increased a CC's likelihood in all groups, though more in successful groups (+68%: 14%→82%) than in unsuccessful groups (+29%: 12% →41%), possibly because the quality of their justifications differed. In successful groups, members often referred to mathematics relationships to justify their ideas.

Ian: Twenty-two plus five is–

Jo: –we're doing times five 'cause it's rate times time.

When Jo corrected Ian, she referred to the "rate times time" (equals distance) formula.

In unsuccessful groups however, students often justified their claims by citing authority (e.g., teacher, textbook, problem statement).

May: Ninety times five is four-fifty.

Kit: Ninety times–

May: – five hours because Ms. T [teacher] said five hours.

Unlike Jo, May justified her computation by incorrectly citing the teacher, "because Ms. T [teacher] said five hours." Justifications based on mathematics might be more valid, relevant, and helpful to group members than those based on authority, which might explain why justifications had larger effects on CCs in successful groups than in unsuccessful groups.

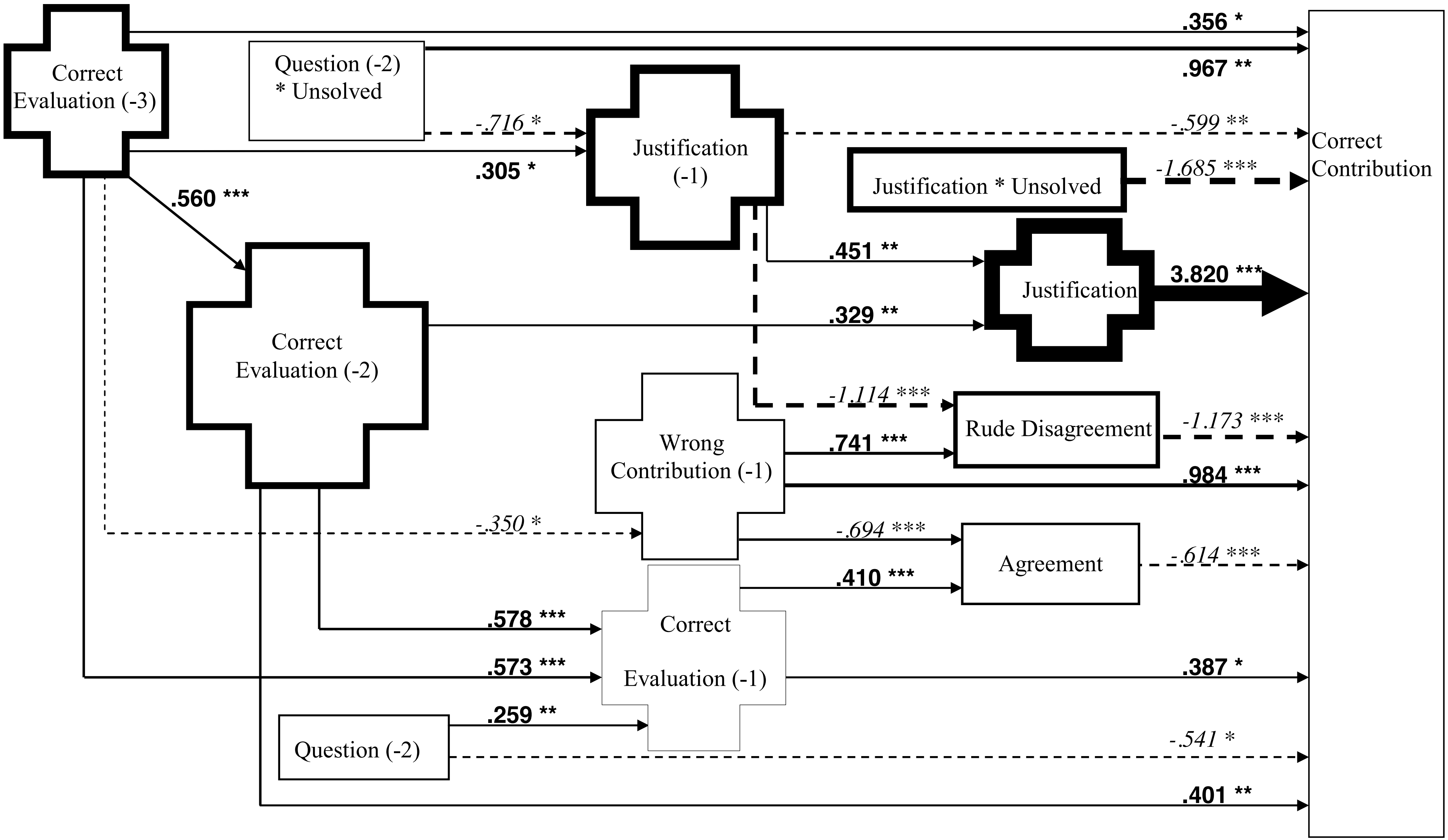

Figure 4. Path analysis of significant predictors of CCs using multilevel Logit. Negative numbers in parentheses (-1, -2, -3) indicate actions that occurred one, two, or three turns ago. Values are standardized parameter coefficients. Crosses (✚) indicate positive overall effects, while rectangles (▫) indicate negative overall effects. Solid arrows (→) indicate positive direct effects, while dashed arrows (- - >) indicate negative direct effects. Thicker lines indicate larger effects. For example, a correct evaluation in the previous turn (-1) has a direct effect on a correct contribution of +0.387 while its indirect effect is –0.252 (= 0.410 x –0.614), yielding a total effect of +0.135. “Unsolved” refers to speakers in groups that did not correctly solve the problem.

Justifications (–1) also raised the likelihoods of a CC (+36%: 15% → 51%) and a subsequent justification (+4%: 14% →18%) while reducing that of a rude disagreement (–5%: 11% → 6%).

Questions (–2) reduced a CC's likelihood (–2%: 21% → 19%). This result suggests that questions typically indicated individual knowledge gaps. As such, other group members could give an immediate, known explanation, which did not stimulate a CC.

Xiao: Why do we multiply twenty-two times five hours?
Ron: Rate times time is how far the ship moves.
Xiao: Oh! Rate times time. Ok.

When Xiao questioned a computation, Ron justified its validity via the rate-time-distance relationship. John understood ("oh!") and accepted it ("ok").

In unsuccessful groups however, a question (–2) further reduced the likelihoods of a justification (–1) and of a CC (respectively, –2%: 14.6% →12.6%; –4%: 17% →13%). Consider this segment.

Beth: Why ninety times five?
Mark: That's what the problem said.
Beth: No, it didn't. It didn't say do ninety times five.

