Statistical Discourse Analysis

Ming Ming Chiu
Purdue University

Using traditional statistical analyses, researchers have examined learning outcomes and have summarized aggregates of processes that are linked to outcomes, but few statistical analyses examine the dynamic relationships among students' and teachers' processes that contribute to these learning outcomes. Consider the following research questions that focus on these dynamic relationships: Which characteristics of a classroom, or a group of students, or an individual, or recent actions might influence the likelihood of a desirable (or undesirable) student action? During group discussions, which student actions radically improve or harm subsequent interactions? How do a teacher's reflections influence his or her teaching practice? In this chapter, I examine how statistical discourse analysis (SDA) can be used to address these questions.

Answers to these classes of questions can help teachers and students improve one another's learning. With these answers, teachers can recognize when classroom conversations are productive or problematic, reflect on specific conversation segments, and develop suitable interventions for specific types of situations. Likewise, teachers can use these answers to help students develop their metacognitive skills (monitoring, self-regulation, other-regulation; Chiu & Pawlikowski, in press). With improved metacognitive skills, students can better evaluate the current classroom conversation, identify problematic areas, and respond accordingly to improve both their own and their classmates' learning.

Although qualitative studies have generated answers to these questions for a few cases, statistical methods are needed to analyze large data sets to test whether these preliminary answers apply across people and contexts. Videotaped face-to-face conversations, online forums, Twitter tweets, and other social interactions can yield data sets with millions of data points that require quantitative methods to analyze in a reasonable amount of time. However, such quantitative analyses face many statistical challenges that were not fully addressed until recently. In particular, the above research questions exemplify several general methodological issues:

- Do variables at multiple levels (conversation turn, person, group, class, school, etc.) account for the likelihoods of desirable (or undesirable) actions?
- Do sequence(s) of prior actions (micro-time context) affect the likelihoods of target actions?
- Do the relationships among explanatory variables and target actions differ across contexts?
- Do the likelihoods of target actions differ across time?

- Do the relationships among explanatory variables and target actions differ across time?
- Do the explanatory variables show indirect effects on a target action through intermediate variables?

Although past studies have tried to address these questions by analyzing data across time with conditional probabilities (e.g., Parks & Fals-Stewart, 2004), sequential analysis (e.g., Gottman & Roy, 1990), logit regressions (e.g., Pevalin & Ermisch, 2004), nonlinear dynamic models (Gorman, Amazeen & Cooke, 2010), and pattern analysis (Stachowski, Kaplan & Waller, 2009), these methods do not adequately address differences across groups, differences across individuals, characteristics of sequences of recent conversation turns, or differences across time (Mercer, 2008; Reimann, 2009). As shown below, SDA addresses all of these issues (Chiu, 2008; Chiu & Khoo, 2005).

In the next section, I discuss these statistical challenges in greater detail, followed by explication of how SDA addresses each of these challenges. I then show how SDA can provide unique insights from face-to-face dialogue data, asynchronous online forum data, and individual teacher reflection data by presenting examples from three different studies.

Addressing Analytic Difficulties With Statistical Discourse Analysis

Statistically analyzing interaction processes requires addressing a range of analytic difficulties regarding the entire data set, dependent variables, and explanatory variables (see Table 1 for a summary). Data issues include missing data, content analysis (coding), parallel talk, and the tree structure of online messages. Difficulties involving outcomes include nested data, differences across time, discrete outcomes, infrequent outcomes, and multiple outcomes. Explanatory variable issues include sequences, false positives, context-dependent effects, indirect effects, and robustness of results. SDA addresses each of these analytic difficulties as follows.

Overall Data Set Issues

SDA addresses the overall data set issues (missing data, content analysis (coding), parallel talk, message trees) with Markov-chain Monte Carlo multiple imputation, multidimensional coding, and storage of the interaction structure. Missing data can (a) bias the results, (b) raise the complexity of the data analyses, and (c) reduce estimation efficiency (Peugh & Enders, 2004). Markov-chain Monte Carlo multiple imputation is used to estimate the values of the missing data, which addresses this issue more effectively than deletion, mean substitution, or simple imputation, according to computer simulations (Peugh & Enders, 2004).

Ideally, the content analysis uses a unit of analysis with clear boundaries (e.g., an online message) and categories that are mutually exclusive, exhaustive, and sufficiently comprehensive to test one's hypotheses. As the number and complexity of categories rise, however, (a) the training time for coders and the overall coding time rise, (b) coding conflicts rise, (c) internal consistency and intercoder reliability fall, (d) degrees of freedom in the explanatory model fall, and (e) precision falls (Chiu & Khoo, 2005).

