Group Problem Solving and the Creation of Correct New Ideas: Effects of Evaluations, Wrong Ideas, Justifications, and Rudeness

Ming Ming Chiu
Department of Educational Psychology
The Chinese University of Hong Kong

Creating new ideas that are correct (NICs) helps groups solve difficult problems, so modeling group processes that affect the likelihood of NICs might help students create NICs. 80 high school students worked in groups of 4 to solve an algebra problem and were videotaped. Group mean mathematics grade and percentage of NICs predicted solution score at the group level. Using a new statistical method for analyzing group processes in large data sets, I examined predictors of a NIC at the speaker turn level. Correct evaluations raised the likelihood of NICs over the next three turns. Furthermore, wrong, new ideas facilitated NICs by the next speaker. Justifications also predicted NICs by both the current speaker and the next speaker. Meanwhile, speakers who agreed or

rudely disagreed were less likely to create a NIC in the same turn. Unlike unsuccessful groups, successful groups' justifications were more likely to accompany NICs, and their questions more often received satisfactory responses. Breakpoint analyses showed that NICs occurred in clusters for each group, averaging 3 time periods per group. Correct evaluation (lag 2) and agreement showed different size effects across time periods. Otherwise, all other predictors showed similar size effects across groups and across time periods.

Students who work together (*cooperative learning*) typically show many positive outcomes, including increased learning, increased motivation, decreased racial tension, etc. (Slavin, 1990). However, such positive results are not universal (e.g., Laughlin, VanderStoep, & Hollingshead, 1991). Why are some groups more successful than others?

Earlier researchers used group structures and group member traits to explain different group *outcomes*. More recently, researchers in psychology, education, organizational behavior and sociology have focused on how group processes affect group outcomes. In this paper, I examine relationships within group processes. Specifically, I focus on group problem solving (GPS) and new ideas that are correct (NIC). GPS is a process by which group members work together to find a consensus solution to specific problem. NICs are correct, new ideas produced during group problem solving that no group member had previously expressed. Theoretical models of GPS highlight the importance of NICs as the building blocks of successful group outcomes (e.g., Chiu, 2000a, 2001). This in turn raises the issues of which group processes affect the *process* of creating of NICs and whether these processes differ across groups or across time periods within each group.

In this study, I contribute to the research literature by systematically examining the group-, individual- and speaker turn-level factors that helped

and hindered the process of creating NICs during GPS in both success-
ful and unsuccessful groups. In particular, I test whether correct evaluations
and wrong ideas by earlier speakers helped the current speaker create a
NIC. I also estimated the effect of NICs on the GPS outcomes. In contrast
to past studies that used only data from questionnaires and group outcomes,
I also analyzed videotape data of students' actual problem-solving behaviors,
using advanced statistical methods. Specifically, I used a new methodol-
ogy to model the group processes at the turn level: dynamic multilevel Probit/
Logit.

Theoretical Perspective

During GPS, students work together to find an answer to a specific
problem. GPS is unlike other types of cooperative learning (e.g., poetry
discussions) that need not yield a consensus interpretation. Ideal GPS
tasks are too difficult and/or complex for one member to do alone, but are
simple enough for the group to solve together. This section describes how
GPS processes can help or hinder creation of NICs.

GPS Processes That Help NICs

Compared to individuals, group members' diverse views and argumenta-
tion tend to create more NICs. Diverse views helps the group create more
ideas and judge them more accurately. Through argumentation, group
members tend to create correct ideas by evaluating ideas, identifying
problems, and justifying NICs.

Diverse views. Group members often have different perspectives and
diverse sources of knowledge (Hastie, 1986). Capitalizing on this diversity,
groups (especially heterogeneous ones) often create more ideas, repre-
sentations and solution proposals (Stasson & Bradshaw, 1995). Thus, at
least one of their ideas is more likely to be correct or optimal.

Group members can express idiosyncratic ideas and build on each

other's ideas to construct new alternatives, through processes such as sparked ideas, jigsaw pieces, and creative misinterpretations (Chiu, 1997). Comments by one person, say a key word, may spark another person to propose a NIC. In addition, two or more members can put together different pieces to create a NIC, like fitting jigsaw pieces together. Finally, a person may creatively misinterpret another person's idea to construct a correct, new one. So, even incorrect, new ideas can lead to NICs.

Group members' diverse views also increase the likelihood that at least one of them will recognize a flaw (Cobb, 1995; Piaget, 1985). So, they are more likely to identify and refine incorrect ideas. By creating more ideas and judging their validity better, groups are more likely to create NICs.

Argumentation. Successful GPS of problems often involves argumentation, a social process by which people explain and justify their own views to convince both themselves and others (Cobb, 1995). During this process, group members evaluate one another's ideas, recognize problems, and justify their views (Cobb).

Evaluations. Evaluations characterize how the current speaker responds to the previous speaker, especially how the current speaker assesses the previous action and the problem solving approach (Chiu, 2000a, 2001; Goodwin & Goodwin, 1987; Pomerantz, 1984). Evaluations can accept, reject, or ignore the previous action, and they can be right or wrong. While agreements encourage continuation of the current problem-solving trajectory, disagreements and changes of topic try to change it, often with new ideas. Meanwhile, correct evaluations support NICs or identify flawed ideas. In contrast, incorrect evaluations reject NICs or accept flawed ideas (Harwood, 1995). A group's collective attention and diverse perspectives can help it evaluate ideas correctly (Cobb, 1995; Hinsz, 1990).