Mark answered Beth's question by referring to the problem statement, "that's what the problem said." Not satisfied with that answer, Beth challenged Mark with a rude disagreement, "no, it didn't," declaring that Mark was wrong as the problem statement did not specify that multiplication, "it didn't say do ninety times five." In short, after a question, unsatisfactory responses might help account for CCs being less likely in unsuccessful groups than in successful groups.

Face, Rudeness, and Status. Polite disagreements increased a CC's likelihood, supporting hypotheses H-3b (+14%; 13% →27%). Consider the transcript segment of Ian and Jo again.

Ian: Twenty-two plus five is–
Jo: –we're doing times five 'cause it's rate times time.

Jo redresses her disagreement by shared positioning, "we," before justifying her disagreement with Ian's idea.

In contrast, rude disagreements reduced a CC's likelihood, also supporting hypothesis H-3c (–4%: 21% → 17%; [4, 2]). Consider the following example.

Eva: Ninety times five, four-fifty.
Ada: That's wrong.
Eva: No, it's not.

Although Eva's idea is arithmetically correct, the computation was not consistent with the problem situation and hence, conceptually incorrect. Ada rudely disagreed with Eva's computation without explaining, "that's wrong." In response, Eva retaliated with a rude disagreement, "no, it's not." Thus, rude arguments hindered creation of CCs, while polite ones aided their creation.

Controlling for correct evaluations, agreements reduced a CC's likelihood, also supporting hypothesis H-3c (–5%: 23% → 18%; [4, 1]). These agreements were often simple confirmations such as "yep" or repetitions such as "one ten, right."

Lana: Uh, ninety times five, four-fifty.
Jack: Uh-huh.

Students like Jack often gave brief confirmations, "uh-huh," without further elaboration or CCs, suggesting that he might have used false agreements to build social relationships by sacrificing problem solving progress. Meanwhile, commands did not significantly affect CC creation.

Status and other effects. None of the other predictors were significant. In particular, larger status differences did not affect CCs (no support for hypothesis H-3d). Neither group mathematics grade nor individual mathematics grade were significant either. Although mean mathematics grade positively correlated with group percentage of CCs in the group level analysis, that group level analysis omitted 99.9% of the CC variance.

Only two predictors showed different effects across time periods (agree and correct evaluation [–2], for detailed regression results, see Appendix D, Table D5). Agreements reduced a CC's likelihood (mean = –5%: 23% → 18%), with the effect varying across time periods from –3% to –21%. Correct evaluation (–2) raised a CC's likelihood (mean = +5%: 19% → 24%), with the effect varying across time periods from –0.3% to +9%. The varying effect sizes of agreement and correct evaluations across time periods suggested that their effects were moderated by unexamined variables that differed across these time contexts. Aside from justifications, agreements, and correct evaluations (–2), the effects of all other predictors did not differ significantly across time periods or across groups. Hence, the other predictors showed no evidence of contextual effects and are candidates for broader, possibly universal effects.

This model had an 84% accuracy rate for predicting whether a CC occurred in any given turn (y_{ijk}* vs. y_{ijk}). Furthermore, the Q-statistics run on the final model showed no significant serial correlation of residuals in any of the twenty groups. So, the time-series model was likely appropriate.

Discussion

Past theoretical models have highlighted the importance of correct, new ideas (correct contributions or CCs) to group problem solving (e.g., Chiu, 2000a, 2001; Hinsz, Tinsdale & Vollrath, 1997). By understanding how group processes help or hinder CC creation, educators can help students engage in beneficial group processes and avoid harmful ones. This study replicated past research by showing that groups with more CCs were more likely to solve the problem successfully. More important, this study extended this line of research in three ways. First, I showed how satisfactory responses to questions and justifications yielded more CCs in groups that successfully solved the problem. Second, I statistically identified three types of crucial events (breakpoints) that divided each group's problem solving into distinct time periods with more CCs vs. fewer CCs. Third, I showed how specific group problem solving processes in the micro-time context helped create CCs.

Differences among successful and unsuccessful groups

Groups that successfully solved the problem were more likely to respond to questions with justifications, showed stronger justification effects on CCs, had higher mathematics grades and had proportionately more CCs. Successful groups often answered their members' questions with satisfactory explanations. Although these explanations did not immediately help create CCs, they helped build partially shared understandings in the problem content space and social solidarity in the social relational space, both of which might have enhanced the micro-time context to help create CCs later. In unsuccessful groups however, inadequate responses to questions were more often rebuffed with rude disagreements, which in turn hindered creation of CCs.

Justifications showed stronger positive effects on CCs in successful groups, in part because these groups used more justifications that referred to mathematics relationships. In contrast, unsuccessful groups used more authority-based justifications, which were less helpful in yielding CCs.

In general, groups that had higher mathematics grades or more CCs were more likely to solve the problem when examining less than 0.1% of the CC variance at the group level. After including the remaining 99.9% of the CC variance at the speaker turn level however, group mathematics grades did not significantly affect CC creation. Thus, these results highlight the importance of also analyzing group processes at the speaker turn level, not only at the group level. In short, groups that answered one another's questions with explanations or that often used justifications (especially mathematically-based ones) created more CCs and were more likely to solve the problem successfully.

Clusters of CCs in time periods separated by breakpoints

Dynamic multilevel analysis's breakpoint estimation method identified watersheds that altered group problem solving processes and their effects on CC creation. 85% of these groups did not uniformly created CCs at random intervals throughout their activity. Instead, watersheds divided their problem solving activity into distinct time periods of many CCs vs. few CCs. In contrast to hypothesis H-2, CCs clustered in most groups, regardless of their problem solving outcome.

An exploratory analysis of the breakpoints between time periods suggested three types of watersheds: off-task ↔ on-task transitions, insights, and critical errors. In half of the breakpoints, groups switched between primarily on-task vs. off-task time periods. During about one quarter of the breakpoints, a group member had an insight that led to many more subsequent CCs. In the remaining breakpoints, a group member made a critical error that sharply reduced the number of subsequent CCs.

Most group processes had similar effects on CC creation across time periods, showing no evidence of context-dependent effects. Only agreement and correct evaluations (–2) had different effects on CCs across time periods, suggesting that one or more unexamined variables might moderate their effects across these different time contexts.

Predicting CCs at the speaker turn level

Group members' recent actions (micro-time context) affected CC creation. Consistent with prior research, correct evaluations, justifications, and disagreements aided CC creation (e.g., Barron, 2003; Cobb, 1995). This study extended this line of research by showing (a) the effects of different types of new ideas, (b) the effect sizes of different aspects of argumentation, (c) the durations of argumentation effects, and (d) the effects of face and rudeness.