Table 1. Addressing Each Analytic Difficulty With Statistical Discourse Analysis

Analytic Difficulty	Statistical Discourse Analysis Strategy
Data set	
• Missing data (0110??10)	• Markov-chain Monte Carlo multiple imputation (Peugh & Enders, 2004)
• Complex categories/codes (A, B, …Q)	• Multidimensional content analysis
Dependent variables	
• Nested data (turns of talk within time periods, students within groups within classes within schools, etc.)	• Multilevel analysis (a/k/a hierarchical linear modeling; Bryk & Raudenbush, 1992; Goldstein, 1995)
• Differences across time periods ($T_1 \neq T_2$)	• Breakpoint analysis (Chiu & Khoo, 2005) and multilevel cross-classification (Goldstein, 1995)
• Serial correlation ($t_3 \sim t_4$)	• I2 index of Q statistics (Huedo-Medina et al., 2006)
• Discrete variable (yes/no)	• Logit (Kennedy, 2008)
• Infrequent dependent variables	• Logit bias estimator (King & Zeng, 2001)
• Multiple dependent variables at the same level (Y_1, Y_2,…)	• Multivariate outcome model (Goldstein, 1995)
• Dependent variables at different levels	• Separate analyses at each level
• Parallel talk ($\rightarrow\rightarrow\rightrightarrows\rightrightarrows$) OR message trees ($\Lambda$)	• Store conversation turn (or message) to which a conversation turn (or message) responds
Explanatory variables	
• Sequences of turns of talk (X_{t-2} or $X_{t-1} \rightarrow Y_0$)	• Vector autoregression (Kennedy, 2008)
• False positives (Type I errors)	• Two-stage linear step-up procedure (Benjamini et al., 2006)
• Moderation effects across levels (e.g., Student × Turn of Talk)	• Random-effects model (Goldstein, 1995)
• Indirect, multilevel mediation effects (X $\rightarrow$ M $\rightarrow$ Y)	• Multilevel M tests (MacKinnon et al., 2004)
• Robustness of relationships	• Single outcome, multilevel models • Testing on subsets of the data • Testing on original data

By using a multidimensional coding scheme with corresponding decision trees, SDA can capture the data's complexity, reduce the number of needed variables, increase intercoder reliability, and thereby model complex phenomena. For example, individual actions can be coded on three dimensions (evaluation, knowledge content, invitational form; Chiu, 2000). Because each dimension has three categories, this framework can capture 27 different types of action. By coding one dimension at a time with a decision tree, a coder uses clear criteria to choose among only three possible codes (instead of 27). Thus, multidimensional coding reduces training time, reduces overall coding time, and likely increases intercoder reliability.

Conversations do not always proceed neatly with one turn of talk following another. Students can split off to engage in parallel conversations, and multiple online messages can respond to one message (see Figure 1). To address this issue, the database includes a variable that identifies and stores the conversation turn to which a conversation turn responds; for example, Turn 4 responds to Turn 2 (not Turn 3). Likewise, storing the message to which a message responds captures the tree structure of online messages.

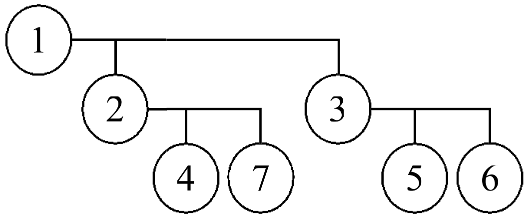

Figure 1. Parallel talk or tree of online messages. A circle indicates a turn of talk or message. The number indicates the temporal order, and the lines indicate the referent. For example, Turn 4 (or Message 4) responds to Turn 2 (not Turn 3).

Dependent Variable Issues

SDA addresses the dependent variable issues (nested data, time, discrete variables, infrequent events, and multiple dependent variables) with multilevel analysis, breakpoint analysis, an I^2 index of Q statistics, multilevel cross-classification, logit, logit bias estimation, and multivariate outcomes models.

Social interactions can involve nested data (conversation turns within time periods, individuals within groups within classrooms, etc.). Failure to account for similarities in actions within the same time period (vs. different time periods), by the same person (vs. different people), within the same group (vs. different groups), or within the same classroom (vs. different classrooms) can underestimate the standard errors (Goldstein, 1995). SDA addresses this issue by modeling nested data with a multilevel analysis (Goldstein, 1995; cf. hierarchical linear modeling, Bryk & Raudenbush, 1992).

Because these outcomes can differ across time, they require identification of distinct time periods, modeling of time period differences, and modeling of recent events (Chiu & Khoo, 2005). SDA statistically identifies the pivotal moments that separate the data into time periods of high versus low frequencies of the outcome variable(s) (Chiu, 2008; Wise & Chiu, 2011).