Perturbations. Group members recognize problems or difficulties (perturbations) and express them through disagreements or questions. Piaget (1985) defined two types of perturbations: (a) obstacles, which give negative feedback (e.g., " that doesn't work because the ship keeps moving")

and (b) lacunae, gaps in a person's understanding (e.g., "how do we find the speed?"). Disagreements indicate obstacles to be overcome and hence motivate the creation of relevant NICs. A question may indicate one's own lacuna or reflect a group lacuna. The first case is a request for help with the likely benefits of assistance to the inquirer and elaborated understanding to the helper (Bargh & Schul, 1980; Webb & Farivar, 1999). So, it likely encourages review of previous ideas rather than new ones. In contrast, questions that reveal group lacunae point to new directions for creating NICs. De Lisi and Goldbeck (1999) argued that group members' diverse views and levels of knowledge facilitate both perturbations and responses to them. By responding to these perturbations, groups tend to create more new ideas, including some NICs (Asch, 1952; Doise, Mugny, & Perret-Clermont, 1975; Piaget, 1985).

Justifications. After perturbations provoke new ideas, justifications often follow them. When group members disagree, the person proposing the idea typically tries to justify it (Pontecorvo, 1993). In response, other members might present different views and justifications (Piaget's, 1985, genuine argument). Likewise, when a group member indicates a gap in understanding by asking a question, other members can respond with explanations and justifications. As justified ideas are more likely than unjustified ideas to be correct, justifications tend to raise the likelihood of NICs (e.g., Goldbeck, in press; Lindow, Wilkinson, & Peterson, 1985).

GPS Processes That Hinder NICs

Concerns about public self-image (face) and other impediments can reduce NICs directly by reducing the expression of new ideas or indirectly by reducing group members' attention to NICs (Harwood, 1995). Concerns about face encourage group members to accept incorrect ideas, discourage NICs that conflict with other group members' ideas, and discourage acceptance of NICs during disagreements.

Rudeness. Group members' concerns about face can discourage NICs

and distort evaluations of NICs. When people interact, they try to maintain their face and use an appropriate level of politeness to do so (Brown & Levinson, 1987; Goffman, 1959). In face-threatening situations, rude behavior is particularly harmful, so a speaker's politeness is likely to increase (Brown & Levinson, 1987). An examination of evaluations shows the dangers of both excessive politeness and inadequate politeness.

Evaluations and rudeness. Each type of evaluation has an implicit level of politeness and affects the previous speaker's maintenance of face (Chiu, 2000a, 2000b, 2001). Evaluations range from polite to rude: agreement, neutral, change of topic, polite disagreements, and rude disagreements. By agreeing with the previous speaker, the responder promotes the previous speaker's face. In contrast, neutral actions do not promote face. Although changes of topic are often neutral in a GPS context, they can be rude if the previous speaker asks a question or otherwise expects a response. Lastly, disagreements are face-threatening because they reduce public perception of a person's competence (Brown & Levinson, 1987).

To reduce the threat to a person's face, people can disagree politely (Brown & Levinson, 1987; Person, Kreuz, Zwaan, & Graesser, 1995). Compare a rude disagreement "You're wrong" with a polite one, "If 6 is multiplied by 2, we don't get 10." The polite disagreement both reduces blame and creates common ground. First, the speaker uses the hypothetical "if," thereby distancing the idea from reality. Second, the speaker avoids assignment of blame by not referring to the previous speaker. Third, the speaker hides agency by using the passive voice. Lastly, the speaker uses the passive circumstantial verb "get," thereby implicating agency in external conditions. The speaker uses the polite criticism to create common ground by repetition and shared positioning. Using repetition, "6 is multiplied by 2 ... 10," the speaker shows shared understanding. Also, the speaker uses shared positioning, "we," to claim a common cause. Holtgraves (1997) showed that politeness during disagreement is the accepted norm and that lack of the above redresses is noticeably rude.

Concerns about face during evaluations. Concerns over face can distort evaluations of NICS and reduce expression of NICs through preference for agreement and through retaliatory disagreements. Agreements promote friendly relationships through the social exchange of positive mutual respect (Brown & Levinson, 1987). This is especially true if the participants invest themselves in their ideas. So, 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, disagreements (even polite ones) threaten face. So, group members who are concerned about their social relationships tend to agree more with one another and disagree less, thereby allowing potential NICs to remain unvoiced and errors to persist. Even authority does not eliminate this effect, as tutors often do not criticize their students' errors (Person et al., 1995). Teenage girls are especially reluctant to disagree with one another (Tann, 1979; Tudge, 1989).

Suppose a person disagrees. Ideally, the target person(s) would try to understand the criticism. However, the threat to the person's face may encourage him or her to retaliate, possibly with a rude, emotionally-loaded response (Chiu & Khoo, 2003; Gottman & Krokoff, 1989). If a spiral of rude disagreements ensues, the collaboration may end. A politer disagreement can reduce the likelihood of retaliation by supporting the relationship with the target person(s). Then, he or she is more likely to try to understand the speaker's criticism(s), recognize the flaw and correct it with a NIC (Chiu & Khoo, 2003). Whereas rude disagreements are face-threatening and can obstruct the GPS process, polite disagreements reduce interpersonal conflict and help create NICs.