New ideas. Group members' wrong contributions aided CC creation, but CCs did not aid creation of subsequent CCs, partially supporting hypothesis H-3a. After a wrong contribution, group members were more likely to disagree, suggesting that group members often detected and corrected flaws to create a CC. Thus, serving as grist for CCs outweighed the danger of accepting wrong ideas. In contrast, a CC did not aid subsequent CC creation, possibly because they were not necessarily recognized as correct. Together, these results showed that wrong contributions were *more* important than correct ones for creating CCs in this study.

Argumentation. Correct evaluations and justifications both immediately aided CC creation while questions did not, partially supporting hypothesis H-3b. Correct evaluations had long lasting effects, helping create CCs over the next three speaker turns. Hence, recognizing as ideas as correct or flawed helps group members build on them accordingly (Barron, 2003). Correct evaluations also facilitated subsequent correct evaluations, justifications, and agreements while yielding fewer wrong contributions. Together, these results support the view that correct evaluations helped creating a valid basis of partially shared understandings in the problem content space for creating CCs.

Justifications had the largest effect on CCs (+68% in successful groups and +29% in unsuccessful groups). Furthermore, the effect of two consecutive justifications on creating CCs was larger than the combined effects of all other predictors. Justifications both elicited further justifications in the problem content space and reduced rude disagreements in the social relational space, thereby facilitating calmer, reason-based discussions that helped create CCs both immediately and in the future.

In contrast, questions yielded fewer CCs, especially in unsuccessful groups. In this study, questions often identified individual knowledge gaps. Thus, other group members could answer these questions with previously discussed ideas, reducing the likelihood of immediately creating a CC. Failing to respond with satisfactory explanations led to more rude disagreements, and eventually, fewer CCs, especially in unsuccessful group. Hence, answering group members' questions satisfactorily likely built partially shared understandings in the problem content space and solidarity in the social relational space, both of which likely helped create future CCs and a correct problem solution.

Face and rudeness. Polite disagreements created more CCs, but rude disagreements and agreements yielded fewer CCs, supporting hypothesis H-3c. These results were consistent with the view that polite disagreements reduced interpersonal conflict, aided understanding of criticisms and fostered CCs. Meanwhile, the results are also consistent with the view that rude disagreements escalated face threats and hindered creation of CCs (Chiu & Khoo, 2003).

Agreements also yielded fewer CCs, suggesting that students had substantial face concerns. Specifically, students' social motives might have inclined them to prefer agreements, sometimes reflexively with simple confirmations (Burgoon, Dillman & Stern, 1993; Chiu, 2001). This result is consistent with that of Chiu and Khoo's (2003) study showing that people tend to agree excessively after controlling for correctness of the previous speaker's idea. In these cases, students sacrificed progress in the problem content space for progress in the social relational space. Together, the disagreement and agreement results support the view that using *politeness* theory to modify *socio-cognitive conflict* theory yields more precise models of group problem solving processes (Brown & Levinson, 1987; Piaget, 1985).

Implications for researchers

This study modeled conceptual relationships among group processes affecting the *processes* of creating CCs and applied new methods for analyzing them. Due to the small number of groups, the data are not necessarily representative of group interactions. If future studies show similar findings, these results have the following implications for researchers, teachers, and students.

There are four implications for researchers. First, CC creation differed mostly due to the micro-time context; group and classroom differences accounted for less than 0.1% of the CC variance. To model successful and unsuccessful group processes more precisely, researchers can focus on group members' actions, not only on their group or individual characteristics.

Second, group processes often differed across time. Researchers can model group processes more accurately by analyzing group processes at various times during an activity. Specifically, watersheds (breakpoints) might radically alter group processes, dividing the activity into distinct time periods. Exploratory analyses of these breakpoints suggested three major categories that future researchers can elaborate or expand (on-task ↔ off-task transitions, insights, or critical errors). Across time periods, relationships among group processes (or between group processes and outcomes) might remain the same or change substantially.

Third, this study highlighted the temporal micro-development of students' interactions (Mercer, 2008), showing how the micro-time context influenced group processes. Specifically, it showed how sequences of actions and interactions by the three most recent speakers affected the problem content space, the social relational space, and the creation of CCs. In addition to an activity's broader macro-context, researchers can develop better understanding of group processes by modeling the micro-time contexts and their effects on group processes.

Lastly, this study showcased a method for systematic, fine-grained analyses of individual or group processes, dynamic multi-level analysis (DMA). Specifically, DMA statistically identified breakpoints (and their respective time periods) and modeled individual actions or social interactions over time (Chiu & Khoo, 2005). The breakpoints are watersheds that altered group processes, dividing the session into distinct time periods. Meanwhile, the relational variables across speaker turns, multi-level Logit/Probit, lag variables, path analyses, and serial correlation tests modeled social interactions within micro-contexts of time, as well as both group and time period differences. Furthermore, DMA modeled explanatory variables at the group level, time period level, and speaker turn level simultaneously. In addition to estimating effect sizes and effect durations of explanatory variables, DMA also identified differences in effects across groups or across time periods, thereby locating possible moderator effects of unexamined variables at the group or time period levels.

Implications for teachers and students

The results suggest that teachers can help create classroom cultural practices to facilitate desirable group interactions. Specifically, teachers can promote a mutually respectful, supportive, accountable, safe, and reflective classroom culture. When students are mutually respectful, supportive, and accountable to one another, they are more likely to answer group members' questions to increase their partially shared understandings. A safe environment reduces students' concerns over loss of face and aids their free expression of new ideas, including wrong ones. Furthermore, an accountable, classroom culture facilitates frequent justifications of ideas. As generalized justifications were more beneficial than appeals to authority, mathematics teachers can help their students develop mathematics norms of discourse so that students can propose, evaluate, and justify their ideas more effectively against mathematics standards of reasoning (Balacheff, 1988; Sfard, 2007). By learning these norms, students can develop mathematics eyes to view the world and acquire and communicate mathematics' structural relationships more easily (Blanton & Kaput, 2003; Franke, et al, 2007).

This study also showed that correct evaluations had long-lasting benefits. This result suggests that a supportive, accountable, and reflective classroom culture can help students evaluate one another's ideas carefully without impulsive confirmations or rude rejections. If the above results are supported by future studies, these changes in classroom culture might help students realize the potential benefits of cooperative learning.

Limitations and Future research

This study's limitations include its small sample sizes of higher level units (groups, schools, countries), limited problem content, setting, and methodological assumptions. Due to the small number of time periods, groups, teachers, and classrooms, the data were not necessarily representative of group interactions in classrooms. Likewise, the results might differ across schools or across countries. Furthermore, these students were not accustomed to working together in groups, so students with substantial experience working together might behave differently. Likewise, these results might not apply to students who do not know each other well (e.g., during the first day of class in a new school). These same students might also behave differently during discussions of different mathematics problems (e.g., geometry), let alone problems in other subjects like history. Furthermore, these students might behave differently in other settings outside of school (e.g., home or playground).