Because adjacent conversation turns are often more similar than conversation turns that are far apart from one another, failure to model this similarity (serial correlation of errors) can bias the results (Kennedy, 2008). An I^2 index of Q statistics tests all groups simultaneously for serial correlation of residuals in adjacent events (Huedo-Medina, Sanchez-Meca, Marin-Martinez, & Botella, 2006). If the I^2 index shows significant serial correlation, adding the value of the outcome variable of the previous conversation turn often eliminates the serial correlation (e.g., when modeling the outcome variable *justification*, add whether it occurs in the previous turn [*justification*(–1)] Chiu & Khoo, 2005). Then, SDA models these differences across time with a multilevel cross-classification (Goldstein, 1995).

The outcomes are often discrete (a justification occurs in a conversation or it does not) rather than continuous (e.g., test scores), so standard regressions such as ordinary least squares can bias the standard errors. To model discrete outcome variables, a logit regression is used (Kennedy, 2008). However, infrequent outcomes can bias the results of a logit regression (King & Zeng, 2001). For infrequent outcomes, the logit bias is estimated and removed (King & Zeng, 2001).

Multiple outcomes can have correlated residuals that underestimate standard errors (Goldstein, 1995). If the outcomes are from different levels, analyzing them in the same model overcounts the sample size of the higher level outcome(s) and results in biased standard errors. If multiple outcomes are from different levels of analysis, separate analyses must be performed at each level. To model multiple outcomes at the same level of analysis, a multivariate outcome, multilevel cross-classification is used (Goldstein, 1995).

Explanatory Variable Issues

SDA addresses the explanatory variable issues (sequences, indirect effects, false positives, and robustness) with a vector autoregression (Kennedy, 2008), multilevel M tests, a two-stage step-up procedure, and robustness methods. A vector autoregression combines characteristics of sequences of recent conversation turns into a local time context (*micro-time context*) to model how they influence the subsequent conversation turns. For example, the likelihood of a new idea in a conversation turn might be influenced by characteristics of earlier conversation turns (e.g., disagreement in the previous turn) or earlier speakers (e.g., past mathematics achievement of the speaker in the previous turn).

Single-level mediation tests can detect indirect effects (rather than direct effects), but applying these tests to nested (multilevel) data can bias results downward. To test whether explanatory variables show indirect effects through intermediate variables properly, SDA uses multilevel M tests (MacKinnon, Lockwood, & Williams, 2004).

Testing many hypotheses also increases the risk of false positives (Type I errors; Benjamini, Krieger, & Yekutieli, 2006). The two-stage linear step-up procedure reduces false positives more effectively than 13 other methods, according to computer simulations (Benjamini et al., 2006).

Last, results from one analysis are not necessarily robust. To test the robustness of the results, three variations of the final model can be used. First, a single-outcome, multilevel analysis can be performed for each outcome variable. Second, subsets of the data (e.g., halves) can be run separately to test the consistency of the results for each subset. Third, the analyses can be repeated for the original data set (without the estimated data).

Three Studies

I now show how SDA can be applied to address the three research questions in the introduction. After explicating one analysis and its results in detail, I summarize the other two analyses and results.

Example 1: Modeling Students' New Ideas and Justifications

In the following example, I test whether recent social metacognitive group processes (evaluations, questions, commands) facilitate or hinder *micro-creative processes* (creating *new ideas* and assessing their utility via *justifications*). Whereas individual metacognition is monitoring and regulating one's own knowledge, emotions, and actions (Hacker & Bol, 2004), *social metacognition* is *group members*' monitoring and regulating *one another's* knowledge, emotions, and actions (Chiu & Kuo, 2009).

Data. Eighty-four ninth grade students (43 girls, 41 boys) completed a survey about social status and were videotaped during their algebra classes in an urban high school in the United States. They worked in groups of 4 on the following problem for 30 minutes and produced a total of 3,296 conversation turns.

> Under the Universal plan, each text message costs \$.10. Budget costs \$.01 per text message, but charges a monthly fee, \$18.
> (1) How many text messages do you send each month?
> (2) Which company should you use?
> (3) How many texts should you send for the Universal plan and the Budget plan to cost the same?

To solve the third part of this problem, one can equate the total costs for each company ($0.10t = 0.01t + 18$), to obtain 200 text messages ($18/[0.10t - 0.01t]$).

Variables. In addition to individual, group, and classroom variables (e.g., respectively, gender, mean of mathematics grades in a group, classroom identifier), two research assistants transcribed the videotapes and coded the transcripts to create variables at the conversation turn level. They divided each transcript into sequences of words and/or actions by a person that are bracketed by those of others (*conversation turns*). Then, they coded each turn along five dimensions: evaluation of the previous action (agree, politely disagree, rudely disagree, ignore, neutral), knowledge content (old idea, new idea, is unrelated to the problem idea [e.g., "Are you hungry?"]), validity (right, wrong, is unrelated to the problem idea), justification (or none), and invitational form (question, command, statement). Interrater reliability was computed with Krippendorff's (2004) α, which can be applied to incomplete data, any sample size, any measurement level, any number of coders or categories, and scale values. See Chiu (2001, 2008) for variable descriptions, coding decision trees, and other details.

