Uncovering what matters: Analyzing transitional relations among contribution types in knowledge-building discourse

Bodong Chen, Monica Resendes
University of Toronto
252 Bloor Street West, Suite 9-118, IKIT
Toronto, Ontario, Canada
bodong.chen@utoronto.ca
monica.resendes@utoronto.ca

ABSTRACT

Temporality matters for analysis of collaborative learning. The present study attempts to uncover temporal patterns that distinguish "productive" threads of knowledge building inquiry. Using a rich knowledge building discourse dataset, in which notes' *contribution types* and threads' *productivity* have been coded, a secondary temporal analysis was conducted. In particular, Lag-sequential Analysis was conducted to identify transitional patterns among different contribution types that distinguish productive threads from "improvable" ones. Results indicated that productive inquiry threads involved significantly more transitions among *questioning*, *theorizing*, *obtaining information*, and *working with information*; in contrast, responding to questions and theories by merely *giving opinions* was not sufficient to achieve knowledge progress. This study highlights the importance of investigating temporality in collaborative learning and calls for attention to developing and testing temporal analysis methods in learning analytics research.

Categories and Subject Descriptors

K.3.1 [**Computer Uses in Education**]: Collaborative learning; I.5.2 [**Design Methodology**]: Pattern analysis

General Terms

Algorithms, Design, Measurement

Keywords

Temporal analysis, Sequential analysis, Discourse analysis, Lag-sequential Analysis, Collaborative learning, Knowledge Building, Evidence-based research

1. INTRODUCTION

Collaborative learning is understood as a dynamic, interactive, and continuous process that evolves over time. From a social constructivist point of view, students learn through social interaction and communication with each other, with deep understanding achieved through exchanges that build on each other in a meaningful social context [23]. Thus, analysis of collaborative learning processes needs to attend to interpersonal interactions where group cognitions emerge [21], as well as interactions among ideas and conceptual artifacts from which knowledge progress occurs [20, 6].

The significance of the temporal dimension for understanding collaborative learning has been recognized by earlier research. However, in contrast to a wide spectrum of other popular techniques, it was only recently that temporality was explicitly recognized as an important angle for analyzing collaborative learning [3, 7, 9, 24]. This phenomenon could be attributed to a theoretical gap, in that learning theories generally do not take temporality into consideration [15]. It could also be caused by a methodological challenge, as little guidance is available for gathering temporal data and studying temporality of collaborative learning [8]. The present study attempts to bridge this gap, by advancing Lag-sequential Analysis as an analytic method emphasizing temporality and by showcasing its application in studying students' knowledge building dialogues.

2. BACKGROUND

2.1 Temporality of learning processes

A defining characteristic of interaction is that it unfolds in time. This notion is reflected in extensive research of collaborative groups [20, 21]. However, it is surprisingly rare that researchers investigate the temporal dimension of learning. In most collaborative learning studies, a quantitative method referred to as "coding and counting" is normally applied [22], which involves coding interaction data, counting occurrences of different types of interactions, and inspecting significant links between occurrences and interested learning outcomes. In some cases this approach is combined with narrative accounts of group interactions attending to causal and temporal links among interaction types. Undoubtedly, such analyses based on "coding and counting" are useful in uncovering patterns attributing to productivity of collaborative learning. However, applying such a "coding and counting" approach often means the loss of temporal or sequential information of interactions. As pointed out in previous research [7], measures produced by this approach tell us a certain proportion of interactional content that was coded in a particular category but nothing about the sequence of

LAK '14, March 24 - 28 2014, Indianapolis, IN, USA
Copyright 2014 ACM 978-1-4503-2664-3/14/03...$15.00.
http://dx.doi.org/10.1145/2567574.2567606

different categories; by aggregating counts over time, data get "flattened out" in the time dimension and information about temporal variation is unfortunately abandoned.