DMA relies on two primary assumptions and requires a minimum sample size. Like other regressions, DMA assumes a linear combination of explanatory variables and independent, identically distributed residuals. (Non-linear functions can be modeled as individual or multiple explanatory variables, e.g., age^2.) DMA also has modest sample size requirements. Green (1991) proposed the following heuristic sample size, *N*, for a multiple regression with *M* explanatory variables and an expected explained variance R^2 of the outcome variable:

$$N > 8 * (1 - R^2) / R^2 + M - 1 \qquad (1)$$

For a large model of 20 explanatory variables with a small expected R^2 of 0.10, the required sample size is 91 speaker turns (= 8*(1 – 0.10)/0.10 + 20 – 1). Less data is needed for a larger expected R^2 or for smaller models. In practice, two groups of students talking for half an hour will often yield over 100 speaker turns, sufficient for DMA. Thus, researchers can analyze seemingly "qualitative" data sets through both qualitative and quantitative methods (e.g., both traditional discourse analysis and DMA's statistical discourse analysis).

In addition to addressing the above limitations, future research can use DMA to model individual or group processes by asking questions of the form: what affects the likelihood of an action at each moment in time? Consider the following research questions. What sequences of recent teacher or student actions facilitate student use of a specific strategy? What influences development of a student's science language use (register) over five lessons? How much do these effects differ (if at all) across different people, time periods, artifacts, activities, etc.?

In general, one can apply DMA to predict people's behaviors with diverse explanatory variables. Researchers can use DMA to examine individual learning/problem solving (e.g., with protocol analysis data, Ericsson, 2001). Or, they can examine verbal and non-verbal interactions among people (students, teachers, parents, computer avatars, therapists, etc.). DMA can be used to predict specific behaviors or their properties (e.g., decisions, gaze, use of metacognitive strategies, vocabulary, etc.). Possible predictors include characteristics of the following: recent events, the time period, the individual, group members, artifacts (e.g., graphs), activity (e.g., discussing a poem), broader contextual factors (e.g., setting), historical factors (e.g., notable school events), or interactions among them.

Careful preparation is needed to perform a DMA. Before doing a DMA, a researcher should clearly delineate outcome variables, explanatory variables, and units of time (e.g., speaker turn). Furthermore, the sample size of time units should be sufficient (see Equation 1). If coding of variables is needed, the researcher then uses or creates a coding framework to yield sufficiently high inter-rater reliability (compute via Krippendorf's alpha). Then, follow the procedure in the methods section: (a) estimate breakpoints and time periods, (b) compute the variance components to identify the number of levels, (c) add predictors, (e) test for serial correlation, and (f) estimate direct and indirect effects in a path analysis. For details and further suggestions, see Chiu and Khoo (2005).

DMA can be simplified to omit the breakpoint estimation, extended to include multiple outcomes, or modified to perform a meta-analysis. If the data is naturally divided into time periods, breakpoint estimation might not be needed (though it could confirm the validity of the division of time periods; Chiu & Khoo, 2005). Also, breakpoint estimation requires only specification of the outcome variable and can be done separately without the other DMA components (Chiu & Khoo, 2005). Although the current study tested only one outcome variable, multiple simultaneous outcome variables can also be modeled by adding the outcome variables at the lowest level of the nested data structure (for details, see Goldstein, 1995).

DMA can be used to do meta-analyses of DMA studies to test the generality of their results. A DMA meta-analysis combines DMA studies to yield larger samples by adding a "studies" variable at the highest level (Goldstein, 1995). This meta-analysis maintains each DMA study's nested level structure (e.g., turns within time periods within groups) to yield more precise results (unlike current meta-analyses that use only the effect size).

Conclusion

This study of eighty high school students working on an algebra problem in groups of four showed that some group processes facilitated creation of correct ideas (correct contributions). As expected, groups with higher past mathematics grades or proportionately more correct contributions were more likely to solve the problem correctly. Also, watershed events separated distinct time periods with many vs. few correct contributions.

Recent actions by group members (micro-time contexts) affected the likelihood of a correct contribution at a given moment in time. Specifically, wrong contributions, correct evaluations, justifications, and polite disagreements increased the likelihood of a correct contribution. Students often detected and corrected flaws in wrong contributions to create correct contributions. Correct evaluations had broad effects, increasing the likelihoods of correct contributions, justifications, and subsequent correct evaluations. Justifications promoted rational discourse, increasing the likelihood of a subsequent justification and reducing the likelihood of a rude disagreement. While justifications had the largest effects, correct evaluations had the longest-lasting effects (across three speaker turns). In contrast, asking questions, disagreeing rudely, and agreeing, reduced the likelihood of a correct contribution.

Some effects differed across groups or across time periods. In groups that successfully solved the problem, justifications had larger effects on correct contributions and questions were more likely to elicit an explanation. Meanwhile, agreements' and correct evaluations' effects on correct contributions differed across time periods. Other variables that showed consistent effects across both groups and time periods are candidates for universal effects.

References

Amason, A. (1996). Distinguishing the effects of functional and dysfunctional conflict on strategic decision making. *Academy of Management Journal*, *39*, 123-148.

Balacheff, N. (1988). Aspects of proof in pupils' practice of school mathematics. In D. Pimm (Ed.), *Mathematics, teachers and children* (pp. 316-230). London: Hodder & Stoughton.

Bales, R. F. (2001). *Social Interaction Systems*. New Brunswick, NJ: Transaction Publishers.

Barron, B. (2003). When smart groups fail. *The Journal of the Learning Sciences, 12*(3), 307-359.

Benjamini, Y., Krieger, A. M., & Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika, 93*, 491-507.

Blanton, M., & Kaput, J. (2003). Developing elementary teachers' "algebra eyes and ears." *Teaching Children Mathematics, 10,* 70–77.

Brown, P, & Levinson, S. C. (1987). *Politeness*. New York: Cambridge University Press.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models*. Newbury Park: Sage.

Burgoon, J. K., Dillman, L., & Stern, L. (1993). Adaptation in dyadic interaction. *Communication Theory, 3*, 295-316.

California Department of Education (2005). *Academic performance index*. Retrieved on May 8, 2007, http://www.cde.ca.gov/ta/ac/ar/index.asp

Chiu, M. M. (1997). *Building on diversity*. Los Angeles: UCLA. (ERIC No. ED 410 325).

Chiu, M. M. (2000a). Group problem solving processes: Social interactions and individual turns. *Journal for the Theory of Social Behavior, 30*(1), 27-50.