Explanatory model. I applied the following multilevel, cross-classification, logit analysis (Goldstein, 1995):

$$\text{Prob}(\textbf{Action}_{yijkl}) = F(\beta_y + f_{yjkl} + g_{ykl} + h_{yl}) + e_{yijkl} \qquad (1)$$

The probability that action y (e.g., new idea) occurs at conversation turn i in time period j by student k in group l is the logit or probit link function (F) of the mean β_y, the unexplained components for each time period f_{yjkl}, student g_{ykl}, group h_{yl}, and conversation turn e_{yijkl}. Next, I tested sets of explanatory variables (vectors):

$$\begin{aligned}\text{Prob}(\textbf{Action}_{yijkl}) &= F(\beta_y + f_{yjkl} + g_{ykl} + h_{yl}\beta_{ys}\textbf{Classroom}_{yl} + \beta_{yt}\textbf{Group}_{yl} \\ &+ \beta_{yujkl}\textbf{Current\_turn}_{yijkl} + \beta_{yvjkl}\textbf{Earlier\_turn}_{y[i-1]jkl} \\ &+ \phi_{yvjkl}\textbf{Earlier\_turn}_{y[i-2]jkl}) + e_{yijkl}\end{aligned} \qquad (2)$$

First, classroom identification variables (*classroom #1*, *classroom #2*, and *classroom #3*) were added as control variables (**Classroom**). Second, I entered group-level variables: *numbers of girls*, *Blacks*, *Asians*, and *Latinos*; *mean past mathematics grade*; *mean social status*; *gender variance*; *racial variance*; and *unsolved* (**Group**). "*Unsolved*" refers to groups that did not solve the problem. Third, I tested for interaction effects among pairs of significant variables in **Group**. All nonsignificant variables and interactions were removed.

Next, I added current conversation turn variables: *gender*, *Black*, *Asian*, *Latino*, *mathematics grade*, *social status*, *relative mathematics status*, *relative social status*, *agree*, *politely disagree*, *rudely disagree*, *ignore*, *question*, and *command* (**Current_turn**). Because a variable cannot be used to predict itself, *new idea* was not added as an explanatory variable to model the dependent variable *new idea* (likewise *justification* was not an explanatory variable for *justification*). To identify cross-level moderation effects, I tested if each conversation turn–level regression coefficient ($\beta_{yujkl} = \beta_{yu} + f_{yujkl} + g_{yukl} + h_{yul}$) differed significantly at the time period level ($f_{yujkl} \neq 0$?), individual level ($g_{yu0kl} \neq 0$?), or group level ($h_{yu00l} \neq 0$?) (Goldstein, 1995).

Next, I tested whether characteristics of earlier conversation turns were linked to the likelihoods of a new idea or a justification in the current turn by entering lag variables for the previous speakers (vector autoregression; Kennedy, 2008): *gender(–1)…command(–1)* (**Earlier_turn**, similar to **Current_turn** except that both *new idea[–1]* and *justification[–1]* are included). Then, I repeated the procedure for lag –2 of the variables in **Earlier_turn** and so on for earlier lags (–3, –4,…) until no explanatory variable was significant.

An α level of .05 was used. I reported the predictive accuracy of the final model: the model's predicted versus actual presence or absence in each conversation turn of a *new idea* (and of a *justification*). To facilitate interpretation of these results, the odds ratio of each variable's total effect (*E*; direct plus indirect) was reported as the percentage increase or decrease (+*E*% or –E%) in the likelihoods of a *new idea* and a *justification* (Kennedy, 2008). With 3,296 turns, the statistical power was greater than 0.99 at the conversation turn level, even for a 0.1 effect size (Konstantopoulos, 2008).

Results. The groups averaged 2.65 new idea pivotal moments (3.65 time periods) and 2.05 justification pivotal moments (3.05 time periods), for a mean of 6.70 time periods per group, showing that *new ideas* and *justifications* varied widely within each group's problem-solving session. Still, different classrooms had similar numbers of new ideas, as did different groups and different individuals. Of the total variance of *new idea*, less than 0.1% occurred across classrooms, groups, or individuals (also less than 0.1% for *justification*). Instead, different time periods had substantially different numbers of *new ideas*; 59% of the variance of new ideas occurred across time periods. The remaining 41% occurred within time periods (across different turns of talk). Likewise, different time periods had substantially different numbers of *justifications*; 60% of the variance occurred across time periods and 40% within time periods.

