

# Statistically Modelling Effects of Dynamic Processes on Outcomes: An Example of Discourse Sequences and Group Solutions

Ming Ming Chiu<sup>1\*</sup>

## Abstract

Learning analysts often consider whether learning processes across time are related 1) to one another or 2) to learning outcomes at higher levels. For example, are a group's temporal sequences of talk (e.g., correct evaluation → correct, new idea) during its problem solving related to its group solution? I show how to address these issues with 1) a higher-level outcome regression and 2) a lower-level process regression, applying both to 3,234 turns of talk by 80 students working in 20 groups to solve an algebra problem. The easy-to-use, outcome-level analysis of group solution score has the following problems: multicollinearity, possibly low statistical power, cannot test for links among sequence components, and cannot model outcomes at multiple levels. The complex, process-level analysis for turns of talk overcomes these shortcomings with multilevel analysis, vector auto-regression, and outcome-level regression residuals. These results suggest a combined procedure. First, run an outcome-level analysis. If the results are significant, then the outcome-level results suffice. Otherwise, non-significant results might reflect multicollinearity, which then requires a process-level analysis. This procedure can help test a comprehensive model of how learning processes or their temporal sequences are related to learning outcomes at the turn-, time period-, individual-, group-, class-, and school-levels.

## Notes for Practice

- Researchers have not determined how to analyze whether student learning processes (e.g., creating an idea or evaluating it) often occur in specific sequences and whether they are linked to learning outcomes at higher level(s) (e.g., individual test score, group final project, etc.).
- After introducing, illustrating, and contrasting two methods to address the above issues on a dataset, I created a statistical procedure using both methods: higher-level outcome analysis followed by lower-level process analysis (if needed).
- This statistical procedure can test whether learning processes form a recurrent sequence (e.g., correct evaluation → correct, new idea).
- This statistical procedure can test whether learning processes or sequences of them are linked to learning outcomes at higher level(s) (e.g., groups with more correct evaluations; correct, new ideas; or correct evaluation → correct, new idea sequences have higher solution scores).
- These results can help teachers assess student learning processes and sequences of them, and then intervene suitably to improve their learning outcomes.

## Keywords

Time, multilevel modelling, hierarchical linear modelling, mathematical proof, sequential analysis.

**Submitted:** 09/08/16 — **Accepted:** 01/11/18 — **Published:** 04/09/18

<sup>1</sup>Email: [mingmingchiu@gmail.com](mailto:mingmingchiu@gmail.com) Address: Department of Special Education and Counseling, The Education University of Hong Kong, Faculty of Education and Human Development, Building D2, 2nd Floor, Room 15, Hong Kong, 10 Lo Ping Rd, Ting Kok, Hong Kong

\*Corresponding author

## 1. Introduction

In the field of learning analytics, educators often collect substantial data on lower-level learning processes across time and are interested in whether these temporal learning processes are 1) related to one another or 2) related to later learning outcomes at higher levels (e.g., Molenaar & Chiu, 2014). For example, do students who pass the same courses but in different orders/sequences (e.g., Trigonometry → Physics vs. Physics → Trigonometry) have different likelihoods of graduating? Do students who read suggested online readings out-of-order rather than following the list order have higher

final essay grades? Do children reading e-books who jump back to earlier pages more often than other children have higher story re-telling scores? During group problem solving, do groups with more turn-of-talk sequences of correct evaluation → correct, new idea have superior subsequent group solutions?

These examples each involve two aspects of time. First, the temporal sequence of the earlier action and the later action(s) embodies a relationship across time. Examples include the following:

1. Take Trigonometry now followed by Physics next semester
2. Read online message 2 then online message 4
3. Read e-book page 5 then page 3
4. A correct evaluation in a turn of talk followed by a correct, new idea in the next turn of talk.

Statistical analyses can test which learning processes tend to precede one or more target learning processes.

Second, specific sequences of learning processes might be linked to later higher-level outcomes. For example, sequences of courses could affect likelihood of graduation. Sequences of messages might influence the final paper grade. Sequences of pages could be linked to story re-telling. Sequences of turns of talk might affect the group solution. Likewise, statistical analyses can determine whether these links between temporal sequences and outcomes are significant.

One common way to analyze relations between variables is via regressions. For a simpler problem of determining how a higher-level fixed variable, such as gender, influences a lower-level process, such as talk, a well-established statistical method is multilevel analysis (Goldstein, 2011), also known as hierarchical linear modelling (Bryk & Raudenbush, 2002). For example, does the gender of a student affect the likelihood of a new idea in each turn of talk? To model time, time-series analysis is integrated into the multilevel analysis (e.g., statistical discourse analysis, Chiu & Fujita, 2014; Chiu & Lehmann-Willenbrock, 2016).

In contrast, researchers have not determined the optimal approach for addressing the reverse set of questions, whether a lower-level explanatory variable, let alone a temporal sequence, is related to a subsequent, higher-level outcome variable (Molenaar & Chiu, 2014), as shown in the examples in the first paragraph. A straight-forward approach aggregates the lower-level variable but loses a lot of information. A more complex approach capitalizes on the non-directionality of mathematics and inverts the analysis by switching the independent and dependent variables' positions. This paper explores the trade-offs between these two regression approaches and shows how to combine them in the context of a specific example of lower-level processes (turns-of-talk) and a higher-level outcome (group solution score), which occur in many learning analytics datasets as shown by the diverse examples above. These analyses can help test a theory of learning processes with outcomes at different levels and integrate their results to create more comprehensive theories of learning encompassing multiple levels and time scales.

To explore these issues, I test whether the temporal sequence of lower-level processes, correct evaluation → correct, new idea, is linked to each group's solution (subsequent higher-level outcome variable) by comparing two statistical approaches at different units of analysis, higher-level outcome versus lower-level process. To showcase these two analyses, I apply each one to test a pair of hypotheses using a dataset representative of many existing datasets collected by education researchers. The dataset has 3,234 turns of talk by 80 students working in 20 groups of 4 to solve an algebra problem. After illustrating the results, I mathematically prove some key relations. Next, I contrast the advantages and disadvantages of each approach, and recommend a procedure involving both.

## 2. Hypotheses

For problems that individual students cannot solve alone, working together might yield a solution (zone of proximal development tasks, Vygotsky, 2011). During group problem solving, students try to develop and put together new ideas to form a correct solution (Chiu, 2008). According to functional group decision-making theory (Orlitzky & Hirokawa, 2001), correct, new ideas (micro-creativity) are the building blocks of successful group solutions for open, ill-defined problems. Furthermore, case studies of group problem solving suggest that correct evaluations raise the likelihood of correct, new ideas (Barron, 2003). In many contexts such as algebra, an evaluation can be objectively classified as correct or incorrect ("you added wrong"; Chiu, 2008). Correct evaluations support correct ideas ("two times four is eight, right?") or reject flawed ideas ("no, two times four is not six"), thereby creating a foundation of partially shared understandings of correct ideas that group members can use to create a correct, new idea. This results in the temporal sequence of correct evaluation → correct, new idea. In contrast, incorrect evaluations reject correct ideas ("nope, two times three isn't six") or accept flawed ones ("two times three is five, yeah"), embedding errors in their partially shared understandings and fostering flawed ideas. Hence, statistical tests might provide evidence of the greater generality of this temporal, sequential relation between correct evaluation and correct, new idea, beyond past case studies (Barron, 2003).