Method

I test the above hypotheses at both the group level and at the student action level. See Figure 1 for a summary of these hypotheses.

Figure 1   Model of the Effects of Student and Group Properties Before the Group Problem Solving Process and of Group Problem Solving Processes on Correct, New Ideas and on Group Solution Score
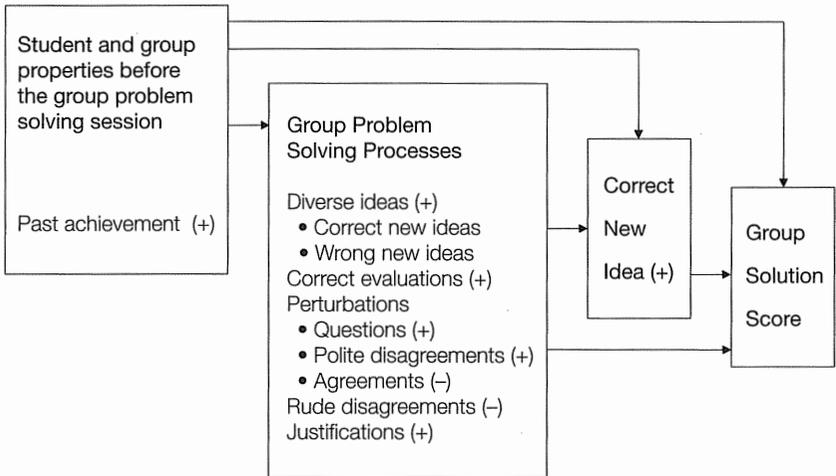
Participants

The participants attended four ninth grade algebra classes in an urban U.S. high school. On several state-wide exams, the school consistently scored between the 40th to 50th percentiles in mathematics (personal communication from teachers). 87 racially diverse students were asked to answer pretest and posttest questionnaires. These students had not previously worked together and had not received any group work training. 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.)

Guided by the teacher and the questionnaires, these students were placed into groups of four so that no members within a group were close friends. There was no same gender or same race group. There were 40 girls and 40 boys and their races were 12 Asian, 27 Black, 28 Hispanic, and 13 White. The group members in this study have known one another for at least 7 months, were aware of one another's grades, and had direct knowledge of one another's mathematics abilities. Hence,

group members' relative mathematics ability was likely the primary status effect and diffuse status characteristics likely have much smaller effects on their interaction.

Procedure

All 80 algebra students who agreed to be part of the study filled out questionnaires. These questionnaires had two questions regarding peer friendship. "Who are the 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've won a cruise from New York to London, but you arrive 5 hours late and the ship has left without you. To catch the ship, you rent a helicopter. If the ship travels at 22 miles an hour and the helicopter moves at 90 miles an hour, how long will it take for 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 method. 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 any procedures for solving this problem in class. Furthermore, the problem involved complex mathematics relationships, non-trivial combinations of multiple operations, and a non-integer solution. The most efficient solution method is equating the distance equations for each vehicle, cruise ship and helicopter (22mph x [Time+ 5 hours] = 90 mph xime), to obtain 1.618 hours or 1 hour 37 minutes.

Without teacher assistance, the students worked in groups for 30 minutes, facing moderate rather than tight time pressure. (Moderate time pressure allowed them to focus on quality rather than simply task completion, Hinsz, Tindale & Vollrath, 1997) They had pens, paper,

and calculators available for their use. All students were videotaped, and the subsequent data were transcribed, coded and analyzed.

Variables

Gender and race. The following variables counted the total members within each group that belonged to the given category: girls, Asians, Blacks, Whites and Latinos.

Mathematics grade. The students' mathematics grades were their mid-year algebra grades. The group mathematics grade was the mean of the group members' grades. The group mathematics grade variance measured the grade differences within each group. The group's highest mathematics grade was the highest mathematics grade of the members in the group.

Peer friendship. Each student's peer friendship measure was the mean number of times a student's name appeared in his or her classmates' answers to the two social status questions. The group means and group variances of peer friendship were also computed.

Solution score. Each group also received a final *solution score* (0–3) based on their problem solving progress toward a correct answer. For the turn-level analysis, the solution score was a binary variable, 0 or 1, as the multi-level analyses failed to converge when using complex interaction terms with four possible solution scores.

Time period. Each group's problem solving session was divided into time periods according to the break point method described below. The number of time periods in a group ranged from two to five.

Coding

I trained two research assistants (RAs) to code using a similar set of data from a pilot study. Blind to the study's hypotheses, the RAs coded all the transcripts. I used Scott's (1955) pi (π) to test for inter-coder reliability.

Speaker turn variables. A speaker turn is a sequence of words by a group member bracketed by the words of other group member(s). Turns

unaccompanied by words (e.g., writing "3x40") were also counted as turns. The RAs coded the following variables: correctness, new idea, NIC, justification, agreement, rude disagreements, polite disagreements, questions, and commands. The baseline for the evaluations was change of topic (or ignores previous speaker). Likewise, the baseline for questions and commands was statements (Chiu, 2001, called this dimension "invitational form"), see Chiu (2000a, 2001) for coding details.