Social status, *command*, *agreement*, and *rude disagreement* were all linked to new ideas (see Figure 2). Students with higher *social status* were more likely to voice *new ideas*. When students *agreed* or *rudely disagreed*, they were less likely to express *new ideas*. After a *command*, a *new idea* was less likely. The effect of *rude disagreement* on the likelihood of a new idea in the next conversation turn depends on whether the group solves the problem or not (*unsolved*) and on whether there was a wrong idea two conversation turns ago (*wrong[–2]*).

Meanwhile, greater *social status*, *higher mathematics grade(–1)*, *polite disagreement*, and *command(–1)* were linked to *justification*. A student whose *social status* exceeded the mean

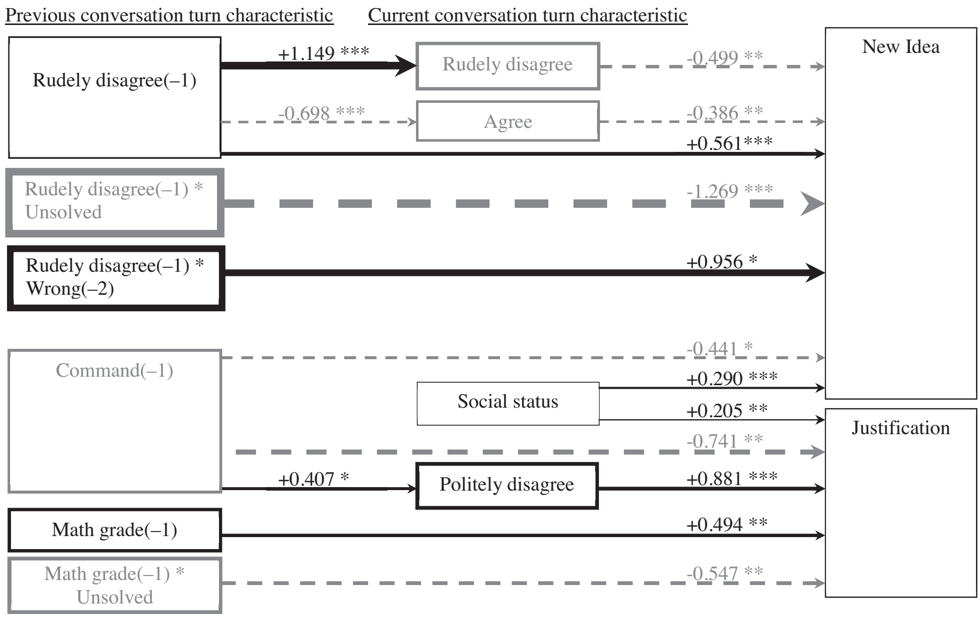

Figure 2. Path analysis of significant predictors of new ideas and justifications using multivariate, multilevel logit. Values are standardized parameter coefficients. Black, solid arrows indicate positive direct effects, and gray, dashed arrows indicate negative direct effects. Thicker lines indicate larger effect sizes. The total effect of an explanatory variable on a dependent variable is the sum of its direct effects and its indirect effects. For an indirect effect of $X \rightarrow M \rightarrow Y$, compute the indirect effects by multiplying the coefficient of $X \rightarrow M$ and the coefficient of $M \rightarrow Y$. For example, the indirect effect of Command(–1) → Politely disagree → Justification is .407 × .881 = .389. For interaction effects, sum the coefficients of the applicable explanatory variables and interaction variables. For example, the relationship between math grade(–1) and justification for groups that did not solve the problem (unsolved) is the sum of the math grade(–1) coefficient (.494) and the math grade(–1) × Unsolved coefficient (–.547), which is –.053. In contrast, the relationship between math grade(–1) and justification for groups that solved the problem is only the math grade(–1) coefficient (+.494).

by 10% expressed more *justifications* (+1%, odds ratio computation using the regression coefficient, +.290). Note that this +1% applies to *each* conversation turn, so the cumulative greater likelihood of a student with 10% higher *social status* than the mean expressing at least one more *justification* than other students during their average of 39 conversation turns is +33%.[1]

Meanwhile, group members whose *mathematics grade(–1)* exceeded the mean by 10% helped others create more *justifications* (+2%) in groups that solved the problem but slightly fewer justifications in groups that did not solve the problem (–0.2%), showing a group context difference. The greater use of *justifications* in successful groups suggests that members adaptively responded to members with higher mathematics status by making more *justifications*. Note that group members with higher *math grades* did not express more *new ideas* or more *justifications* than other group members. Thus, in successful groups, the greater number of *justifications* stemmed from students with *lower grades* responding to group members with higher *mathematics grades*. Together with the positive links between *social status* and both *new ideas* and *justifications*, these results suggest the importance of other skills (e.g., emotional and social skills) in addition to mathematics skills for successful collaborative solutions to mathematics problems.