Chiu, M. M. (2000b). Status effects on solutions, leadership, and evaluations during group problem solving. *Sociology of Education, 73*(3), 175-195.

Chiu, M. M. (2001). Analyzing group work processes: Towards a conceptual framework and systematic statistical analyses. In F. Columbus (Ed.), *Advances in psychology research* (vol. 4, pp. 193-222). Huntington, NY, US: Nova Science Publishers.

Chiu, M. M., & Khoo, L. (2005). A new method for analyzing sequential processes: Dynamic multi-level analysis. *Small Group Research, 36*, 600-631.

Chiu, M. M., & Khoo, L. (2003). Rudeness and status effects during group problem solving *Journal of Educational Psychology, 95*, 506-523.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127-149). Washington, D.C.: American Psychological Association.

Cobb, P. (1995). Mathematical learning and small-group interaction: Four case studies. In P. Cobb & H. Bauersfeld (Eds.), *The emergence of mathematical meaning* (pp. 25-130). Hillsdale, N.J.: Erlbaum.

Cohen, E. G. (1982). Expectation states and interracial interaction in school settings. *Annual Review of Sociology, 8,* 109-235.

Cohen, E. G. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research, 64*(1), 1-35.

Coleman, E. (1998). Using explanatory knowledge during collaborative problem solving in science. *The Journal of Learning Sciences, 7*, 387-427.

De Lisi, R., & Goldbeck, S. L. (1999). Implications of Piagetian theory for peer learning. In A. M. O'Donnell & A. King's (Eds.), *Cognitive perspectives on peer learning* (pp. 3-37). Mahwah, NJ: Erlbaum.

Dembo, M. H., & McAuliffe, T. J. (1987). Effects of perceived achievement and grade status on social interaction and influence in cooperative groups. *Journal of Educational Psychology, 79,* 415-423.

Doise, W., Mugny, G., & Perret-Clermont, A. N. (1975). Social interaction and the development of cognitive operations. *European Journal of Social Psychology, 5*, 367-383.

Ellis, D. G., & Fisher, A. B. (1994). *Small group decision making* (4th ed.). New York: McGraw-Hill.

Ericsson, K. A. (2001). Protocol analysis in psychology. In N. Smelser and P. Baltes (Eds.), *International Encyclopedia of the Social and Behavioral Sciences* (pp. 12256-12262). Oxford, UK: Elsevier.

Franke, M., Carpenter, T. & Battey, D. (2007). Content Matters. In Kaput, J. & Blanton, M. (Eds.) *Algebra in the Early Grades* (pp 333-359). Mahwah, New Jersey: Erlbaum.

Gersick, C. J. G. (1988). Time and transition in work teams: Toward a new model of group development. *Academy of Management Journal, 31,* 9-41.

Goldbeck, S. (1998). Peer collaboration and children's representation of the horizontal surface of liquid. *Journal of Applied Developmental Psychology, 19*, 571-592.

Goldstein, H. (1995). *Multilevel statistical models*. Sydney: Edward Arnold.

Grasa, A. A. (1989). *Econometric model selection*. Boston: Kluwer Academic.

Green, S. B. (1991). How many subjects does it take to do a regression analysis? *Multivariate Behavioral Research*, *26*, 499-510.

Hackman, M. Z., & Johnson, C. E. (2000). *Leadership*. Prospect Heights, IL: Waveland.

Hinsz, V. B. (1990). Cognitive and consensus processes in group recognition memory performance. *Journal of Personality and Social Psychology, 59*, 705-718.

Hinsz, V. B., Tindale, R. S., Vollrath, D. A. (1997). The emerging conceptualization of groups as information processors. *Psychological Bulletin, 121*, 43-64.

Hirokawa, R. Y. (1983). Group communication and problem-solving effectiveness: An investigation of group phases. *Human Communication Research, 9*, 291-305.

Holtgraves, T. (1997). YES, BUT … Positive politeness in conversation arguments. *Journal of Language and Social Psychology, 16*, 222-239.

Johnson, D. W., & Johnson, R. T. (1994). *Learning together and alone: Cooperative, competitive, and individualistic learning* (4th ed.). Boston: Allyn & Bacon.

Johnson, D. W., & Johnson, R. T. (2002). Cooperative learning methods: A meta-analysis. *Journal of Research in Education, 12*(1), 5-14.

Kennedy, P. (2004). *Guide to econometrics*. Cambridge: Blackwell.

Kieran, C. (1992). The learning and teaching of school algebra. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 390–419). NewYork: Macmillan.

Krippendorff, K. (2004). *Content analysis*. Thousand Oaks, CA: Sage.

Kuhn, D., Shaw, V., & Felton, M. (1997). Effects of dyadic interaction on argumentative reasoning. *Cognition and Instruction, 15*, 287-315.

Labov, W. (2001). *The social stratification of English in New York City*. Cambridge: Cambridge University Press.

Lilien, D. M., Startz, R., Ellsworth, S., Noh, J, & Engle, R. (1995). *Eviews*. Irvine, CA: Quantitative Micro Software.

Ljung, G., & Box, G. (1979). On a measure of lack of fit in time series models. *Biometrika, 66,* 265-270.

Maddala, G. S., & Kim, I. M. (1998). *Unit roots, cointegration and structural change*. New York: Cambridge University Press.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka's (Ed.), *Frontiers in Economics* (pp. 105-142). New York: Academic.

Mercer, N. (2008). The seeds of time. *The Journal of the Learning Sciences, 17*, 33-59.

Michaels, S., O'Connor, C., & Resnick, L.B. (in press). Deliberative discourse idealized and realized. *Studies in Philosophy and Education*.

Nijstad, B. A., Diehl, M., & Stroebe, W. (2003). Cognitive stimulation and interference in idea generating groups. In P. B. Paulus & B. A. Nijstad (Eds.), *Group creativity: Innovation through collaboration* (pp. 137-159). New York: Oxford University Press.

Orlitzky, M., & Hirokawa, R. Y. (2001). To err is human, to correct for it divine. *Small Group Research*, *32*, 313-341.

Paulus, P. B., & Brown, V. (2003). Ideational creativity in groups. In P. B. Paulus, & B. A. Nijstad (Eds.), *Group creativity* (pp. 110-136). New York: Oxford University Press.

Pavitt, C., & Johnson, K. K. (2001). The association between group procedural MOPs and group discussion procedure. *Small Group Research, 32*, 595-624.

Pavitt, C. (1993). What (little) we know about formal group discussion procedures: A review of relevant research. *Small Group Research, 24*(2), 217-235.

Person, N. K., Kreuz, R. J., Zwaan, R., & Graesser, A. C. (1995). Pragmatics and pedagogy *Cognition and Instruction, 13*, 161-188.