On-task turns were judged for correctness. A speaker turn was coded as correct if the expressed knowledge content was consistent both with the problem situation and with mathematical concepts and relationships. On-task turns that were both correct and gave new information relative to the problem solving sessions were coded NICs. Additional computed variables included the ratios of these variables to total speaker turns per group.

Relative variables. Relative variables indexed how the individual compared to the other group members with regard to a particular property. Consider a person with a mathematics grade of 90 and two groups. In one group, the other three members' grades are all 70. In the other group, their grades are all 95. As the person's relative grade differs in these two groups, he or she may behave differently. The relative variables were computed as follows: relative variable = individual value – group mean value. Relative variables were computed for the following measures: mathematics grade, peer friendship, and rudeness of disagreements.

Predicting Solution Score

Hierarchical regressions and path analyses were used to test for total, direct and indirect effects on solution score. I entered the predictors sequentially into the regression in order of temporal occurrence and theoretical importance. The order was: group mathematical grade (mean and/or highest), peer friendship, mathematical grade variance, peer friendship variance, words, on-task words, and % NICs over total group turns.

A path analysis tested for direct and indirect effects. Temporal order constrained causal relationships. So, the final model's predictors were entered in temporal order into the path analysis. These computations were performed with the statistical software, E-views.

Predicting NICs at the Speaker Turn Level

Statistical analysis of group processes at the speaker turn level can be problematic in at least three ways. First, group members' behaviors and effects differ across groups and across time. Second, the outcome variables are often discrete (e.g., correct vs. incorrect) rather than continuous. Third, events are often similar to recent events, so variables tend to depend on values from recent turns (serial correlation). I address all of these problems with Dynamic Multilevel Probit/Logit (Chiu, 2001; Chiu & Khoo, 2003). This procedure includes identifying distinct time periods, testing for group and time period differences, building an explanatory model for NICs, testing for serial correlation, and modeling direct and indirect effects.

Building an explanatory model for NICs. I used multi-level Probit regressions to estimate a general time-series model for the binary outcome variable NIC (Goldstein & Rasbash, 1996). First, I tested whether NICs depended on whether or not a group solved the problem. I then entered the following sets of variables in sequence: gender, race, mathematics grade, and peer status. For mathematics grade, and peer status, I entered the following variables in sequence: (a) group average, (b) group variance, (c) relative property of the current speaker, and (d) relative property of the previous speaker.

Then I entered current speaker turn-level variables in the following sets: (a) correct evaluation, agreement, polite disagreement, and rude disagreement, (b) justification, and (c) question and command. Next, I entered lag variables for the previous speakers (lag1), then at lag 2, then at lag 3 and lastly, at lag 4. These sets of variables were:(a) correct evaluation,

agreement, polite disagreement, and rude disagreement, (b) NIC, c) wrong contribution, (d) correct old idea, (e) justification, and (f) question and command. In this paper, I use the convention that speaker (lag 1) refers to the previous speaker; speaker (lag 2) refers to the speaker from two turns ago; and so on.

Only significant variables were retained for subsequent sequential regressions to avoid inflated risk of Type I error. Doing many tests on one set of data increased the likelihood of a spurious correlation. To address this issue, I used Hochberg's (1988) variation on Holm's (1979) method.

If groups or time periods showed significant variation, explanatory variable effects on NICs can vary across groups or across time periods. So, I estimated the variations of significant explanatory variables' effects across groups and across time periods using random parameters for each explanatory variable at the group or time period level as needed in MLn (Goldstein, 1995).

Total effects of each predictor. 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. Temporal order constrained causal relationships, so the final model's explanatory variables were entered in temporal order into the path analysis.

To facilitate interpretation of these results, I convert the total effects of each predictor to odds ratios (Judge, Griffiths, Hill, Lutkepohl, & Lee, 1985). Furthermore, I present representative examples to illustrate non-intuitive predictors effects. I also estimated the predictive accuracy of the final model by testing how accurately the final model predicted a NIC at each speaker turn in each group. An alpha level of .05 was used for all significance tests.

Results

First, I showed that the percentage of NICs predicted solution scores at the group level. Then, I examined the predictors of NICs at the speaker

turn level. Due to space considerations, I include only the main results here. All results are available upon request.

Predicting Solution Scores

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.

Mean group mathematics grade and percentage of NICs positively predicted solution score (b= 0.256, SE = 0.089; and b = 11.377, SE = 5.033, respectively, R^2 = .374). None of the other predictors significantly affected solution score. Group mean mathematics grade also predicted percentage of NICs (b= .012, SE = .005, $p < .05$, R^2 = .20).

Predicting NICs

The 2951 turns in the data included 19% NICs, 10% wrong contributions, 30% correct evaluations, 15% justifications, and 50% on-task turns. Evaluations included 54% agreements, 0.3% neutral, 16% change of topic turns, 18% polite disagreements and 9% rude disagreements. Of these, 30% were correct on-task evaluations and 20% were incorrect on-task evaluations. Questions were 24% of the turns, and commands were 7% of the turns. The inter-rater reliabilities as measured by Scott's π for evaluations, new ideas, correctness of ideas, correctness of evaluations, and invitational forms were 0.95, 0.89, 0.99, 0.95, and 0.91 respectively.

Identifying distinct time periods. The number of breakpoints for each group ranged from zero to five, yielding one to six time periods. The number of breakpoints for successful groups (M = 3.2, SD = 2.1) did not differ significantly from that of unsuccessful groups (M = 2.9, SD = 1.7).