These *justification* results also support both politeness theory and status theory (Berger & Fisek, 2006; Brown & Levinson, 1987; Chiu & Khoo, 2003). Unlike *rude disagreements*, *polite disagreements* yielded more *justifications* (+13%, largest effect size), and *commands(–1)* yielded fewer *justifications* (–6%). *Polite disagreements* were also more likely to follow *commands* (+6%), suggesting a mutual recognition of a status difference in which the lower status participant responds to a higher status participant's *command* by being polite and giving a *justification* (Berger & Fisek, 2006; Brown & Levinson, 1987).

No other variable showed significant effects, and only the effect of *mathematics grade(–1)* on the likelihood of a *justification* in the next conversation turn was higher in some time periods than in others. When *mathematics grade(–1)* exceeded the mean by 10%, the effect on *justifications* ranged across time periods from –1% to +3% in groups that solved the problem and from –2% to +1% in groups that did not solve the problem. This model had a 77% accuracy rate for correctly predicting the occurrence or absence of a *new idea* in each conversation turn. For *justification*, the accuracy rate was 89%.

Example 2: Pivotal Moments During Online Forums

SDA can also test whether specific messages during asynchronous online discussions radically changed the interaction patterns, dividing the discussion into distinct message sequences that differ substantially (I call such messages *pivotal messages*). Consider Wise and Chiu's (2011) study of 21 participants in two online discussion groups at a university in western Canada. The 8 women and 13 men were students in a Foundations of Educational Technology course that met in a classroom and online. Each week, both groups worked collectively to create a suitable educational design (e.g., an activity plan). For example, during 1 week, they designed a set of activities to help a group of 10-year-olds become "experts" in the Chinese zodiac.

All 252 posts in the discussions were coded using a five-phase knowledge construction scheme (Gunawardena, Lowe, & Anderson, 1997). Then, SDA was applied to identify pivotal messages that initiated new discussion phases (different knowledge construction phases). The results in Figure 3 show that students playing one of two summarizing roles (*Synthesizer* and *Wrapper*) often wrote *summaries* that were pivotal messages, elevating the discussion to more advanced phases of knowledge construction.

Example 3: Teacher Reflections and Practice

In addition to testing micro-relationships among group processes, SDA can also be used to test relationships among an individual's behaviors over time. For example, SDA can also test whether teacher reflection and practice occur in specific sequences. Hayden and Chiu (2011) examined the relationship between novice teachers' teaching practices and their reflections written immediately after their teaching. These teachers taught once a week in a

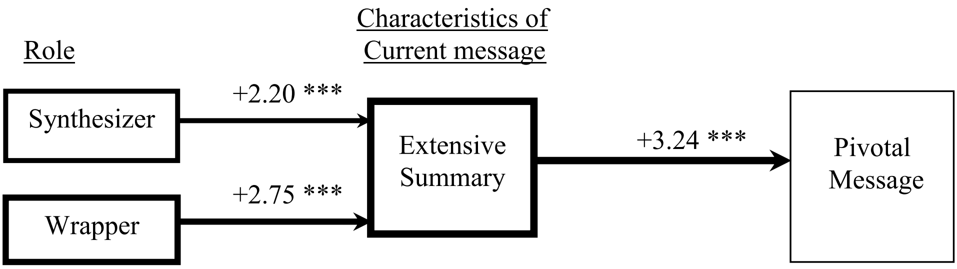

Figure 3. Path diagram of final model predicting pivotal messages. Numbers shown are regression coefficients. Thicker box and arrow lines indicate larger, positive effect sizes. *$p < .05$. **$p < .01$. ***$p < .001$.

reading clinic for 8 weeks. Weekly written reflections ($N = 175$) from 23 preservice teachers were examined to test whether a *Reflective Problem Exploration → Teaching Adaptation → Problem Resolution* sequence occurred. SDA showed that a novice teacher who engaged in more *reflective problem explorations* or *teaching adaptations* in the previous week, or more *teaching adaptations* in the current week, reported more *problem resolutions* (see Table 2). These results show that *reflective problem exploration* and *teaching adaptation* both preceded *problem resolution*, and that *reflective problem exploration* both preceded and co-occurred with *teaching adaptation*, showing some support for the *Reflective Problem Exploration → Teaching Adaptation → Problem Resolution* sequence.

Discussion

Complementing traditional statistical analyses of discussion characteristics with individual learning outcomes, SDA examines the dynamic relationships among students' and teachers' observable behaviors during learning episodes. In the studies discussed here, SDA tested models of desirable (or undesirable) actions using explanatory variables at multiple levels (conversation turn, student, teacher, school) with sequences of prior actions (micro-time context). Furthermore, SDA examined whether these dependent variables differ across time, whether these relationships differ across contexts or across time, and whether indirect effects occur. In short, SDA tests systematic, explanatory models of multiple target actions and processes that can account for differences across contexts and across time.