A growing body of research is starting to attend to the temporal dimension of collaborative learning. For example, Wise and Ming analyzed temporal patterns of knowledge construction in online discussions, focusing on consequences of role assignments [24]. By applying statistical discourse analysis attending to the progressive nature of knowledge construction, they identified "pivotal posts"—posts initiating new segments of discussion of higher knowledge construction quality—and found that assigning a summarizing role mid-discussion can facilitate group knowledge construction. In another study, Dyke and colleagues incorporated interactive sliding window visualizations to understand how interactions develop over time and provided means to detect temporal irregularities in collaborative learning [3]. Also interested in applying visualization techniques, Halatchliyski and colleagues developed a novel approach of tracing idea flow in an online learning community based on scientometric methodology [6]. By tracing temporal flow of ideas, their visualizations could help characterize contribution profiles of different authors. In summary, analyzing the temporal dimension of collaborative learning is fruitful and deserves more attention in learning analytics research.

2.2 Ways of contributing to knowledge building discourse

The need for analyzing temporality of discourse could be well demonstrated in research of *knowledge building*, which is a renowned constructivist pedagogy in the learning sciences [20]. To help students see ideas as knowledge builders, knowledge building pedagogy puts ideas at the center and engages students to take *cognitive responsibility* for improving them through communal discourse [18]. Such discourse is normally supported by Knowledge Forum, an online community space where students can contribute ideas in the form of *notes* to shared knowledge spaces, or *views* [19]. Previous research has extensively documented various epistemic roles in knowledge building classrooms, including posing questions, drafting hypotheses, and designing experiments to collect data in order to test hypotheses [18, 25]. From a research perspective, an intriguing question is how the synergy of different types of contributions could bring about productive discourse that effectively advances knowledge. To start tackling this question, researchers focusing on mapping dialogue moves important to knowledge building discourse have developed an inventory of "ways of contributing to explanation-seeking discourse" by applying the "grounded theory" approach on multiple years of Knowledge Forum notes produced by Grade 1-6 students [2]. Table 1 shows the current version of this scheme, which includes six major contribution categories and 24 sub-categories. This scheme has been useful in guiding content analysis of student online discourse and eliciting ways of contributing measures potentially useful for diagnosing discourse [13]. However, previous research in this area has neglected the temporal relations among different contribution types. Because knowledge building discourse is thought to be a complex, collaborative process of theory building that involves interactions among facts, theories built to explain facts, and various types of actions to improve theories [20], without any understanding of the temporal relations among different contribution types, an important piece of the picture remains missing. The traditional "coding and counting" approach flattens plural, complex discourse data and falls short in addressing interactions among different contribution types.

Table 1: Ways of contributing to knowledge building discourse coding scheme.

Questioning	1. Formulating an explanatory question 2. Asking a design question 3. Asking a factual question
Theorizing	4. Proposing an explanation 5. Supporting an explanation 6. Improving an explanation 7. Seeking an alternative explanation
Obtaining information	8. Asking for evidence 9. Designing experiment to test hypothesis 10. Reporting experiment results 11. Introducing facts from sources 12. Introducing facts from experience 13. Identifying design problems 14. Improving design problems
Working with information	15. Providing evidence to support a theory 16. Providing evidence to discard a theory 17. Weighing explanations 18. Accounting for conflicting explanations
Syntheses and Analogies	19. Synthesizing available ideas 20. Creating analogies 21. Initiating a rise-above
Supporting Discussion	22. Using diagrams to communicate ideas 23. Giving an opinion 24. Acting as a mediator

To bridge the gap of analyzing temporality in discourse-centric learning such as knowledge building, the present study advances Lag-sequential Analysis as a temporal analytic technique and applies it in knowledge building discourse analysis. More specifically, this study attempts to answer the following research question: Suppose we could reliably assess the productivity of a knowledge building dialogue, what underlying temporal patterns could distinguish dialogues which promote knowledge advancement?

3. METHODS

To tackle the research question, we performed secondary analysis of knowledge building discourse data already coded in previous research.

3.1 Data sources

Data analyzed in this study consisted of one year of Knowledge Forum notes, collected from Grade 1-6 students who were doing knowledge building in science. These students were from a knowledge building school in downtown Toronto. In this school, there is only one class in each grade, with approximately 22 students in each class. In this school, knowledge building pedagogy is introduced to students from Junior Kindergarten and Knowledge Forum is used from Grade 1. Thus, all students and teachers involved in this study were comfortable with this pedagogy and the supportive technology. In a typical school year, science learning in all grades is organized around an overarching topic, such as "water" or "trees," which further expands to a few big ideas in that content area. Students bring in their authentic problems of understanding, build explanations together based on their real ideas, and make constructive use of authoritative sources when necessary to improve their ideas. To make their ideas public and open to be improved by the community, students write notes in Knowledge Forum extensively.