Group or time period differences? The likelihood of NICs varied significantly across time-periods (M = 3.457, SE = 0.682) and turns (M = 0.908, SE = 0.024), but not across groups (M= 0.000, SE = 0.001). Whereas the group level results showed that the percentage of NICs in successful groups exceeded that of unsuccessful groups, the variance components model

showed that NICs varied much more within a group than across groups. Of the total variance, 79% was across time periods while 21% was within time periods. Thus, the NICs clustered in some time periods. As groups did not show significant NIC variance, a 2-level model (time period and turns) with group interaction terms was sufficient.

Explanatory model, serial correlation, total effects. Properties of both the current speaker and previous speakers affected the likelihood of a NIC, with two predictors showing different effects across time periods (agree and correct evaluation [lag 2]). Aside from these two predictors, the effects of all other predictors did not differ significantly across time periods. The Q-statistics run on the final model showed no significant serial correlation of residuals in any of the ten groups. So, the time-series model used was likely to be appropriate. Altogether, the accuracy of this model for predicting NICs is 84%. For path analysis results, see Figure 2.

Students who justified their ideas were much more likely to introduce NICs, 4,460% more likely in successful groups and 745% more likely in unsuccessful groups (odds ratios computed from the total effects of each predictor). Meanwhile, students who disagreed rudely were 69% less likely to make a NIC. Students who agreed were 45% less likely to make a NIC on average, with the effect varying across time periods from 2% less likely to 67% less likely.

The local temporal context also affected the likelihood of a NIC as shown by the effects of previous speakers' actions within the last three turns. If the previous speaker (lag 1) justified an idea, a student was over 1,137% more likely to express a NIC. Also, if the previous speaker gave a wrong new idea or evaluated correctly, the current speaker was more likely to give a NIC, by 72% or 15% respectively. If the speaker from two turns ago (lag 2) made a correct evaluation, the current speaker was 467% more likely to make a NIC on average, with the effect varying across time periods from 260% more likely to 1007% more likely. (The current speaker and the speaker two turns ago were the same person 58% of the time.)

Figure 2   Path Analysis of Significant Predictors of NICS Using Multilevel Logit. Values are Standardized Parameter Coefficient.

Creating Correct, New Ideas
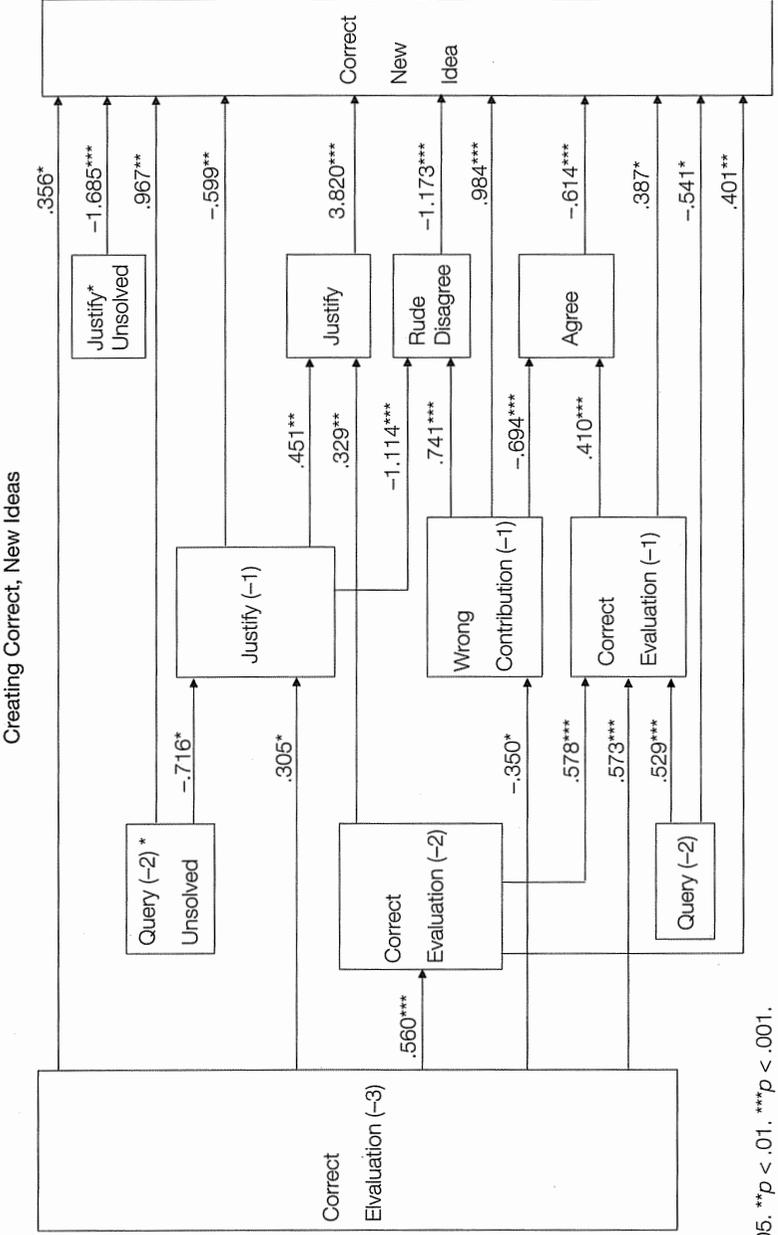
*p < .05. **p < .01. ***p < .001.