More importantly, this correct evaluation → correct, new idea temporal sequence might be linked to a correct group solution, even after accounting for the effects of its components (Molenaar & Chiu, 2017). As many students have poor metacognitive skills, they often have difficulty evaluating the quality of their individual problem solving (Chiu & Klassen, 2010), so a potential benefit of working together during group problem solving is improved evaluations of one another's work, a component of social metacognition theory (Chiu & Kuo, 2009). Whereas individual metacognition is monitoring and controlling one's own knowledge, emotions, and actions (Hacker & Bol, 2004), social metacognition is group members' monitoring and control of one another's knowledge, emotions, and actions (Chen & Chiu, 2008).

Whereas an isolated action can be ignored by groupmates and thereby have limited value, a sequence suggests that a group member operated on another's reasoning (transactivity; in this case, the correct evaluation facilitated a correct, new idea), which tends to generate more mutual complex understanding and to do so more quickly (Teasley, 1997). Hence, a sequence might be more likely than its isolated components to form the basis for solving a problem (Weinberger & Fischer, 2006). Specifically, the temporal sequence correct evaluation → correct, new idea might be linked to solution score. If so, it can help teachers evaluate student discussions and intervene accordingly.

Hypothesis H-1. Groups with proportionally more correct evaluation → correct, new idea temporal sequences produce superior solutions compared to other groups.

To help clarify analyses of a possible link between a lower-level process and a higher-level outcome as in hypothesis H-1, I also test a contrasting link between two variables at the same group-level in hypothesis H-2 below. Specifically, past studies have shown that individual students or groups of students with high past academic achievement often outperform their lower-achieving counterparts on academic tasks (e.g., Chiu & Chow, 2015).

Hypothesis H-2. Groups with higher mean mathematics grades in the last semester are more likely than groups with lower ones to produce superior solutions.

I test these two hypotheses with the following data.

### 3. Data

Eighty 9th grade students (40 girls, 40 boys) were videotaped during their algebra classes in an urban high school in the United States. This school's mathematics test scores were at the 40th percentile in the state (maximum = 100; California Department of Education, 2005). Their races were 12 Asian, 27 Black, 28 Hispanic, and 13 Caucasian.

These students worked in groups of four. All groups had at least one boy and one girl. Also, all groups had members from different races. These students attended the same algebra class for seven months, had not received any group work training, and had not previously worked together in groups. A student's mathematics grade in the last semester indicated past achievement; a group's mean mathematics grade is the mean of each member's mathematics grade.

During the first lesson of a new unit on two-variable algebraic equations, groups of students worked on the following problem for 30 minutes and altogether produced a total of 3,234 conversation turns.

Under the Universal plan, each text message costs \$.10. Budget costs \$.01 per text message, but charges a monthly fee of \$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?

As our study focuses on only the third part of this problem, the hypotheses likewise refer only to the third part. To solve the third part of this problem, one can equate the total costs for each company, ( $0.10 \times \text{texts} = 0.01 \times \text{texts} + 18$ ), to obtain 200 texts ( $= 18 / [0.10 \times \text{texts} - 0.01 \times \text{texts}]$ ). In addition to a correct solution, credit was also given to partial solutions. Hence, possible solution scores were: (a) correct: 3 points, (b) correct method but incorrect answer (e.g., arithmetic error): 2 points, (c) incorrect method but understands the problem situation: 1 point, and (d) does not understand the problem situation: 0 points.

Group-level Variable	Mean	SD	Min	Max	$\alpha^a$	Description	Example
<i>Group solution score</i>	1.90	1.25	0.00	3.00	N/A <sup>b</sup>	Score of group’s final solution. <i>Correct</i> = 3; <i>Correct method but wrong answer</i> (e.g., arithmetic error) = 2; <i>Wrong method but understands the problem situation</i> = 1; <i>Does not understand the problem situation</i> = 0.	<i>Correct</i> : 200 texts; <i>Wrong answer</i> : 20 texts = 18/.09; <i>Wrong method</i> : 180 = 18/.10; <i>Does not understand problem</i> : Universal always costs more (.10 > .01).
<i>Group mean math grade</i>	83.95	7.27	70.50	94.00	N/A	Mean of all students’ last semester’s mathematics grades within a group.	80 = (100 + 90 + 70 + 60)/4
<i>% Correctly evaluate (-1) → Correct, new idea</i>	0.15	0.13	0.01	0.41	<sup>a</sup>	The proportion of temporal sequences in which a <i>correctly evaluation in a previous turn of talk</i> is followed by <i>correct, new idea in the current turn of talk</i> over all sequences. (See correct evaluation and correct, new idea below)	Eighteen divided by point o nine isn’t two. → Oh, it’s two hundred.
<b>Turn-level Variable</b>							
<i>Correctly evaluate</i>	0.30	0.46	0.00	1.00	.93	Agreeing with the previous speaker’s correct idea or disagreeing with the previous speaker’s incorrect idea	Eighteen divided by point o nine isn’t two.
<i>Correct, new idea</i>	0.20	0.40	0.00	1.00	.98	A mathematically correct idea consistent with the problem situation that has not been mentioned earlier during the group problem solving session or in the problem situation	Oh, it’s two hundred.

<sup>a</sup> The reliability of this variable is computed from its component turn-level variables (correct, new idea and correctly evaluate) whose Krippendorff’s (2012)  $\alpha$  reliabilities are .98 and .93, respectively.

<sup>b</sup> Both coders agreed on all solution scores for the 20 groups.

Each group was videotaped. Two research assistants who did not know the research hypotheses coded each turn of talk for (a) validity (correct vs. incorrect vs. null), (b) novelty (new vs. old idea vs. null), and (c) evaluation (agree vs. disagree vs. ignore; for coding details, see Chiu, 2008). Null indicates that the turn of talk did not include content to evaluate as correct or incorrect, such as an off-task comment about a recent movie. A turn of talk was coded as correct only if it was both mathematically correct and consistent with the above problem situation. For example,  $20 \times 20 = 400$  is mathematically correct but is not related to the above problem. A new idea is an idea that was not previously mentioned in the group conversation. As noted above, a correct, new idea is an idea that is both correct and new, and a correct evaluation is either agreeing with a correct idea or disagreeing with an incorrect idea.

#### 4. Outcome-Level (Higher) vs. Process-Level (Lower) Analyses

To test whether sequences of lower-level processes are related to higher-level outcomes, two approaches at different units of analysis are used: higher-level outcome versus lower-level process; in this study, they are group versus turn-of-talk. A regression at the group-level unit of analysis uses the group solution score as the dependent variable. As statistical regressions do not assume directionality, an alternate regression at the turn-level unit of analysis uses the group solution score as an independent variable.