The SDA results also show that groups can be compared not only at the group level but at the process level by testing whether relationships among processes differ across groups. These differences in processes not only uncovered different relationships among processes across groups but also detected subtle relationships that are not visible in coarser grained analyses at the individual or group levels. For example, in groups whose members had higher mathematics grades, the greater number of justifications stemmed from students with *lower grades* responding to group members with higher grades, not from the group members with higher grades.

Although *rude disagreement* has only short-term effects on *new idea* (in the first study), a *summary* can be a pivotal message that radically changes a group's interactions and yields

Table 2. Summary of Multilevel Regression Models Predicting Problem Resolutions With Unstandardized Regression Coefficients (Standard Errors in Parentheses)

	Three Multilevel Regression Models of Problem Resolutions		
Explanatory Variable	**Model 1**	**Model 2**	**Model 3**
Teacher characteristics			
Some teaching experience	.29 (.24)		
Total years in school	.06 (.06)		
Graduate student	–.27 (.14)*	–.19 (.09)*	–.21 (.10)*
Student characteristics			
Black student	–.27 (.45)		
White student	–.22 (.39)		
Female	.26 (.09)**	.15 (.07)*	.16 (.08)*
Free/reduced-price lunch	–.19 (.08)*	–.19 (.08)*	–.18 (.09)*
Grade	.08 (.03)*	.06 (.03)*	.07 (.03)*
Reading level	–.01 (.04)		
School characteristics			
Private school	.09 (.18)		
Low income	.01 (.02)		
Previous reflection note (–1)			
Reflective problem exploration (–1)		.00 (.06)	
Teaching adaptation (–1)		.15 (.05)**	.18 (.05)***
Earlier reflection note (–2)			
Reflective problem exploration (–2)			.14 (.05)**
Teaching adaptation (–2)			.15 (.05)**
Variance at each level			
Teacher (21%)	.22	.31	.31
Note (79%)	.06	.05	.18
Total variance explained	.09	.11	.20

Note. Each regression model included a constant term.
*$p < .05$. **$p < .01$. ***$p < .001$.

medium-term effects (in the second study). Rude disagreements affect the likelihood of a new idea for two conversation turns. In contrast, an extensive summary pivotal post often ignites a new discussion phase and elevates the online discussion during the new phase.

The study of teachers' reflections on their practices shows that SDA can also be applied to individual changes over time. The results showed evidence of two types of sequences of teacher behaviors. Teachers often engaged in *reflective problem exploration*, which ignited a *teaching adaptation*, which in turn led to a *problem resolution*. In addition, a *teaching adaptation* was often followed by another *teaching adaptation* that led to a *problem resolution*.

Limitations

SDA requires a minimum sample size and is especially suitable for analyses of processes. Like all regressions, SDA requires a minimum sample size. 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:

$$N > 8 \text{ x } (1 - R^2)/R^2 + M - 1 \qquad (3)$$

For a large model of 25 explanatory variables with a small expected R^2 value of .10, the required sample size is 96 conversation turns (96 = 8 x [1 – 0.10]/0.10 + 25 – 1). Fewer data are needed for a larger expected R^2 value or for smaller models. (In practice, two groups of students talking for half an hour will often yield more than 100 conversation turns.)

SDA is particularly useful for examining process variables across time. For individual outcome variables across time, less complex statistical models will typically suffice, such as multilevel cross-classifications (Goldstein, 1995) or multilevel growth models (Muthén & Asparouhov, 2011). For example, if a researcher wants to test whether students who express more justifications during group discussions tend to have higher mathematics grades over time, a multilevel cross-classification is sufficient.

Future Research

Future challenges for statistical analyses of processes include how to code large data sets and how to model social networks across time. Participant coding and computer coding can facilitate coding of large data sets. In specially designed online environments (e.g., Fujita, 2009), participants use online cues such as "my theory is…" or "different opinion…," effectively providing codes for analysts to use. Another possibility is computer coding based on either a set of fixed decision rules or human codes for similar data (Erkens & Janssen, 2008). Whether participant coding or computer coding is comparable to external human coding remains an open research area. (As sampling a portion of the data omits substantial data, it can bias the results [as noted earlier in the discussion of missing data]; hence, sampling is not a viable option.)

Another potential area of study is the nature of the participants' social network within each time period. Although SDA's breakpoint analysis component identifies pivotal conversation turns or posts, integration of SDA with social network analysis can help model (a) how the social network of participants influences the likelihood of participant actions, and conversely, (b) how sequences of conversation turns or posts modify the participants' social network.

Future research in these areas can enhance the power of SDA to test more hypotheses. Their results can then help teachers, students, and classmates improve the quality of learning interactions.

Acknowledgment

I appreciate the research assistance of Choi Yik Ting.