Meanwhile, if a speaker (lag 2) asked a question, the current speaker was less likely to make a NIC, 40% less likely in successful groups and 72% less likely in unsuccessful groups. Lastly, if the speaker (lag 3) evaluated correctly, the current speaker was 608% more likely to make a NIC.

Groups that successfully solved the problem were more likely to create NICs than unsuccessful groups, but this effect was not significant after adding other predictors. None of the other predictors were significant. In particular, individual mathematics grade and correct ideas were not significant. Recall that group mean mathematics grade positively correlated with group percentage of NICs. Together, these results suggest that the group mean effect on NICs stemmed from interactions among members, not simply the sum of more NICs from individuals with high mathematics grades. Moreover, with recent correct evaluations in the regression, neither recent NICs nor recent correct ideas were significant. Yet, correct evaluations remained significant, suggesting that evaluations had a greater effect on NICs than recent NICs or recent correct ideas.

Several of these predictors affected one another are seen in Figure 2. While justifications had the strongest effects on NICs, correct evaluations had the broadest and longest lasting effects. If the speaker (lag 3) evaluated correctly, students were 38% more likely to justify their idea, 84% more likely to agree, and 45% less likely to disagree rudely. If the speaker (lag 2) evaluated correctly, students were 39% more likely to justify their idea and 27% more likely to agree. Students were 51% more likely to agree if the previous speaker evaluated correctly. Furthermore, correct evaluations raised the likelihood of subsequent correct evaluations, in the next turn by 78% and in two turns by 145%.

Justifications, questions, and wrong new ideas also affected other predictors. If the previous speaker justified an idea, a student was 57% more likely to justify an idea and 67% less likely to disagree rudely. If speaker (lag 2) asked a question, students in unsuccessful groups were 122% more likely to disagree rudely and 28% less likely to justify an idea. In successful groups, if speaker (lag 2) asked a question, speaker (lag 1) was 30% more likely to evaluate

correctly, and the current speaker was 11% more likely to agree. If the previous speaker gave a wrong, new idea, as student was 50% less likely to agree and 110% more likely to disagree rudely.

Examples of predictor effects. The following transcript segments help illustrate the functional mechanisms of the following predictor effects: rude disagreements, agreements, wrong new ideas, questions and justifications. After a disagreement, especially a rude one, a student often disagreed rudely without introducing new ideas. Consider the following example.

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

Ada rudely disagrees with Eva's computation, but does not politely redress it or justify it. In response, Eva retaliates immediately with a rude disagreement of her own.

Agreements consistently showed negative effects on NICs in all time periods. These agreements were often in the form of simple confirmations such as "uh-huh" or repetitions such as "one ten, right."

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

Students like Jack often gave brief confirmations without further elaboration or NICs, suggesting that they were almost reflexively agreeing out of social preference.

As argued earlier, wrong new ideas can lead to NICs if the group can identify the flaws. Wrong ideas by the previous speaker reduced the likelihood of agreements and increased the likelihood of rude disagreements, suggesting that groups generally recognized their flaws. In the following segment, a student multiplies 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 recognizes 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 helps Rex detect and correct the flaw in the number of hours from five to two. Building on Amy's partially correct idea to multiply the helicopter speed by five hours, Rex creates a NIC.

Group members' questions tended to reveal their own knowledge gaps more than identify gaps in the group's knowledge as these questions tended to lower the likelihood of a NIC, especially in unsuccessful groups. Compared to unsuccessful groups, a respondent to a question in a successful group was more likely to evaluate the ideas in the question correctly, resulting in agreement by the following speaker. Consider the following example.

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

When John questioned a computation, Ron affirmed its validity via the rate-time-distance relationship. John understood ("oh!") and agreed with it ("ok"). In unsuccessful groups however, the second respondent to a question (the same person 59% of the time) not only was less likely to give a NIC or a justification but also more likely to disagree rudely, suggesting dissatisfaction with the answer to the question. 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. Not satisfied with that answer, Beth challenged Mark with a blunt rejection, declaring that Mark was wrong as the problem statement did not specify that multiplication. In short, the lower likelihood of NICs in unsuccessful groups than in successful groups after a question (lag 2) might be due to less satisfactory responses.

Lastly, students who justified their ideas were far more likely to have NICs in successful groups than in unsuccessful ones, possibly due to the

quality of their justifications. Successful groups often referred to mathematical relationships to justify their ideas ("rate times time is how far the ship moves" from 2 transcript segments earlier). In contrast, students in unsuccessful groups often justified their claims by citing authority, such as the teacher, the textbook, or the problem statement ("that's what the problem said" from previous transcript segment). Justifications based on mathematics might be more likely to be valid than those based on authority, which would explain why students who justified their answers were more likely to have NICs in successful groups than in unsuccessful groups.

Discussion

Past theoretical models have highlighted the importance of correct, new ideas to group problem solving (e.g., Chiu, 2000a, 2001). By understanding how group processes affect a group's creation of correct, new ideas, educators can help students engage in beneficial processes to create correct, new ideas and avoid harmful processes that hinder their creation. This study replicates past studies by showing the effect of correct, new ideas on group outcomes, solution score in this case. More important, this study extends this line of research by systematically analyzing the group *processes* that affected this process, the creation of correct, new ideas. In particular, this study highlighted the importance of correct evaluations, incorrect ideas and the recent temporal context.