##### 4.1. Outcome-Level Specification

In this outcome-level analysis, group solution score is the dependent variable, and group-level independent variables are created from turn-level processes. In this case, group *solution* is the dependent variable. As *solution* is an ordered variable (0, 1, 2, 3) rather than a continuous variable (e.g., height), an ordinary least squares regression would yield biased standard errors; hence, an ordered Logit or ordered Probit regression is needed for unbiased results (Kennedy, 2008). Ordered Logit and ordered Probit have different distribution assumptions but often yield similar results in practice (Kennedy, 2008).

Group-level independent variables such as *group's mean mathematics grade* are entered as usual in the group-level regression, but turn-level process variables are not allowed. Hence, to enable testing of hypothesis H-1, aggregate each turn-level process variable into a group-level, proportion variable, by computing the proportion of each turn-level process (e.g., *correctly evaluate* → *correct, new idea* sequence) in each group *g*. For each group's data, we compute the following proportion:

$$\frac{\text{Total instances of } \textit{correctly evaluate} \rightarrow \textit{correct, new idea} \text{ sequence in group } g}{\text{Total 2-turn sequences in group } g}$$

For example, group 4 has 201 turns of talk and 30 *correctly evaluate* → *correct, new idea* sequences. For *T* turns of talk, there are  $T - L + 1$  temporal sequences of length *L*, so group 4 has 200 sequences with a length of 2 turns ( $200 = 201 - 2 + 1$ ), and its proportion of *correctly evaluate* → *correct, new idea* sequences is 15% ( $= 30 / 200$ ). This proportion is computed for each remaining group.

In addition to the sequence *correctly evaluate* → *correct, new idea*, a group's solution score might also be related to either component of the sequence: *correctly evaluate* or *correct, new idea*. Hence, we compute the proportion of *correctly evaluate* turns in each group and the proportion of *correct, new idea* turns in each group.

Note that the % *correctly evaluate* → *correct, new idea* is correlated with both % *correctly evaluate* and % *correct, new idea*. A turn of talk that does not have a *correct evaluation* cannot be part of the *correctly evaluate* → *correct, new idea* sequence. Likewise, a turn of talk that does not have a *correct, new idea* also cannot be part of this sequence. Hence, groups with a quantity of *correctly evaluate* → *correct, new idea* sequences must have at least the corresponding quantities of *correctly evaluate* and *correct, new idea*. For example, a group with 42 sequences of *correctly evaluate* → *correct, new idea* must have at least 42 instances of *correctly evaluate* and at least 42 instances of *correct, new idea*. For a mathematical proof showing that a composite variable and any of its component variables are always positively correlated, see Appendix B, proposition B1. When the correlations between a temporal sequence and its components are large, adding them as explanatory variables in a regression can yield *multicollinearity*, which negatively biases the results (Yoo et al., 2014).

To test hypothesis H-2, enter the group-level outcome and explanatory variables into an ordered Logit regression (*solution* and *mean math grade*).

$$\textit{Solution}_g = \beta_0 + \beta_1 \textit{Mean\_math\_grade}_g + e_g \tag{1}$$

The *Solution* score of group  $g$  has intercept  $\beta_0$ , has an unexplained component (*residual*)  $e_g$ , and might be related to the group's *Mean\_math\_grade*.

To test hypothesis H-1, add the variable, *%\_Correctly\_evaluate →\_Correct, new idea*.

$$Solution_g = \beta_0 + \beta_1 Mean\_math\_grade_g + \beta_2 \%\_Correctly\_evaluate \rightarrow\_Correct, new\ idea_g + e_g \quad (2)$$

To control for each component of the *correctly evaluate → correct, new idea* sequence, the percentages of *correctly evaluate* turns and *correct, new idea* turns are entered.

$$Solution_g = \beta_0 + \beta_1 Mean\_math\_grade_g + \beta_2 \%\_Correctly\_evaluate \rightarrow\_Correct, new\ idea_g + e_g + \beta_3 \%\_Correctly\_evaluate + \beta_4 \%\_Correct, new\ idea_g + e_g \quad (3)$$

(Note that *%\_Correctly\_evaluate →\_Correct, new idea* is an aggregated proportion, not an interaction between *%\_Correctly\_evaluate* and *%\_Correct, new idea*. Hence, the latter two variables need not be entered into a regression before entering *%\_Correctly\_evaluate →\_Correct, new idea*.)

As testing multiple hypotheses increases the likelihood of a false positive, the false discovery rate (FDR) is reduced via the *two-stage linear step-up procedure*, which outperformed 13 other methods in computer simulations (Benjamini, Krieger & Yekutieli, 2006).

In short, this model tests whether a group-level outcome, *solution*, is related to a group-level covariate *Mean\_math\_grade*, a sequence *%\_Correctly\_evaluate →\_Correct, new idea*, or its components *%\_Correctly\_evaluate* or *%\_Correct, new idea*.

While this example has only one outcome (group solution), multiple outcomes can also be modelled, such as group solution and individual test score. If all outcomes are continuous variables, then a *system of equations* is suitable (e.g., *Zellner's method*; Kennedy, 2008). If the outcomes are all discrete (binary or ordered), then a *multilevel, multivariate outcome (ordered) Logit/Probit analysis* is appropriate (Goldstein, 2011). Lastly, if modelling both continuous and discrete outcomes, a *multilevel, mixed response* model is suitable (Goldstein, 2011).

#### 4.2. Outcome-Level Results

Analyses with *MLwiN* software (Rasbash, Steele, Browne, & Goldstein, 2015) yielded the results in Table 2. They show that groups with higher *mean\_math\_grades* tend to have significantly higher solution scores in all models (1–4), supporting hypothesis H-2.

Meanwhile, the results for *%\_correctly\_evaluate →\_correct, new idea* and hypothesis H-1 are less clear; it is significant in model 2 but not in model 3 when controlling for its components, *%\_correctly\_evaluate* and *%\_correct, new idea*. Group solution score is highly correlated with both *%\_correctly\_evaluate* ( $r = .70$ ) and *%\_correct, new idea* ( $r = .64$ ), so the significant *%\_correctly\_evaluate →\_correct, new idea* result might be a spurious correlation (see Table 3). To test this possibility, I determine whether the two components *%\_correctly\_evaluate* and *%\_correct, new idea* account for most of the variance in *solution* via another analysis with all the variables except *%\_correctly\_evaluate →\_correct, new idea*. The turn-level independent variables are not significant, and the explained variance is substantially less than model 3 with all the independent variables, which suggests that *%\_correctly\_evaluate →\_correct, new idea* accounts for substantial variance (at least  $0.13 = 0.62 - 0.49$ ).

Hence the results are not conclusive regarding hypothesis H-1. *%\_correctly\_evaluate →\_correct, new idea* might be significantly related to group solution score and account for substantial variance, but its results are rendered non-significant due to extremely high correlations with *%\_correctly\_evaluate* ( $r = .96$ ) and *%\_correct, new idea* ( $r = .94$ ; multicollinearity, Yoo et al., 2014). Analyses with ordered Probit show similar results (see Appendix A).