In this study, a total of 1101 notes produced by six grades were analyzed. An overview of the grades, science units, and number of analyzed notes in each unit is provided in Table 2. Primary data analysis on this dataset was conducted in previous research concerning "ways of contributing to knowledge building discourse" and knowledge advancement in science learning [2, 13, 12]. This analysis included three aspects: (1) coding *types of contribution* in each note—based on content analysis using the aforementioned ways of contributing scheme; (2) identifying *inquiry threads*—through grouping sets of notes by shared principal problems [25]; and (3) determining *productivity* of each inquiry thread—based on whether there was any occurrence of *improving an explanation* in the ways of contributing scheme (see Table 1), inquiry threads were divided into *productive* and *improvable* categories. Two raters coded all notes and achieved a cumulative agreement rate of 95.52% (Grade 1, 99.27%; Grade 2, 98.65%; Grade 3, 82.5%; Grade 4, 99.57%; Grade 5/6, 97.63%). Previous reports of this dataset have mainly focused on percentages of contribution types and derived measures such as *contribution diversity* and *contribution richness* [1, 12].

3.2 Data analysis

To recognize temporality as an aspect of analysis on this dataset, we examined the transitional relations among contribution types in knowledge building discourse. Toward this end, we applied a temporal analysis technique named Lag-sequential Analysis (LsA) [16] on the dataset. Generally, one central goal of LsA is to determine whether there is cross-dependence between a specified behavior and another behavior that occurs earlier in time [4]. It identifies contingency relationships in a sequence of observed behaviors and enables researchers to explore cross-dependencies occurring in complex interactive sequences of behavior. As a simple and valuable method for summarizing interactions between behaviors, it has been found useful for studying interactions in many different contexts, including disruptive behaviors in classrooms and small group interactions [5, 11]. Comparing to "counting" measures in content analysis, LsA takes transitional relationships between contribution types into account. Thus, it could reveal significant insights into the temporal differences between productive and improvable knowledge building dialogues that are otherwise neglected.

To conduct LsA in this study, and to make LsA more accessible to other researchers as well, we implemented an R program based on previous studies [10, 17].[1] Using this R program, we computed a number of LsA measures, including: (1) *transitional frequencies* among all major contribution types—how often a particular transition occurred for a specified sequential lag; (2) *expected transitional frequency*—the expected number of times a transition would occur under the null hypothesis of independence or no relation between the codes; (3) *transitional probabilities*—a conditional probability indicating the likelihood of, for example, behavior B occurring, given that behavior A occurred (the probability of B, given A, see [10]); (4) *adjusted residuals*—z scores representing the statistical significance of particular transitions; and (5) *Yule's Q*—standardized measure ranging from -1 to +1 denoting strength of association. To uncover possible temporal patterns, we further compared these LsA measures between productive and improvable threads. *Yule's Q* was finally adopted to represent the strength of transitional association because it controls for base numbers of contributions and is descriptively useful (with a range from -1 to +1 and zero indicating no association).

[1] Code is available at https://github.com/dirkchen/LsA under the GNU General Public License, version 3.

Table 3: Descriptive statistics of basic contribution measures in two types of inquiry threads.

Measures	Productive	Improvable
Number of units	20.90 (9.15)	23.84 (12.44)
Number of units (merged)	14.23 (7.58)	15.74 (9.68)
Questioning	4.77 (3.33)	5.53 (3.47)
Theorizing	9.19 (5.49)	11.89 (6.34)
Obtaining information	2.42 (1.50)	1.89 (2.08)
Working with information	1.32 (2.06)	0.84 (1.64)
Synthesizing and Analogies	0.42 (0.92)	0.58 (1.02)
Supporting discussion	2.77 (3.04)	3.11 (2.66)