Group Processes Affecting the Creation of Correct, New Ideas

The first analysis showed that groups with more new ideas that were correct tended to solve the problem correctly, controlling for past mathematics achievement. Having verified the importance of correct, new ideas in this data set, the second analysis examined their determinants during the problem solving process. When students justified their ideas, evaluated other group members' ideas, proposed wrong ideas and asked questions, they

affected the likelihood of a correct, new idea. Students who justified their ideas tended to create correct, new ideas and help the next speaker create one, suggesting that justifications both helped students validate their own ideas and provided the bases for other students' correct and new ideas. Justifications also facilitated rational discussions by eliciting further justifications and by reducing rude disagreements.

Evaluations also affected the likelihood of a correct, new idea. When students evaluated ideas correctly, other group members were more likely to create correct, new ideas and less likely to create wrong, new ideas. Students also tended to recognize and agree with correct evaluations, as well as build on them with further correct evaluations and justifications. Furthermore, the effects of correct evaluations on correct, new ideas also lasted longer than other actions. Recent correct evaluations were also more important than recent correct ideas to creating correct, new ideas. Together, these results support the view that correct evaluations identified flaws or verified correct ideas, thereby creating a valid basis of shared understanding for making further correct, new ideas.

Disagreements may have identified obstacles for the group to address (De Lisi & Goldbeck, 1999; Piaget, 1985), but only group members who did so politely tended to express correct, new ideas. Those who disagree rudely were less likely to do so. Furthermore, Chiu and Khoo (2003) showed that other group members were also more likely to respond to rude disagreements by retaliating with rude disagreements. These results suggest that rude disagreements hindered the introduction of NICs.

Students' concerns over face also showed in their agreements. When students agreed, they tended not to create correct, new ideas consistent with Chiu and Khoo's (2003) study showing that people tend to agree even when the proposed idea is wrong. This result suggests that their social motives inclined them to prefer agreements, sometimes reflexively with simple confirmations (Burgoon et al., 1993; Chiu, 2001).

Group members' wrong ideas often led to correct, new ideas. After a new, wrong idea, students also tended to disagree rudely and not to agree.

These results suggest that group members tended to recognize flaws and incorporate the useful parts into a correct, new idea. Thus, incorrect ideas often served as grist for creating correct new ideas, and this benefit outweighed the danger of accepted wrong ideas reducing the number of correct, new ideas.

Questions typically showed individual lacunae rather than group lacunae as group members tended to explain earlier ideas to one another rather than create correct, new ideas. After a group member asked a question, successful groups were more likely than unsuccessful groups to give explanations. This result suggested that members in successful groups were more likely to understand the current problem solving and hence more likely to create a correct, new idea later.

The local temporal context of a group's recent actions affected the likelihood of a correct, new idea, whereas a priori student properties and group did not. For example, correct evaluations from three speaker turns ago raised the likelihood of a correct, new idea in the current turn. These results highlight the importance of examining actual student actions and interactions rather than inferring them from coarser measures, such as student or group properties.

Differences Across Groups and Across Time Periods

This analysis showed differences across groups and across time periods. Successful groups differed from unsuccessful groups both cognitively and socially. Cognitively, successful groups had higher mean past mathematics achievement, more new ideas that were correct and justifications that were more effective in yielding correct, new ideas. Socially, questions in successful groups were more likely to yield satisfactory responses (correct evaluations and agreement) rather than rude disagreements.

The frequency of correct, new ideas and the effects of two predictors varied across time periods. Most of the correct, new ideas were clustered together as time period variation accounted for 79% of the total variance.

Also, the size of the effects of agreement and of correct evaluations (lag 2) varied across time periods. Otherwise, all other predictors showed consistent effects of similar sizes across groups and across time periods. These results showed that the effects of some predictors varied across contexts while others showed consistent effects across contexts.

Conceptual and Methodological Implications for Researchers

This study models conceptual relationships among group processes affecting the process of creating correct, new ideas within the local temporal context and introduces new methods for analyzing them, suggesting several implications for researchers. This study highlights the importance of correct evaluations from several turns ago and wrong ideas as key elements for researchers to consider when building a comprehensive theory of interactions among group processes. Moreover, the model shows the relative effect sizes of justifications, evaluations, wrong ideas and questions on correct new ideas. Furthermore, this study highlighted the importance of local temporal context by showing how recent actions by the three most recent speakers affected a student's creation of correct, new ideas. This model also captures similarities and differences of predictor effects across groups and time periods. In particular, breakpoints suggested that critical events divided the problem solving session into distinct time periods in which correct, new ideas were prevalent or rare.

In this study, I analyzed group processes systematically with a new method, dynamic multi-level Logit/Probit, that identifies breakpoints and models social interactions (Chiu, 2001). The breakpoint method helps researchers identify critical events that radically change the nature of a group's problem solving, such as on-task ↔ off-task transitions, insights or points of confusion. Meanwhile the use of multi-level Logit/Probit, lag variables and serial correlation tests allows researchers to model social interactions within the local context of recent group actions while accounting for both group and time period differences.