#### 4.3. Advantages and Disadvantages of Outcome-Level Analysis

This example shows an outcome-level analysis's advantage of simplicity and its disadvantages of multicollinearity and small sample size. Simple, arithmetic conversion of turn-level process variables into group-level proportion variables facilitate their direct entry into the statistical model of group solution score. As noted above, any temporal sequence variable of the form *% A → B in an outcome-level unit* is correlated with its components, *% A in an outcome-level unit* and *% B in an outcome-level unit* (see mathematical proof in Appendix B, proposition B1). For longer temporal sequences (e.g., *% A → B → C in an outcome-level unit*), their correlations with their components (*% A [or B or C] in an outcome-level unit*) might be smaller but remain substantial. This multicollinearity stems from using both a composite and its component(s)

simultaneously as explanatory variables and can occur in single-level contexts, as well as multilevel contexts (Yoo et al., 2014).

**Table 2.** Summary of 4 ordered logit regression models of group solution score showing unstandardized regression coefficients and standard errors in parentheses (N = 20 groups)

Independent variable	Group solution score			
	Model 1	Model 2	Model 3	Model 4
<i>Group mean mathematics grade</i>	0.15** (0.05)	0.12* (0.05)	0.08* (0.04)	0.07* (0.04)
<i>% Correctly evaluate → Correct, new idea</i>		11.39* (5.62)	7.31 (6.11)	
<i>% Correctly evaluate</i>			6.78 (3.98)	4.02 (3.29)
<i>% Correct, new idea</i>			1.05 (3.97)	-1.33 (3.48)
<i>McFadden's R<sup>2</sup></i>	.25	.35	.62	.49
<i>Cohen's f<sup>2</sup></i>	0.33	0.54	1.63	0.96

\*  $p < .05$ ; \*\*  $p < .01$

**Table 3.** Correlation, variances and covariances of variables in the lower left triangle, diagonal (in bold), and upper right triangle of the matrix below

Variable	1	2	3	4	5
1 <i>Group solution score</i>	<b>1.49</b>	5.96	.10	.16	.13
2 <i>Group mean mathematics grade</i>	.69	<b>5.18</b>	.59	.84	.74
3 <i>% Correctly evaluate → Correct, new idea</i>	.62	.65	<b>.02</b>	.02	.02
4 <i>% Correctly evaluate</i>	.70	.65	.96	<b>.03</b>	.03
5 <i>% Correct, new idea</i>	.64	.63	.95	.94	<b>.03</b>

Also, the small sample size of the group-level analysis yields low statistical power, which raises the likelihood of false negatives. When a dataset allows multiple levels of analysis, each level has its own sample size. As there are 20 groups and 3,234 turns of talk in this dataset, the group-level sample size is 20, and the turn-level sample size is 3,234. As the information from the 3,234 turns of talk are aggregated into group-level variables, all variables in the analysis are group-level variables, so the sample size is 20. During this aggregation process, much of the information about each turn of talk is not used, so the statistical power falls. In this example, the sample size of 20 groups yields a statistical power of only 0.25 for detecting a medium effect size of 0.30 (Cohen, West, Aiken, & Cohen, 2003). The regression coefficient of an explanatory variable is its effect size (Kennedy, 2008). For an 80% likelihood of successfully detecting this medium effect size, 84 groups are needed, and for a 90% likelihood, 112 groups are needed. In analyses of small samples, statistical power is low, so significant results for smaller relations are not expected; hence, any significant result is statistically meaningful, while non-significant results might be false negatives (Kennedy, 2008).

#### 4.4. Process-Level Specification

The process-level analysis uses the turn of talk as the unit of analysis and models the outcome group solution score as an independent variable. Whereas the group-level specification asks, “whether the proportion of a sequence *% Correctly evaluate → Correct, new idea* is related to the group solution score,” the turn-level specification asks, “when groups with higher solution scores use a correct evaluation, whether a correct, new idea is more likely to follow than otherwise.” (The group solution occurs after the talk sequences, and time cannot flow backwards, so the group solution cannot influence the talk sequences.) More generally, a regression does not mathematically dictate the direction of causality, so an explanatory factor can be a dependent variable and an outcome can be an independent variable.

First, suitable variables are created from the turns of talk. See example in Table 4.

**Table 4.** Correct, new idea; correctly evaluate; and its lagged variable, correctly evaluate (-1), which is created from the shifted values (bolded)

Turn	Person	Talk	Correct, New idea	Correctly Evaluate	Correctly Evaluate (-1)	...
1	Ana:	I send 100 texts, so 100 times ten cents	1	N/A	N/A	
2	Ben:	Yeah, ok, 100 times ten is a thousand dollars	0	<b>1</b>	N/A	
3	Eva:	A thousand dollars [writes \$1000]	0	<b>0</b>	1	
4	Ana:	No, it's a thousand cents	1	1	0	
...						...

Ana expresses a correct, new idea (“100 times ten cents”), so the *correct, new idea* variable has a value of 1 (true). As there is no turn of talk to evaluate before the first turn, “N/A” is placed into *correctly evaluate*. Ben agrees with Ana’s correct idea (“Yeah, ok”; *correctly evaluate* = 1) but creates a wrong idea (“a thousand dollars”; *correct, new idea* = 0 [false]). Then, Eva repeats and writes down Ben’s wrong answer, indicating that she incorrectly agrees with it (*correctly evaluate* = 0) and shows no *correct, new idea* (= 0). Ana correctly disagrees with Eva’s wrong idea (“No”; *correctly evaluate* = 1), and creates another correct, new idea (“a thousand cents”; *correct, new idea* = 1).

The value of *correctly evaluate* in the previous turn is its *lag value*, and the corresponding *lag variable* is denoted *correctly evaluate (-1)* to show that it occurred in the previous turn of talk. As *correctly evaluate (-1)* is the previous turn’s value of *correctly evaluate*, we copy the latter’s values from the previous row (previous turn of talk). Specifically, we shift the values of the original variable (N/A, 1, and 0 of *correctly evaluate*) down one row, copy them and paste them into *correctly evaluate (-1)*. As the first turn has no previous turn, its *correctly evaluate (-1)* value is N/A.

In this turn-level analysis, the current action in the process is the dependent variable in the regression. The value of a prior action, indicated with a (-1) sign, is an explanatory variable. In these data, the dependent variable is the dichotomous variable *correct, new idea*, and *correctly evaluate (-1)* is an explanatory variable.

As an ordinary least squares regression on a dichotomous, dependent variable would yield biased standard errors, a binary Logit or binary Probit regression is needed for unbiased results (Kennedy, 2008). As these nested data have two levels (1: turn of talk; 2: group), a single-level analysis can yield biased standard errors, so a multilevel analysis is needed, specifically multilevel binary Logit/Probit (Goldstein, 2011).

$$Correct, new idea_{gt} = \beta_{00} + f_g + e_{gt} \tag{4}$$

Whether there is a *correct, new idea* in turn of talk *t* in group *g* depends on the grand mean  $\beta_{00}$  and the *residuals* at the group- and turn-of-talk-levels,  $f_g$  and  $e_{gt}$ .