4. RESULTS

4.1 Were productive and improvable threads different in basic contribution measures?

The first comparisons between productive and improvable inquiry threads were made on a set of basic *ways of contributing* measures. These measures included: (1) number of discourse units (each unit is defined by a unique type of contribution rather than Knowledge Forum note), (2) number of discourse units after merging adjacent units of a same contribution type, (3) occurrences of each main contribution category, and (4) percentage of each main contribution category. Table 3 shows descriptive statistics of these basic contribution measures. To compare differences of productive and improvable inquiry threads, t-tests were conducted on these measures. Results found none of these descriptive measures significantly different between two types of threads. Thus, "counting" measures of contribution types were not adequate in predicting productivity of knowledge building discourse.

4.2 Uncovering sequential patterns: Transitions among contribution types

Analysis of the basic measures did not identify any significant difference between productive and improvable threads. So what has made them different? Although previous studies found numbers of *theorizing* and *working with information* contributions predicted individual knowledge advancement [2], these measures fall short in predicting progress at the group level. If occurrences of contribution types were not proper indicators, was it because of the interplay among different contribution types? Were some contribution types following each other more often in productive threads? Without analyzing temporality of knowledge building discourse an important piece of the story remains missing.

To tackle this problem, Lag-sequential Analysis measures were computed with sequential *lag* set as 1, focusing on direct transitions. To compare temporal difference between

Table 2: An overview of dataset.

Grades	Units	Note count	Thread count	Productive threads	Improvable threads
Grade 1	Water	298	12	9	3
Grade 2	Trees	117	6	4	2
Grade 3	Fungus	193	8	5	3
Grade 4	Rocks and Minerals	262	11	6	5
Grade 5/6	Astronomy	231	13	7	6

two types of inquiry threads, t-tests were conducted on *Yule's Q* scores (see Table 4).[2] Results indicated productive threads had significantly higher number of transitions from *working with information* to *theorizing* than improvable threads ($t(48) = 2.10$, $p < .05$), and marginally more transitions from *theorizing* to *questioning* ($t(42) = 1.79$, $p = .08$). It could be inferred that in productive threads students worked more constructively with resources and engaged in deepen questioning, as they incorporated information into theorizing and questioned proposed theories more often. Meanwhile, productive threads had less transitions from *theorizing* to *supporting discussion* ($t(44) = -2.23$, $p < .05$), as well as slightly less *supporting discussion* to *questioning* ($t(40) = -1.83$, $p = .08$). Detailed analysis found most supporting discussion fell into the *providing an opinion* subcategory. It appeared responding to theorizing and questioning by merely giving opinions was not sufficient for advancing knowledge.

To expand the analysis from immediate transitions to indirect transitions, we set sequential *lag* as 2 and applied the same Lag-sequential Analysis process. Results indicated productive threads had more indirect bidirectional transitions between *questioning* and *obtaining information* ($t(45) = 1.88$, $p < .05$ and $t(40) = 1.99$, $p = .05$), as well as between *theorizing* and *obtaining information* ($t(46) = 1.91$, $p = .06$ and $t(42) = 1.83$, $p = .07$). Therefore, the integration of *obtaining information* with *questioning* and *theorizing* is also contributing to idea improvement. At the same time, similar to the analysis of immediate transitions, significantly less indirect transitions from *questioning* to *supporting discussion* ($t(34) = -2.00$, $p = .05$) and from *supporting discussion* to *theorizing* ($t(35) = -2.06$, $p < .05$) were identified in productive threads.

Table 4: Means and standard deviations of *Yule's Q* in productive and improvable threads.

	Productive		Improvable	
Pairs	Lag = 1	Lag = 2	Lag = 1	Lag = 2
Q-OI		-0.23 (0.9)		-0.64 (0.7)
Q-SD		-0.67 (0.7)		-0.21 (0.8)
T-Q	0.37 (0.7)		0.05 (0.6)	
T-OI		-0.31 (0.8)		-0.70 (0.6)
T-SD	-0.22 (0.9)		0.32 (0.8)	
OI-Q		0.11 (0.9)		-0.39 (0.8)
WI-T	-0.35 (0.9)		-0.77 (0.6)	
SD-Q	-0.54 (0.8)		-0.14 (0.7)	
SD-T		-0.46 (0.8)		-0.27 (0.7)