Practical Implications for Teachers

This study also identified several student actions during GPS that raised or reduced the likelihood of correct, new ideas, suggesting several implications for teachers. Teachers can encourage students to express and justify their own ideas while evaluating those of others carefully and politely. This study showed that group problem solving benefits from free expression of new ideas, even wrong ones. Thus, teachers might facilitate this free expression with a safe and supportive classroom culture. As justifications often accompany correct, new ideas, teachers can foster classroom conversations among students in which they are expected to justify their answers.

This study also showed that correct evaluations had far-reaching beneficial consequences, suggesting that students can benefit from regular evaluations of one another's ideas. Lastly, teachers can encourage students to consider new ideas slowly to discourage immediate confirmations or rude rejections. Together, these changes might help students realize the potential benefits of cooperative learning.

References

Asch, S. E. (1952). *Social Psychology.* Englewood Cliffs, NJ: Prentice Hall.

Bargh, J. A., & Schul, Y. (1980). On the cognitive benefit of teaching. *Journal of Educational Psychology, 72*, 593–604.

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage.* New York: Cambridge University Press.

Burgoon, J. K., Dillman, L., & Stern, L. (1993). Adaptation in dyadic interaction: Defining and operationalizing patterns of reciprocity and compensation. *Communication Theory, 3*, 295–316.

Chiu, M. M. (1997). *Building on diversity: A moment-to-moment analysis of students collaboratively solving mathematics problems.* Los Angeles: University of California. (ERIC Documentation Reproduction Service 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: Nova Science.

Chiu, M. M., & Khoo, L. (2003). Rudeness and status ef fects during group problem solving: Do they bias evaluations and reduce the likelihood of correct solutions? *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, DC: 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, NJ: Erlbaum.

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

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

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.

Goffman, E. (1959). *The presentation of self in everyday life*. Garden City, NY: Doubleday.

Goldbeck, S. (in press). Peer collaboration and children's representation of the horizontal surface of liquid. *Journal of Applied Developmental Psychology, 19*, 573–594.

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

Goldstein, H., & Rasbash, J. (1996). Improved approximations for multi-level models with binary responses. *A Journal of the Royal Statistical Society, 159*, 505–513.

Goodwin, C., & Goodwin, M. H. (1987). Concurrent operations on talk: Notes on the interactive organization of assessments. *IPRA Papers in Pragmatics, 1*(1), 1–54.

Gottman, J. M., & Krokoff, L. J. (1989). Marital interaction and satisfaction: A longitudinal view. *Journal of Consulting and Clinical Psychology, 57*(1), 47–52.

Harwood, D. (1995). The pedagogy of the world studies 8-13 project: The influence of the presence/absence of the teacher upon primary children's collaborative group work. *British Educational Research Journal, 21*, 587–611.

Hastie, R. (1986). Experimental evidence on group accuracy. In B. Grofman & G. Owen (Eds.), *Information pooling and group decision making* (pp. 129–157). Greenwich, CT: JAI.

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 emergining conceptualization of groups as information processors. *Psychological Bulletin, 121*, 43–64.

Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika, 75*, 800–802.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandanavian Journal of Statistics, 6*, 65–70.

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.

Judge, G. G., Griffiths, W. E., Hill, R. C., Lutkepohl, H., & Lee, T. C. (1985). *The theory and practice of econometrics* (2nd ed.). New York: Wiley.

Laughlin, P. R., VanderStoep, S. W., & Hollingshead, A. B. (1991). Collective versus individual induction: Recognition of truth, rejection of error, and collective information processing. *Journal of Personality and Social Psychology, 61*, 50–67.

Lindow, J. A., Wilkinson, L. C., & Peterson, P. L. (1985). Antecedents and consequences of verbal disagreements during small group learning. *Journal of Educational Psychology, 77*, 658–667.

Person, N. K., Kreuz, R. J., Zwaan, R., & Graesser, A. C. (1995). Pragmatics and pedagogy: Conversational rules and politeness strategies may inhibit effective tutoring. *Cognition and Instruction, 13*, 161–188.

Piaget, J. (1985). *The equilibrium of cognitive structures: The central problem of*

intellectual development (T. Brown & K. L. Thampy, Trans.). Chicago: University of Chicago Press.

Pomerantz, A. (1984). Agreeing and disagreeing with assessments: Some features of preferred/dispreferred turn shapes. In J. M. Atkinson & J. Heritage (Eds.), *Structures of social action: Studies in conversation analysis* (pp. 57–101). Cambridge: Cambridge University Press.

Pontecorvo, C. (1993). Forms of discourse and shared thinking. *Cognition and Instruction, 11*, 189–196.

Scott, W. (1955). Reliability of content analysis: The case of nominal scale coding. *Public Opinion Quarterly, 17*, 321–325.

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

Stasson, M. F., & Bradshaw, S. D. (1995). Explanation of individual-group performance differences: What sort of "bonus" can be gained through group interaction? *Small Group Research, 26*, 296–308.

Tann, S. (1979). A Study of group work. *Forum for the Discussion of New Trends in Education, 21*(3), 91–95.

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., & Farivar, S. (1999). Developing productive group interaction in middle school mathematics. In A. M. O'Donnell & A. King (Eds.), *Cognitive perspectives on peer learning* (pp. 117–149). Mahwah, NJ: Erlbaum.