Unlike the group-level analysis, the turn-level analysis tests whether the components of a temporal sequence are linked; specifically, after a correct evaluation (-1), is a correct, new idea more likely to occur? Hence, the explanatory variable *correctly\_evaluate(-1)* is entered into the model.

$$Correct, new idea_{gt} = \beta_{00} + f_g + e_{gt} + \beta_{g2}Correctly\_evaluate(-1)_{gt} \tag{5}$$

A regression with lag variables is a *vector autoregression* (VAR; Kennedy, 2008). This integration of multilevel, binary logit and VAR is a simple version of *statistical discourse analysis* (SDA; Chiu & Fujita, 2014; Chiu & Khoo, 2005).

To test hypothesis H-1, whether groups that solve the problem correctly are more likely to use this *correctly evaluate* → *correct, new idea* sequence, we add two more independent variables, (a) *solution* and (b) the interaction *solution \* correctly\_evaluate(-1)*. If *solution\_score* has a significant, positive, regression coefficient, then groups with higher solution scores are more likely to show a *correct, new idea*. If *solution \* correctly\_evaluate(-1)* has a significant, positive, regression coefficient, then groups with higher solution scores are more likely to have *correctly evaluate* → *correct, new idea* sequences. The potential link between *solution \* correctly\_evaluate(-1)* and *correct, new idea* in this process-level analysis corresponds to the link between % *Correctly evaluate* → *Correct, new idea* and *Group solution score* in the outcome-level analysis, but they are not identical.

$$Correct, new idea_{gt} = \beta_{00} + f_g + e_{gt} + \beta_{g2}Correctly\_evaluate(-1)_{gt} + \beta_3Solution_g + \beta_{g4}Correctly\_evaluate(-1)*Solution_g \tag{6}$$

However, this model does not account for group-level explanatory variables, such as *group mean mathematics grade*, that might be related to *group solution score*, namely hypothesis H-2. Hence, I control for differences in *group mean mathematics grades* that are related to the *group solution score* as follows. First, I test whether the mean mathematics grade of a group is significantly related to its solution score using the outcome-level equation (1) above (see Table 1). If they are significantly related, I remove this overlap between mean mathematics grade and *group solution score* by storing the latter's unexplained component  $e_g$  from equation (1) into *solution\_residual*. Then, a revised version of equation (6) is used to test hypothesis H-1, in which all instances of *solution* are replaced with *solution\_residual*.

$$Correct, new\ idea_{gt} = \beta_{00} + f_g + e_{gt} + \beta_{g2}Correctly\_evaluate(-1)_{gt} + \beta_3Solution\_residual_g + \beta_{g4}Correctly\_evaluate(-1)*Solution\_residual_g \tag{7}$$

As above, the two-stage linear step-up procedure reduces the likelihood of a false positive (Benjamini et al., 2006). Also, multiple dependent variables can be modelled with systems of equations (e.g., Zellner's method; Kennedy, 2008); multilevel, multivariate outcome analysis (Goldstein, 2011); or multilevel, mixed response models (Goldstein, 2011), depending on the type(s) of dependent variables. For example, to test whether a group's *solution score* is related to *correct evaluation*, the latter can be modelled as an additional dependent variable along with *correct, new idea*.

Unlike the group-level analysis, this turn-level analysis can model outcomes at multiple levels via explanatory variables and interactions among the explanatory variables. For example, a sequence of repetitions/paraphrases of earlier statements (summary) during an online discussion can enhance two types of outcomes: (a) ignite a new time period of elevated discussion quality among group members (Wise & Chiu, 2011) or (b) improve later individual performance outcomes (Wise & Chiu, 2014).

**4.5. Process-Level Results**

The results of equations (6) and (7) show that all independent variables are significantly related to the dependent variable *correct, new idea* in all models in Table 5. After correctly evaluate (-1), a correct, new idea is more likely in the next turn of talk than otherwise. Hence, they are significantly linked to one another, and the sequence *correctly evaluate (-1) → correct, new idea* is unlikely to be random (see models 1, 2, and 3). The likelihood of *correct, new idea* in each turn of talk is greater in groups with higher solution scores than in groups with lower solution scores (model 2). Also, the interaction term *correctly evaluate (-1) \* solution* in model 2 has a significant, positive regression coefficient, indicating that in groups with higher solution scores, *correctly evaluate (-1)* is more likely to be followed by *correct, new idea* in the next turn of talk than otherwise; this result shows that the sequence *correctly evaluate → correct, new idea* is related to higher group solution scores, supporting hypothesis H-1.

**Table 5.** Summary of 3 statistical discourse analysis models of correct, new idea using multilevel Logit (N = 3,234 turns of talk)

Independent variable	Correct, new idea		
	Model 1	Model 2	Model 3
<i>Correctly evaluate (-1)</i>	0.98 (0.09)***	1.37(0.25)**	0.96(0.10)***
<i>Solution</i>		0.47(0.06)***	
<i>Correctly evaluate (-1) * Solution</i>		0.23(0.10)*	
<i>Solution residual</i>			0.40(0.08)***
<i>Correctly evaluate (-1) * Solution residual</i>			0.22(0.10)*
Variance at each level	Variance explained		
Group (1%)	.00	.24	.22
Turn of talk (99%)	.04	.06	.04
Explained variance	.04	.06	.04
<i>Cohen's f<sup>2</sup></i>	0.04	0.06	0.04

\*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$

The *solution residual* results (model 3) resemble the *solution* results (model 2). After controlling for the group-level control variable *group mean mathematics grade*, the solution residual is positively related to the likelihood of *correct, new idea*. Likewise, *correctly evaluate (-1) \* solution residual* is also positively related to the likelihood of correct, new idea. These two results show that controlling for group mean mathematics grade, groups with higher solution scores have greater likelihoods of both *correct, new idea* and sequences of *correctly evaluate → correct, new idea*. As the turns of talk occur before the group solution, these results are consistent with the claim that *correctly evaluate → correct, new idea* sequences yield higher group solution scores. If this evidence is further buttressed in future studies, it suggests that teachers consider whether such sequences occur during group problem solving when assessing the effectiveness of their group processes, and to foster such sequences if needed.

**Table 6.** Correlation, variances, and covariances of variables in the lower left triangle, diagonal (in bold), and upper right triangle of the matrix below.