Piaget, J. (1985). *The equilibrium of cognitive structures* (T. Brown & K. L. Thampy, Trans.). Chicago: University of Chicago Press.

Poole, M. S. (1981). Decision development in small groups: A comparison of two models. *Communication Monographs, 48*, 1-24

Robinson, P. M. (1982). On the asymptotic properties of estimators with limited dependent variables. *Econometrica, 50*, 27-42.

Roschelle, J. (1992). Learning by collaborating: Convergent conceptual change. *The Journal of the Learning Sciences, 2*, 235-276.

Sfard, A. (2007). When the rules of discourse change, but nobody tells you. *The Journal of the Learning Sciences, 16*, 567–615.

Sharan, S., & Shachar, H. (1988). *Language and learning in the cooperative classroom*. New York: Springer.

Slavin, R. E. (1990). *Cooperative learning*. Boston: Allyn & Bacon.

Sowder, L., & Harel, G. (1998). Types of students' justifications. *Mathematics Teacher, 91*, 670–674.

Stasser, G., & Taylor, L. (1991). Speaking turns in face-to-face discussions. *Journal of Personality and Social Psychology, 60*, 675-684.

Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making. *Journal of Personality and Social Psychology, 53*, 81-93.

Toulmin, S. E. (2003). *The uses of argument*. Cambridge: Cambridge University Press.

Tracy, K., & Tracy, S. J. (1998). Rudeness at 911: Reconceptualizing face and face attack. *Human Communication Research, 25*, 225-251.

Tsui, A. (1992). A functional description of questions. In Coulthard, M. (Ed). *Advances in Spoken Discourse Analysis*. 89-110. London: Routledge.

Tudge, J. R. H. (1989). When collaboration leads to regression: Some negative consequences of socio-cognitive conflict. *European Journal of Social Psychology, 19*, 123-138.

Webb, N. M. (1984). Sex differences in interaction and achievement in cooperative small groups. *Journal of Educational Psychology, 76*, 33-44.

Webb, N. M. (1991). Task-related verbal interaction and mathematics learning in small groups. *Journal of Research in Mathematics Education, 22*, 366-389.

Webb, N. M., & Palincsar, A. S. (1996). Group processes in the classroom. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 841-873). New York: MacMillan.

Appendix A: Problem, coding, and solutions

Problem

You won a cruise from New York to London, but you arrive 5 hours late. So, the ship left without you. To catch the ship, you rent a helicopter. The ship travels at 22 miles an hour. The helicopter moves at 90 miles an hour. How long will it take you to catch the ship?

Goal

Find the time at which the ship and the helicopter are in the same location

Key problem situation understanding:

After 5 hours, both vehicles move simultaneously at their respective speeds

Solution Score

Correct answer:	3 points
Articulated at least one of the solution methods below:	2 points
Articulated the correct goal and problem situation:	1 point
None of the above:	0 points

Solution #1: Write the distance expression for each vehicle and equate them

Ship distance	= Helicopter distance
Ship speed × ship time	= Helicopter speed × helicopter time
22 mph × (T + 5) hours	= 90 mph × T hours
22 × T + 110 – 22 × T	= 90 × T – 22 × T
110	= 68 × T
110 / 68	= 68 × T / 68
1.6176	= T

Solution #2:

Compute current gap between ship and helicopter, distance ship traveled in 5 hours at 22 mph:

5 hours × 22 mph = 110 miles

Compute net closing speed, helicopter speed minus ship speed:

90 mph – 22 mph = 68 mph

Obtain time by dividing current gap by net closing speed.

110 miles / 68 mph = 1.6176 hours

Solution #3

Estimate the additional time needed by iteratively computing the extra time needed for the helicopter to travel to the ship's momentary new position.

(a) Compute ship movement after 5 hours
(b) Compute helicopter time needed to travel that distance
(c) Compute distance ship travels in that time
Repeat (b) and (c) until the helicopter and the plane are in the same place

Ship movement	Helicopter movement
22 mph × 5 hours =	110 miles
	110 miles / 90 mph = 1.22222 hours
22 mph × 1.222 hours =	26.8888 miles
	26.8888 miles / 90 mph = 0.29876 hours
22 mph × 0.29876 hours =	6.57273 miles
	6.57273 miles / 90 mph = 0.07303 hours

And so on …

Time = 1.22222 + .29876 + .07303 + … = 1.6176 hours

Appendix B: Coding Speaking Turns

In the case of overlapping speech/nonverbal behaviors, the interrupter's speech/behavior is coded as a separate turn after the interrupted person. Let's say Ron interrupts Ana. If Ana stops talking before Ron stops talking, Ana's turn ends at that point. However, if Ana continues speaking through and after Ron stops speaking, then Ana's speech consists of 2 turns. The first turn ends at the end of Ron's speech, and a second turn begins after Ron's speech.

If there are multiple interrupters, the interrupters' turns are sequenced according to who spoke first. If multiple people (say Ada and Ben) begin and end at the same time, we coded the speech as follows. Each simultaneous speaker is coded as responding to the previous single speaker.

a) If the simultaneous speakers say the same thing, the following speakers are coded as responding to all of these simultaneous speakers and the sum of their properties (such as peer status)

b) If the simultaneous speakers say different things, one of them usually continued arguing his/her position after a brief silent pause. Let's say Ada and Ben both speak, then Ada keeps talking. The turn order would be Ben, then Ada.

(In this data of high school students, a third person never spoke after simultaneous disagreeing speakers stopped. One of the disagreeing speakers always continued.)

<u>Decision trees for each speaker turn dimension</u>

Evaluation:

Does the speaker respond to the previous speaker?
- No, code as *unresponsive / ignore*
- Yes, does the speaker fully agree with the previous speaker?
  - Yes, code as *agree*
  - No, does the speaker disagree with the previous speaker?
    - No, code as *neutral*
    - Yes, does the speaker redress threats to face?
      - Yes, code as *politely disagree*
      - No, code as *rudely disagree*

Actions that redress threats to face during disagreements:

Hypothetical (if, let's say)
Indirect responsibility
Passive verbs (get, have),
Passive voice (is multiplied)
Citing other people
First-person plural pronouns (we, our)

(Disagreeing with a question rejects it as not useful; e.g. "you're asking the wrong question.")

Knowledge content, Validity, and Justification:

Does the speaker express any mathematics or problem-related information?

    No, code as *null content*
    Yes, is all of this information on the group's log/trace of problem solving?
            Yes, code as *repetition*
            No, code as *contribution*
                and write specific information in this group's log
        Does this information violate any mathematics or problem constraints?
            Yes, code as *incorrect*
            No, code as *correct*
        Does the speaker justify his or her idea?
            Yes, code as *justification*
            No, code as *no justification*

Invitational form

Do any of the clauses proscribe an action?