5. DISCUSSION AND CONCLUSIONS

The present study argues for the importance of examining temporality in collaborative learning. Using a rich Knowledge Forum dataset, it attempts to investigate temporal patterns that can predict productivity of knowledge building discourse. To go beyond the "coding and counting" approach applied in many learning sciences studies, this study applies Lag-sequential Analysis to study transitional relations among different ways of contributing in productive and improvable dialogues. As results indicated, while traditional counting measures fall short in distinguishing two types of dialogues, a few sequential patterns were identified. In particular, productive threads of inquiry involved significantly more transitions among *questioning*, *theorizing*, *obtaining information*, and *working with information*, while improvable inquiry threads showed more transitions involving *giving opinions*. Therefore, responding to questioning and theorizing by merely giving opinions is not sufficient to achieve knowledge progress in knowledge building. These findings are consistent with existing literature in recognizing the importance of "constructive use of authoritative sources" as an important component of productive knowledge building [18, 25]. The bidirectional linkages between *obtaining information* and *theorizing* or *questioning* also highlight the progressive characteristic involving deepening inquiry as an important feature of effective discourse [20].

This study builds on previous research that highlights the importance of temporality. It contributes to learning analytics literature by introducing knowledge building perspectives and to knowledge building research by uncovering temporal patterns worth further investigation. By developing an R program, it also makes Lag-sequential Analysis more accessible for researcher. For future directions, we will explore possibilities of developing embedded knowledge building analytic tools to boost identified favorable discourse behaviors. Given manual content analysis applied in this study appears non-feasible for developing large-scale, real-time assessment tools, substantial efforts is needed to automate the process of coding by applying recent computational linguistics techniques [14]. Since previous studies have showed efficacy of fairly simple tools in facilitating metadisourse and boosting students' competencies [13], future work of knowledge building analytics based on the present study shows great promise in promoting collaborative learning.

[2]Considering the space limit, only pairs of contribution types that were found significantly different between two types of threads are presented in the table. A-B denotes transition from contribution type A to type B. Abbreviations of names of contribution types are used: Q–questioning; T–theorizing; OI–obtaining information; WI–working with information; SD–supporting discussion.

6. REFERENCES

[1] Y.-C. Chen, S. Park, and B. Hand. Unpacking the use of talk and writing in argument-based inquiry: Instruction and cognition. In J. van Aalst,

K. Thompson, M. J. Jacobson, and P. Reimann, editors, *The future of learning: Proceedings of the 10th international conference of the learning sciences (ICLS 2012) âĂŞ Volume 1, Full Papers*, pages 159–166. ISLS, Sydney, Australia, 2012.

[2] M. Chuy, J. Zhang, M. Resendes, M. Scardamalia, and C. Bereiter. Does Contributing to a Knowledge Building Dialogue lead to Individual Advancement of Knowledge? In H. Spada, G. Stahl, N. Miyake, and N. Law, editors, *Connecting Computer-Supported Collaborative Learning to Policy and Practice: CSCL2011 Conference Proceedings. Volume I - Long Papers*, pages 57–63. International Society of the Learning Sciences, 2011.

[3] G. Dyke, R. Kumar, H. Ai, and C. Rosé. Challenging assumptions: Using sliding window visualizations to reveal time-based irregularities in CSCL processes. In J. van Aalst, K. Thompson, M. J. Jacobson, and P. Reimann, editors, *The future of learning: Proceedings of the 10th international conference of the learning sciences (ICLS 2012) - Volume 1, Full Papers*, pages 363–370. ISLS, Sydney, Australia, 2012.

[4] S. V. Faraone and D. D. Dorfman. Lag Sequential Analysis: Robust Statistical Methods. *Psychological Bulletin*, 101(2):312–323, 1987.

[5] P. L. Gunter, S. L. Jack, R. E. Shores, D. E. Carrell, and J. Flowers. Lag Sequential Analysis as a Tool for Functional Analysis of Student Disruptive Behavior in Classrooms. *Journal of Emotional and Behavioral Disorders*, 1(3):138–148, July 1993.