Variable	1	2	3	4	5	6	7
1 <i>Correct, new idea</i>	<b>.16</b>	.04	.07	.26	.09	.03	.01
2 <i>Correctly evaluate (-1)</i>	.19	<b>.21</b>	.09	.33	.47	.04	.03
3 <i>Group solution score</i>	.17	.19	<b>1.09</b>	4.52	.47	.42	.11
4 <i>Group mean math grade</i>	.09	.11	.63	<b>47.53</b>	1.94	1.04	.42
5 <i>Correctly evaluate (-1) * Group solution score</i>	.20	.89	.39	.25	<b>1.33</b>	.18	.16
6 <i>Group solution score residual (Group mean math grade)</i>	.11	.11	.58	.21	.22	<b>.49</b>	.15
7 <i>Correctly evaluate (-1) * Group solution score residual (Group mean math grade)</i>	.06	.14	.27	.16	.35	.56	<b>.15</b>

#### 4.6. Advantages and Disadvantages of Process-Level Analysis

In contrast to the outcome-level analysis, the process-level analysis shows less multicollinearity, has more statistical power, tests sequential links among processes, and can model multiple outcomes at multiple levels; however, it is more complex and requires more statistics knowledge and skill (see Table 7). As the group-level analysis converts process-level variables into percentages, it yields correlated variables (see Appendix B, proposition B1), greater multicollinearity, and unclear results (Yoo et al., 2014). As the turn-level analysis does not create composite variables, it does not have this source of multicollinearity and its corresponding problems.

Meanwhile, the turn-level analysis uses much more of the turn-level information in the data, and its turn-level sample size is much greater than that of the group-level analysis (in this case, 3,234 turns of talk >> 20 groups), resulting in greater statistical power in the turn-level analysis (.99 >> .25 for a medium effect size of 0.30) and more confidence in non-significant results. Note that group-level results still have the same low statistical power as before. For a mathematical proof showing higher statistical power at lower levels than at higher levels, see Appendix B, proposition B2.

Unlike the higher, learning outcome-level analysis, the learning process-level analysis can test a broader variety of theoretical models involving temporal sequences of learning processes or learning outcomes at multiple levels. The learning process-level analysis tests whether the learning process components of the sequence are linked together and whether they are more likely to occur in a specific order than by chance. In this dataset, a correct evaluation in the previous turn is more likely to be followed by a correct, new idea than otherwise. As noted above, the learning process-level analysis can also model multiple learning outcomes at multiple levels (turn of talk, time period, individual, group, classroom, school, and so on) via explanatory variables and interactions among the explanatory variables (Wise & Chiu, 2011, 2014).

Lastly, the learning process-level analysis requires more knowledge and skill than the learning outcome-level analysis to implement. Specifically, the learning process-level analysis requires knowledge and use of multilevel analysis, vector auto-regression, and higher-level regression residuals. Hence, collaboration with statisticians or improved statistical training is needed for widespread use of this learning process-level analysis.

**Table 7.** Comparison of outcome- vs. process-level analyses along 5 dimensions

Dimension	Unit of analysis	
	Outcome level (higher level)	Process level (lower level)
<i>Multicollinearity</i>	More for converted process-level variables	Less
<i>Statistical power</i>	Less	Much more
<i>Test link(s) among sequence components</i>	No	Yes
<i>Model outcomes at multiple levels</i>	No	Yes
<i>Ease of analysis</i>	Simpler (arithmetic conversion of process-level variables, ordinary least squares, Logit, Probit)	More Complex (multilevel analysis, multilevel Logit/Probit, vector auto-regression, residuals as variables)

**4.7. Recommended Procedure**

Given the advantages and disadvantages of outcome- and process-level analyses, I recommend the following procedure (expressed at group- and turn-levels to aid understanding). The group-level analysis is fairly easy to implement, might suffice, and is needed for the turn-level analysis. Thus, the group-level analysis can be done first. If the converted turn-of-talk variables are significant, the group-level analysis might suffice, as low statistical power does not change the interpretation of significant results. The turn-level analysis can be run to test the robustness of the results.

If the converted turn-of-talk variables are not significant and the group-level sample size is small, multicollinearity might be the culprit. Then, the turn-level analysis is needed. As the turn-level analysis uses more of the information in the data, its results are more credible than the group-level results.

If the dependent variable is discrete rather than continuous, binary or ordered Logit is used. As Logit and Probit have different distribution assumptions, all binary or ordered Logit specifications should be run again with binary or ordered Probit, respectively. If both yield consistent results, then (a) report the one that accounts for more variance in the main table and put the other one in an appendix (or link to a website with the results), and (b) state that both analyses yield similar results. If the results differ, report both sets of results and indicate that the results differ due to the different distribution assumptions of Logit versus Probit.

In short, the recommended procedure is as follows:

- Run an outcome-level analysis.
- If converted process-level variables are significant, outcome-level results might suffice. (You may run a process-level analysis for robustness.)
- If converted process-level variables are not significant, check for multicollinearity, and run a process-level analysis.

**5. Discussion**

As learning analysts often have both substantial lower-level learning process data and subsequent learning outcome data, they can answer the following categories of questions regarding time (e.g., Molenaar & Chiu, 2014):

- Which learning processes are linked to the likelihood of a subsequent target learning process?
- Do these learning processes tend to occur in specific sequences?
- Is a learning process or a sequence of learning processes linked to learning outcomes at higher levels?
- What is the impact of multiple, temporal sequences of learning processes on learning outcomes at multiple levels?

The first two questions ask about the temporal relationships among learning processes, while the last two questions ask about the temporal relationships between learning processes and subsequent outcomes. If we can answer the first question, we can identify partial orderings of multiple learning processes. By answering the second question, we can identify sequences of learning processes that serve as candidate learning mechanisms. Answering the third question provides evidence that a sequence of learning processes is related to learning outcomes. Answering the last question can provide evidence that multiple sequences of learning processes fit together to yield learning outcomes at multiple levels and can support a comprehensive theory of learning processes and outcomes.

To address these questions, I compared learning outcome-level and learning process-level statistical analyses by applying them to test a pair of hypotheses on a single dataset, and then combined them into a single procedure. Unlike past methods, this statistical procedure can simultaneously address all of the above questions in a single model. Rather than assuming that a

temporal sequence of learning processes is coherent and has impact, this statistical procedure tests whether they differ significantly from a random concatenation of learning processes in these respects. In this study, *correct evaluation* was significantly more likely than other processes to precede a *correct, new idea*. Moreover, this statistical procedure tests whether one or more temporal sequences of learning processes are related to a later learning outcome at a higher level. In this study, both *correct, new idea* and the *correct evaluation* → *correct, new idea* sequence were linked to *group solution score*.

Whereas past statistical methods test student learning outcomes at only one level, this statistical tool can test a comprehensive, theoretical model of how temporal sequences of learning processes are related to learning outcomes at multiple levels and time scales. For example, this procedure can test whether temporal sequences of turns of talk among group members are related to their discussion quality in a later time period (Wise & Chiu, 2011), individual student's subsequent performance (Wise & Chiu, 2014), the group's subsequent performance (e.g., group presentation), the class's unit test scores, and so on. Such analyses can help test single-level theories of learning, integrate their results to create more comprehensive theories of learning encompassing multiple levels and time scales, and then test these multilevel theories of learning across time within a single statistical model. Hence, this statistical tool can accelerate the integration and testing of comprehensive, multilevel theories of learning processes and outcomes across time.