    Yes, code as *command*
    No, is the subject the addressee?
        No, are any of the clauses in the form of a question?
            No, code as *statement*
            Yes, code as *question*
        Yes, is the verb a modal?
            No, should the described action have been performed, but not done?
                Yes, code as a *command*
                No, code as a *question*
            Yes, Is it a Wh- question (who, what, where, why, when, how)?
                Yes, code as an *question*
                No, is the action feasible?
                    Yes, code as a *command*
                    No, code as an *question*

	Examples	**Invitational Form**
1.	Can you do it on the calculator, John? (John can use a calculator)	Command
2.	Can John do it on the calculator?	Question
3.	Can you do it on the calculator, John? (John might not know)	Question
4.	I hear someone joking around. (Stop joking)	Command
5.	Is someone joking around? (Stop joking)	Command
6.	What are you joking about? (Stop joking)	Command
7.	What are you joking about? (joking is ok in class)	Question
8.	The board is not erased	Command
9.	Did you erase the board?	Command
10.	Did you erase the board? (speaker can't see the board)	Question

The *invitational form* decision tree and the accompanying examples are based on Labov (2001) and Tsui (1992).

Appendix C .Mathematics Equations underlying the Speaker Turn Analyses

When do CCs occur?

To identify distinct time periods, I created a simple univariate time-series model (an auto-regressive order 1 model):

$$y_t = C + \beta\, y_{t-1} + \varepsilon_t \tag{2}$$

The parameter y_t indicates the value of the outcome variable y at speaker turn t. The parameter y_{t-1} indicates the value of the outcome variable in the previous turn, and β is its regression coefficient, indicating its relationship with the outcome variable in the current turn t. Meanwhile, C is a constant and ε_t is the residual at turn t. With breakpoints this model becomes:

$$y_t = C + C_2\, d_2 + C_3\, d_3 + \ldots + C_p\, d_p + \beta\, y_{t-1} + \varepsilon_t \tag{3}$$

The number of time periods is p, and d_p is the dummy variable associated with time period p. Likewise, C_p is the regression coefficient associated with time period p.

The BIC is defined as:

$$-\frac{2L}{n} + \left(\frac{k \ln(n)}{n}\right) \tag{4}$$

where k is the number of estimated parameters, n is the number of observations, and L is the value of the log likelihood function using the k estimated parameters.

Multilevel logit model

Conceptually, a multi-level Logit model can be divided into its multi-level part and its Logit part. Consider a 3-level model with an outcome variable, y_{ijk} (CC) at speaker turn i of time period j in group k and a Logit link function (F):

$$\mathrm{y}_{ijk} = \beta_{000} + \mathrm{e}_{ijk} + \mathrm{f}_{0jk} + \mathrm{g}_{00k} \tag{5}$$

$$\pi_{\mathrm{ijk}} = \mathrm{p}\,(\mathrm{y}_{\mathrm{ijk}} = 1) = \mathrm{F}\,(\beta_{000} + \mathrm{f}_{0jk} + \mathrm{g}_{00k}) = \frac{1}{1+e^{-(\beta_{000}+f_{0jk}+g_{00k})}} \tag{6}$$

The level-2 variation parameter f_{0jk} represents the deviation of time period j from the overall mean while g_{00k} represents the deviation of group k from the overall mean β_{000}. The probability (π_{ijk}) that an event (e.g. a CC) occurs at turn i of time period j in group k is determined by the expected value of the outcome variable and the Logit link function (F). The level-1 variation, e_{ijk}, does not contribute to the fixed components and is a random variable only at level-1. So, I constrain the variance of e_{ijk} to 1 without loss of generality.

Therefore, the observed outcome variable y_{ijk} is:

$$\mathrm{y}_{ijk} = \pi_{ijk} + \mathrm{e}_{ijk}\mathrm{z}_{ijk} \tag{7}$$

$$\sigma_{\mathrm{e}}^2 = 1 \tag{8}$$

$$\mathrm{z}_{ijk} = [\ \pi_{\mathrm{ijk}}(\ 1 - \pi_{\mathrm{ijk}})\ ]^{0.5} \tag{9}$$

Then, I added a vector of s classroom variables as control variables: classroom identification numbers (**S**).

$$\pi_{ijk} = p\,(y_{ijk} = 1 \mid \mathbf{S}_{00k}, \beta_{00s}) = F\,(\beta_0 + \beta_{00s}\mathbf{S}_{00k} + f_{0jk} + g_{00k}) \tag{10}$$

I tested whether this set of predictors was significant with a nested hypothesis test (χ^2 log likelihood, Kennedy, 2004).

Next, I entered a vector of *t* variables at the group level: correct group solution, mean of group members' mathematics grades, mean of group members' peer statuses, variance of group members' mathematics grades, and variance of group members' peer statuses (**T**).

$$\pi_{ijk} = p\,(y_{ijk} = 1 \mid \mathbf{S}_{00k}, \beta_{00s}, \mathbf{T}_{00k}, \beta_{00t}) = F\,(\beta_0 + \beta_{00s}\mathbf{S}_{00k} + \beta_{00t}\mathbf{T}_{00k} + f_{0jk} + g_{00k}) \tag{11}$$

I tested whether this set of predictors was significant with a nested hypothesis test (χ^2 log likelihood, Kennedy, 2004). Then, I tested for interaction effects among pairs of significant variables in **U**. Non-significant variables and interactions were removed from the specification.

Next, I added *u* current speaker variables at the speaker turn level: gender, race, mathematics grade, peer status, correct evaluation, agree, politely disagree, rudely disagree, justify, question and command (**U**).

$$\pi_{ijk} = F(\beta_0 + \beta_{00s}\mathbf{S}_{00k} + \beta_{00t}\mathbf{T}_{00k} + \beta_{ujk}\mathbf{U}_{ijk} + f_{0jk} + g_{00k}) \tag{12}$$

Likewise, I applied the procedure for **T** to **U**. Next, I tested if the *u* speaker turn level regression coefficients ($\beta_{ujk} = \beta_{u00} + f_{ujk} + g_{u0k}$) differed significantly (Goldstein, 1995) at the time-period level ($f_{ujk} \neq 0$?) or group level ($g_{u0k} \neq 0$?). If yes, I kept these parameters in the model. Otherwise, I removed them.