[6] I. Halatchliyski, T. Hecking, T. Göhnert, and H. U. Hoppe. Analyzing the flow of ideas and profiles of contributors in an open learning community. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge - LAK '13*, page 66, New York, USA, 2013. ACM Press.

[7] M. Kapur. Temporality matters: Advancing a method for analyzing problem-solving processes in a computer-supported collaborative environment. *International Journal of Computer-Supported Collaborative Learning*, 6(1):39–56, 2011.

[8] K. Littleton. Productivity through interaction: An overview. *Learning with computers: Analysing productive interactions*, pages 179–196, 1999.

[9] N. Mercer. The Seeds of Time: Why Classroom Dialogue Needs a Temporal Analysis. *Journal of the Learning Sciences*, 17(1):33–59, Feb. 2008.

[10] B. P. O'Connor. Simple and flexible SAS and SPSS programs for analyzing lag-sequential categorical data. *Behavior Research Methods, Instruments, & Computers*, 31(4):718–726, Dec. 1999.

[11] L. L. Putnam. Small Group Work Climates A Lag-Sequential Analysis of Group Interaction. *Small Group Research*, 14(4):465–494, 1983.

[12] M. Resendes. *Enhancing knowledge building discourse in early primary education: Effects of formative feedback*. PhD thesis, University of Toronto, 2013.

[13] M. Resendes, B. Chen, A. Acosta, and M. Scardamalia. The Effect of Formative Feedback on Vocabulary Use and Distribution of Vocabulary Knowledge in a Grade Two Knowledge Building Class. In N. Rummel, M. Kapur, M. Nathan, and S. Puntambekar, editors, *To See the World and a Grain of Sand: Learning across Levels of Space, Time, and Scale: CSCL 2013 Conference Proceedings Volume 1 - Full Papers & Symposia*, pages 391–398. International Society of the Learning Sciences, 2013.

[14] C. Rosé, Y.-C. Wang, Y. Cui, J. Arguello, K. Stegmann, A. Weinberger, and F. Fischer. Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning*, 3(3):237–271, 2008.

[15] W.-M. Roth. *Learning science: A singular plural perspective*. Sense Publishers, 2006.

[16] G. P. Sackett. The lag sequential analysis of contingency and cyclicity in behavioral interaction research. *Handbook of infant development*, 1:623–649, 1979.

[17] G. P. Sackett, R. Holm, C. Crowley, and A. Henkins. A FORTRAN program for lag sequential analysis of contingency and cyclicity in behavioral interaction data. *Behavior Research Methods & Instrumentation*, 11(3):366–378, May 1979.

[18] M. Scardamalia. Collective cognitive responsibility for the advancement of knowledge. In B. Smith, editor, *Liberal education in a knowledge society*, pages 67–98. Open Court, Chicago, IL, 2002.

[19] M. Scardamalia. CSILE/Knowledge Forum. In A. Kovalchick and K. Dawson, editors, *Education and technology: An encyclopedia*, pages 183–192. ABC-CLIO, Santa Barbara, CA, 2004.

[20] M. Scardamalia and C. Bereiter. Knowledge building: Theory, pedagogy, and technology. In R. K. Sawyer, editor, *The Cambridge handbook of the learning sciences*, pages 97–115. Cambridge University Press, 2006.

[21] G. Stahl. *Group cognition*. The MIT Press, Cambridge, Massachusetts; London, England, 2006.

[22] D. D. Suthers. Technology affordances for intersubjective meaning making: A research agenda for CSCL. *International Journal of Computer-Supported Collaborative Learning*, 1(3):315–337, Aug. 2006.

[23] L. S. Vygotsky. *Mind in society: The development of higher psychological processes*. Harvard University Press, Cambridge, MA, 1978.

[24] A. F. Wise and M. M. Chiu. Analyzing temporal patterns of knowledge construction in a role-based online discussion. *International Journal of Computer-Supported Collaborative Learning*, 6(3):445–470, May 2011.

[25] J. Zhang, M. Scardamalia, M. Lamon, R. Messina, and R. Reeve. Socio-cognitive dynamics of knowledge building in the work of 9- and 10-year-olds. *Educational Technology Research and Development*, 55(2):117–145, Sept. 2007.