### 5.1. Limitations

The outcome-level and process-level analyses typically yield different results, and various data attributes can influence the results of the above statistical procedure. While the group-level and turn-level analyses address similar issues, they are neither identical nor symmetric. Across analyses, the key dependent and independent variables are switched, so the residuals have different values and meanings. As a result, the regression coefficients and standard errors also differ. Future studies can employ Monte Carlo methods to test the degree to which these two methods yield similar or different results and the conditions in which the results differ substantially.

Inadequate sample size, non-significant variance, or non-normal distributions can influence the above statistical procedure. For small sample sizes, statistical power is low, which reduces the likelihood of detecting a significant effect. Computer simulations suggest that stable multilevel analysis results at the highest level often require at least 50 units at that level (Maas & Hox, 2005). As this dataset only has 20 groups at the highest level, cautious interpretation is needed for group-level results. Also, non-significant variance at a level of analysis removes the need to model that level (Chiu & Lehmann-Willenbrock, 2016). If only the lowest level of analysis shows significant variance, then a multilevel analysis is not needed (Maas & Hox, 2005). Instead, simpler analyses suffice: an ordinary least squares regression for continuous dependent variables, Logit/Probit for dichotomous ones, and ordered Logit/Probit for ordered ones (Goldstein, 2011). Lastly, variables with non-normal distributions can bias the results, which can be addressed by time models of non-normal distributions (Shadish, 2014).

## 6. Conclusion

To explore how to test whether lower-level learning processes are related to higher-level learning outcomes, I examined two statistical approaches at these two units of analysis: group level versus turn-of-talk level. Each analysis has its strengths and weaknesses; hence, I suggest a procedure that combines both analyses. The learning outcome-level analysis is simple to implement but can cause multicollinearity. Also, learning outcome-level analyses might have small samples and low statistical power. Unlike learning outcome-level analyses, learning process-level analyses do not have these two weaknesses, can test whether the components of a sequence are significantly linked, and can model learning outcomes at multiple levels. However, learning process-level analyses are more complex (requiring multilevel analysis, vector auto-regression, and higher-level regression residuals). Hence, a procedure combining both methods is suggested when applying statistical methods to model lower-level learning processes, especially sequences across time, and higher-level learning outcomes.

## Acknowledgments

In 2011, Dan Suthers raised a question about how to analyze data at different levels that motivated this journal article (yes, it's published in 2018, so I'm a bit slow). I also appreciate Yik Ting Choi's research assistance and Tanya Christ's suggestions on an earlier version.

## Declaration of conflicting interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) declared no financial support for the research, authorship, and/or publication of this article.

## References

- Barron, B. (2003). When smart groups fail. *Journal of the Learning Sciences*, 12(3), 307–359. [http://dx.doi.org/10.1207/S15327809JLS1203\\_1](http://dx.doi.org/10.1207/S15327809JLS1203_1)
- Benjamini, Y., Krieger, A. M., & Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 93, 491–507. <http://dx.doi.org/10.1093/biomet/93.3.491>
- Bryk, A. S., & Raudenbush, S. W. (2002). *Hierarchical linear models*. London: Sage.
- California Department of Education (2005). *Academic performance index*. Retrieved 8 May 2007: <http://www.cde.ca.gov/ta/ac/ar/index.asp>
- Chen, G., & Chiu, M. M. (2008). Online discussion processes: Effects of earlier messages' evaluations, knowledge content, social cues and personal information on later messages. *Computers and Education*, 50, 678–692. <http://dx.doi.org/10.1016/j.compedu.2006.07.007>
- Chiu, M. M. (2008). Flowing toward correct contributions during groups' mathematics problem solving: A statistical discourse analysis. *Journal of the Learning Sciences*, 17(3), 415–463. <http://dx.doi.org/10.1080/10508400802224830>
- Chiu, M. M., & Chow, B. W. Y. (2015). Classmate characteristics and student achievement in 33 countries: Classmates' past achievement, family SES, educational resources and attitudes toward reading. *Journal of Educational Psychology*, 107(1), 152–169.
- Chiu, M. M., & Fujita, N. (2014). Statistical discourse analysis: A method for modeling online discussion processes. *Journal of Learning Analytics*, 1(3), 61–83. <http://dx.doi.org/10.18608/jla.2014.13.5>
- Chiu, M. M., & Khoo, L. (2005). A new method for analyzing sequential processes: Dynamic multi-level analysis. *Small Group Research*, 36, 600–631. <http://dx.doi.org/10.1177/1046496405279309>
- Chiu, M. M., & Klassen, R. M. (2010). *Learning and Instruction*, 20, 2–17. Relations of mathematics self-concept and its calibration with mathematics achievement. <http://dx.doi.org/10.1016/j.learninstruc.2008.11.002>
- Chiu, M. M., & Kuo, S. W. (2009). From metacognition to social metacognition: Similarities, differences, and learning. *Journal of Education Research*, 3(4), 1–19.
- Chiu, M. M., & Lehmann-Willenbrock, N. (2016). Statistical discourse analysis: Modeling sequences of individual behaviors during group interactions across time. *Group Dynamics: Theory, Research, and Practice*, 20(3), 242–258. <http://dx.doi.org/10.1037/gdn0000048>
- Cohen, J., West, S. G., Aiken, L., & Cohen, P. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: Lawrence Erlbaum.
- Goldstein, H. (2011). *Multilevel statistical models*. Sydney: Edward Arnold.
- Hacker, D. J., & Bol, L. (2004). Metacognitive theory: Considering the social-cognitive influences. In D. M. McInerney & S. Van Etten (Eds.), *Big theories revisited*, vol. 4. (pp. 275–297). Greenwich, CT: Information Age Publishing.
- Kennedy, P. (2008). *Guide to econometrics*. New York: Wiley-Blackwell.
- Krippendorff, K. (2012). *Content analysis*. Thousand Oaks, CA: Sage.
- Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 1, 86–92. <http://dx.doi.org/10.1027/1614-2241.1.3.86>
- Molenaar, I., & Chiu, M. M. (2014). Dissecting sequences of regulation and cognition: Statistical discourse analysis of primary school children's collaborative learning. *Metacognition and Learning*, 9, 137–160. <http://dx.doi.org/10.1007/s11409-013-9105-8>
- Molenaar, I., & Chiu, M. M. (2017). Effects of sequences of cognitive actions on group performance over time. *Small Group Research*, 48(2), 131–164. <http://dx.doi.org/10.1177/1046496416689710>
- Orlitzky, M., & Hirokawa, R. Y. (2001). To err is human, to correct for it divine. *Small Group Research*, 32, 313–341. <http://dx.doi.org/10.1177/104649640103200303>
- Rasbash, J., Steele, F., Browne, W. J., & Goldstein, H. (2015). *MLwiN 2.33*. Bristol, UK: Centre for Multilevel Modeling, University of Bristol.
- Shadish, W. R. (2014). Analysis and meta-analysis of single-case designs. *Journal of School Psychology*, 52, 41–70. <http://dx.doi.org/10.1016/j.jsp.2013.11.005>
- Teasley, S. D. (1997). Talking about reasoning: How important is the peer in peer collaboration? In L. B. Resnick, C. Pontecorvo, R. Saljo, & B. Burge (Eds.), *Nato ASI subseries F: Discourse, tools and reasoning* (pp. 361–384). Berlin: Springer-Verlag. [http://dx.doi.org/10.1007/978-3-662-03362-3\\_16](http://dx.doi.org/10.1007/978-3-662-03362-3_16)