Using a vector autoregression (VAR, Hamilton, 1994), I entered lag variables for the previous speakers, first lag 1 (indicating the previous turn and denoted -1), then at lag 2 (denoted -2), then at lag 3, and so on until none of the variables in the last lag were significant (lag 4 in this case). First, I added *v* previous speaker variables at the speaker turn level: gender (–1), race (–1), mathematics grade (–1), peer status (–1), correct evaluation (–1), agree (–1), politely disagree (–1), rudely disagree (–1), CC (–1), wrong contribution (–1), correct old idea (–1), justify (–1), question (–1), and command (–1) (**V**).

$$\pi_{ijk} = F(\beta_0 + \beta_{00s}\mathbf{S}_{00k} + \beta_{00t}\mathbf{T}_{00k} + \beta_{ujk}\mathbf{U}_{ijk} + \beta_{vjk}\mathbf{V}_{(i-1)jk} + f_{0jk} + g_{00k}) \tag{13}$$

Likewise, I applied the procedure for **U** to **V**. Then, I repeated the procedure for lags -2, -3, and -4 of the variables in **V**. Like β_{vjk}, the following symbols ϕ_{vjk}, γ_{vjk}, and η_{vjk} denote the regression coefficient matrices for the variables in **V** but at lags –2, –3, and –4 respectively.

$$\pi_{ijk} = F(\beta_0 + \beta_{00s}\mathbf{S}_{00k} + \beta_{00t}\mathbf{T}_{00k} + \beta_{ujk}\mathbf{U}_{ijk} + \beta_{vjk}\mathbf{V}_{(i-1)jk} + \phi_{vjk}\mathbf{V}_{(i-2)jk} + \gamma_{vjk}\mathbf{V}_{(i-3)jk} + \eta_{vjk}\mathbf{V}_{(i-4)jk} + f_{0jk} + g_{00k}) \tag{14}$$

Appendix D: Correlation Tables and Additional Analyses

Table D1
Summary of speaker turns coded in successful and unsuccessful groups

	Groups	
Speaker turns	Successful	Unsuccessful
Original	1,372	1,781
Poor sound quality	21	28
Omitted due to lags	74	79
Total used	1,277	1,674

Table D2
Inter-rater reliability of each coding dimension

Coding dimension	Agreement %	Krippendorf's alpha
Evaluation of previous action	96	0.93
Knowledge content	98	0.98
Correct idea	99	0.99
Invitational form	96	0.91

Table D3
The following matrix show the correlations, variances, and covariances of the outcome variables and the significant predictors at the group level. The lower left triangle of each matrix contains the correlations. The bold numbers along the diagonal are the variances, and the upper right triangle contains the covariances.

Group level variable	1	2	3
1. Solution score	**1.490**	5.314	0.142
2. Group mean mathematics grade	0.658	**43.715**	0.519
3. % Correct contributions	0.671	0.453	**0.030**

Table D4
The following matrix shows the correlations, variances, and covariances of the outcome variables and the significant predictors at the speaker turn level. The lower left triangle of each matrix contains the correlations. The bold numbers along the diagonal are the variances, and the upper right triangle contains the covariances.

	Speaker turn variable	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1.	Correct contribution	**0.158**	-0.009	-0.002	0.058	0.008	0.035	0.037	0.030	0.010	-0.004
2.	Rude disagreement	-0.079	**0.087**	-0.054	-0.009	-0.009	-0.004	-0.006	-0.003	0.004	-0.005
3.	Agreement	-0.020	-0.368	**0.246**	-0.009	0.002	0.029	0.012	0.019	-0.012	-0.002
4.	Justification	0.412	-0.083	-0.050	**0.127**	0.025	0.010	0.014	0.003	0.005	-0.004
5.	Justification (lag 1)	0.058	-0.086	0.006	0.193	**0.127**	0.006	0.009	0.014	0.014	0.001
6.	Correct evaluation (lag 1)	0.195	-0.030	0.127	0.059	0.038	**0.210**	0.047	0.048	-0.001	0.008
7.	Correct evaluation (lag 2)	0.211	-0.046	0.055	0.093	0.060	0.230	**0.208**	0.047	0.004	-0.016
8.	Correct evaluation (lag 3)	0.164	-0.021	0.080	0.022	0.091	0.231	0.225	**0.209**	-0.003	0.001
9.	Wrong contribution (lag 1)	0.093	0.048	-0.090	0.047	0.142	-0.009	0.016	-0.028	**0.088**	0.000
10.	Question (lag 2)	-0.022	-0.043	-0.020	-0.023	-0.001	0.038	-0.079	0.003	0.008	**0.184**

Table D5
Significant, unstandardized parameter coefficients of hierarchical set multilevel Logit regressions predicting correct contributions at the speaker turn level (with standard errors in parentheses)

	6 Multi-level Logit regressions predicting correct contributions					
Predictor	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Unsolved	-0.907 *	-0.581	-0.526	-0.69	-0.647	-0.43
	(0.463)	(0.558)	(0.555)	(0.552)	(0.546)	(0.477)
Rude		-0.970 ***	-1.177 ***	-1.198 ***	-1.238 ***	-1.173 ***
Disagreement		(0.256)	(0.26)	(0.265)	(0.265)	(0.26)
Agreement		-0.633 ***	-0.642 ***	-0.647 ***	-0.66 ***	-0.614 ***
		(0.131)	(0.131)	(0.134)	(0.133)	(0.123)
Justification		3.725 ***	3.827 ***	3.803 ***	3.849 ***	3.82 ***
		(0.276)	(0.276)	(0.277)	(0.276)	(0.284)
Justification		-1.568 ***	-1.651 ***	-1.654 ***	-1.69 ***	-1.685 ***
X unsolved		(0.338)	(0.338)	(0.341)	(0.339)	(0.346)
Lag 1 predictors						
Correct			0.447 ***	0.427 **	0.402 **	0.387 *
Evaluation			(0.125)	(0.128)	(0.128)	(0.127)
Wrong			0.928 ***	0.952 ***	0.977 ***	0.984 ***
contribution			(0.178)	(0.182)	(0.181)	(0.18)
Justification			-0.550 **	-0.530 **	-0.555 **	-0.599 **
			(0.163)	(0.167)	(0.166)	(0.164)
Lag 2 predictors						
Correct				0.414 **	0.387 **	0.401 **
evaluation				(0.125)	(0.125)	(0.108)
Question				-0.615 **	-0.613 **	-0.541 *
				(0.198)	(0.197)	(0.194)
Question x				1.008 ***	1.002 ***	0.967 **
unsolved				(0.277)	(0.276)	(0.272)
Lag 3 predictors						
Correct					0.353 *	0.356 *
evaluation					(0.127)	(0.126)
Random variation of predictors at the time period level						
Agreement						0.445 *
						(0.149)
Correct						0.555 ***
evaluation						(0.135)

Note. *Significant fixed constant term, random time period constant term, and random speaker turn constant terms are omitted*.
*p < .05, **p < .01, ***p< .001