Note

1. This is estimated by computing the unions of the probabilities for the mean number of conversation turns per student (39), $33\% = 1 - (100\% - 1\%)^{39}$, 39 turns = 3,296 total turns/84 total students.

References

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

Berger, J., & Fisek, H. M. (2006). Diffuse status characteristics and the spread of status value: A formal theory. *American Journal of Sociology*, *111*, 1038–1079.

Brown, P., & Levinson, S. C. (1987). *Politeness* (Vol. 4). New York: Cambridge University Press.

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

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

Chiu, M. M. (2001). Analyzing group work processes. In F. Columbus (Ed.), *Advances in psychology research* (Vol. 4, pp. 193–222). Huntington, NY: Nova Science.

Chiu, M. M. (2008). Flowing toward correct contributions during groups' mathematics problem solving. *Journal of the Learning Sciences*, *17*, 415–463.

Chiu, M. M., & Khoo, L. (2003). Rudeness and status effects during group problem solving: Do they bias evaluations and reduce the likelihood of correct solutions? *Journal of Educational Psychology*, *95*, 506–523.

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

Chiu, M. M., & Kuo, S. W. (2009). From metacognition to social metacognition: Similarities, differences, and learning. *Journal of Education Research*, *3*, 1–19.

Chiu, M. M., & Pawlikowski, M. J. (in press). Social metacognition and micro-creativity. In E. G. Carayannis (Ed.), *Encyclopedia of creativity, invention, innovation, and entrepreneurship*. New York: Springer.

Erkens, G., & Janssen, J. (2008). Automatic coding of dialogue acts in collaboration protocols. *Computer-Supported Collaborative Learning*, *3*, 447–470.

Fujita, N. (2009). *Group processes supporting the development of progressive discourse in online graduate courses*. Unpublished doctoral dissertation, University of Toronto, Toronto, Canada.

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

Gorman, J. C., Amazeen, P. G., & Cooke, N. J. (2010). Team coordination dynamics. *Nonlinear Dynamics Psychology and Life Sciences*, *14*, 265–289.

Gottman, J. M., & Roy, A. K. (1990). *Sequential analysis: A guide for behavioral researchers*. New York: Cambridge University Press.

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

Gunawardena, C. N., Lowe, C. A., & Anderson, T. (1997). Analysis of a global online debate and the development of an interaction analysis model for examining social construction of knowledge in computer conferencing. *Journal of Educational Computing Research*, *17*, 397–431.

Hacker, D. J., & Bol, L. (2004). Metacognitive theory. In D. M. McInerney & S. Van Etten (Eds.), *Big theories revisited* (Vol. 4, pp. 275–297). Greenwich, CT: Information Age.

Hayden, E., & Chiu, M. M. (2011, April). *Reflective teaching via a problem exploration-teaching adaptations-resolution cycle*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Huedo-Medina, T. B., Sanchez-Meca, J., Marin-Martinez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I^2 index? *Psychological Methods*, *11*, 193–206.

Kennedy, P. (2008). *Guide to econometrics*. New York: Wiley-Blackwell.

King, G., & Zeng, L. (2001). Logistic regression in rare events data. *Political Analysis*, *9*, 137–163.

Konstantopoulos, S. (2008). The power of the test in three-level cluster randomized designs. *Journal of Research on Educational Effectiveness*, *1*, 66–88.

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

MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, *39*, 99–128.

Mercer, N. (2008). The seeds of time: Why classroom dialogue needs a temporal analysis. *Journal of the Learning Sciences*, *17*, 33–59.

Muthén, B., & Asparouhov, T. (2011). Beyond multilevel regression modeling: Multilevel analysis in a general latent variable framework. In J. Hox & J.K. Roberts (Eds.), *Handbook of advanced multilevel analysis* (pp. 15–40). New York: Taylor & Francis.

Parks, K. A., & Fals-Stewart, W. (2004). The temporal relationship between college women's alcohol consumption and victimization experiences. *Alcoholism: Clinical and Experimental Research*, *28*, 625–629.

Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research. *Review of Educational Research*, *74*, 525–556.

Pevalin, D. J., & Ermisch, J. (2004). Cohabiting unions, repartnering and mental health. *Psychological Medicine*, *34*, 1553–1559.

Reimann, P. (2009). Time is precious: Variable- and event-centered approaches to process analysis in CSCL research. *International Journal of Computer Supported Collaborative Learning*, *4*, 239–257.

Stachowski, A. A., Kaplan, S. A., & Waller, M. J. (2009). The benefits of flexible team interaction during crises. *Journal of Applied Psychology*, *94*, 1536–1543.

Wise, A., & Chiu, M. M. (2011). Analyzing temporal patterns of knowledge construction in a role-based online discussion. *International Journal of Computer-Supported Collaborative Learning*, *6*, 445–470.