- Vygotsky, L. S. (2011). The dynamics of the schoolchild's mental development in relation to teaching and learning. *Journal of Cognitive Education and Psychology*, 10(2), 198.
- Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education*, 46, 71–95. <http://dx.doi.org/10.1016/j.compedu.2005.04.003>
- 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. <http://dx.doi.org/10.1007/s11412-011-9120-1>
- Wise, A. F., & Chiu, M. M. (2014). The impact of rotating summarizing roles in online discussions: Effects on learners' listening behaviors during and subsequent to role assignment. *Computers in Human Behavior*, 38, 261–271. <http://dx.doi.org/10.1016/j.chb.2014.05.033>
- Yoo, W., Mayberry, R., Bae, S., Singh, K., He, Q. P., & Lillard Jr., J. W. (2014). A study of effects of multicollinearity in the multivariable analysis. *International Journal of Applied Science and Technology*, 4(5), 9–19.

**APPENDIX A: PROBIT RESULTS**

**Table A1.** Summary of 3 ordered Probit regression models of group solution score showing unstandardized regression coefficients and standard errors in parentheses (N = 20 groups)

Independent variable	Group solution score					
	Model 1		Model 2		Model 3	
<i>Group mean mathematics grade</i>	0.15	**	0.12	*	0.07	*
	(0.05)		(0.05)		(0.04)	
<i>% Correctly evaluate → Correct, new idea</i>			11.39	*	-7.31	
			(5.62)		(6.11)	
<i>% Correctly evaluate</i>					6.78	
					(3.98)	
<i>% Correct, new idea</i>					1.05	
					(3.97)	
<i>McFadden's R<sup>2</sup></i>	0.25		0.35		0.62	
<i>Cohen's f<sup>2</sup></i>	0.33		0.54		1.63	

**Table A2.** Summary of 3 statistical discourse analysis models of correct, new idea using multilevel Probit rather than multilevel Logit (N = 3,234 turns of talk)

Independent variable	Correct, new idea					
	Model 1		Model 2		Model 3	
<i>Correctly evaluate (-1)</i>	0.98	***	3.36	**	0.96	***
	(0.09)		(1.22)		(0.09)	
<i>Solution</i>			0.41	***		
			(0.11)			
<i>Correctly evaluate (-1) * Solution</i>			0.30	*		
			(0.13)			
<i>Solution residual</i>					0.40	***
					(0.08)	
<i>Correctly evaluate (-1) * Solution residual</i>					0.23	*
					(0.11)	
Variance at each level	Variance explained					
Group (1%)	.00		.23		.21	
Turn of talk (99%)	.04		.05		.04	
Explained variance	.04		.05		.04	
<i>Cohen's f<sup>2</sup></i>	0.04		0.05		0.04	

## APPENDIX B: PROPOSITIONS AND PROOFS

**Proposition 1.** *The correlation between a composite variable and one of its non-trivial component variables is positive.*

$n$  = sample size

$r_{ab}$  = Correlation between variable  $a$  and variable  $b$

$$r_{ab} = \frac{n \sum a_i b_i - \sum a_i \sum b_i}{[n \sum a_i^2 - (\sum a_i)^2]^{0.5} [n \sum b_i^2 - (\sum b_i)^2]^{0.5}}$$

Let  $a$  be a dichotomous variable such that  $a = 1$  only if dichotomous variable  $b = 1$ , and variable  $c = 1$ .  
 $b$  is non-trivial such that  $b$  is not always 0, and  $b$  is not always 1.

Proof.

$$\sum a_i = \sum (a_i = 1)$$

$a = 1$  if and only if  $b = 1$  and  $ab = 1$

$$\sum a_i = \sum b_i a_i$$

$$\sum (a_i = 1) = \sum b_i a_i$$

$$n \sum (a_i = 1) = n \sum b_i a_i$$

$$\sum b_i \sum a_i = \sum b_i \sum (a_i = 1)$$

$$n > \sum b_i$$

$$n \sum (a_i = 1) > \sum b_i \sum (a_i = 1)$$

$$n \sum b_i a_i > \sum b_i \sum a_i$$

$$n \sum b_i a_i - \sum b_i \sum a_i > 0$$

$$\frac{n \sum a_i b_i - \sum a_i \sum b_i}{[n \sum a_i^2 - (\sum a_i)^2]^{0.5} [n \sum b_i^2 - (\sum b_i)^2]^{0.5}} > 0$$

**Proposition 2.** *Statistical power at the lower level is greater than statistical power at the higher level.*

Let  $n_L$  be the sample size at the lower level, and  $n_H$  be the sample size at the higher level.

Let  $\theta$  indicate the expected effect size.

$B_L(\theta)$  is the statistical power at the lower level, and  $B_H(\theta)$  is the statistical power at the higher level.

$\Phi$  is the quantile function.

$\sigma_\theta$  is the expected standard deviation of the effect size.

$B(\theta)$  can be approximated as follows:

$$B(\theta) \approx 1 - \Phi(1.64 - \theta n^{0.5} / \sigma_\theta)$$

Proof:

The sample size at the lower level must be larger than the sample size at the higher level,  $n_L > n_H$ .

$$\begin{aligned} n_L &> n_H \\ n_L^{0.5} &> n_H^{0.5} \\ \theta n_L^{0.5} &> \theta n_H^{0.5} \\ \theta n_L^{0.5} / \sigma_\theta &> \theta n_H^{0.5} / \sigma_\theta \\ -\theta n_L^{0.5} / \sigma_\theta &< -\theta n_H^{0.5} / \sigma_\theta \\ 1.64 - \theta n_L^{0.5} / \sigma_\theta &< 1.64 - \theta n_H^{0.5} / \sigma_\theta \end{aligned}$$

$\Phi$  is a monotonically increasing function.

$$\begin{aligned} \Phi(1.64 - \theta n_L^{0.5} / \sigma_\theta) &< \Phi(1.64 - \theta n_H^{0.5} / \sigma_\theta) \\ -\Phi(1.64 - \theta n_L^{0.5} / \sigma_\theta) &> -\Phi(1.64 - \theta n_H^{0.5} / \sigma_\theta) \\ 1 - \Phi(1.64 - \theta n_L^{0.5} / \sigma_\theta) &> 1 - \Phi(1.64 - \theta n_H^{0.5} / \sigma_\theta) \\ B_L(\theta) &> B_H(\theta) \end{aligned}$$