

# Evaluating the Distribution of Students' Contributions in Theorizing: Idea Evenness in Knowledge Building Communities

Gaoxia Zhu, Marlene Scardamalia, Ahmad Khanlari, and Haipeng Wan  
gaoxia.zhu@mail.utoronto.ca, marlene.scardamalia@utoronto.ca, a.khanlari@mail.utoronto.ca,  
haipengwan@gmail.com  
University of Toronto

**Abstract:** Asymmetric collaboration in CSCL environments may lead to exclusion of some students and less idea diversity than needed for a productive, inclusive community. To uncover the degree of asymmetric collaboration, the social entropy index (the sum of relative participation proportions of each individual) was adopted in some studies, but only quantitative indicators (e.g., the number of characters) were considered in the measurement. In this study, we define participation evenness plus the dimension of the quality of students' notes as "idea evenness". Adopting the entropy and using the depth of understanding within students' theorizing notes as the parameter, we analyzed the idea evenness of five Knowledge Building communities. We found that the idea evenness values are high in four of the communities and can reflect the different distributions of students' depth of understanding in different communities. The results indicate that students participated in theorizing evenly in most knowledge building communities.

## Introduction

Knowledge Building aims to facilitate students to work on real ideas and address their authentic problems by taking collective responsibility (Scardamalia, 2002). By joining efforts, people may achieve something new that could only emerge as a result of their interactions (Broadbent & Gallotti, 2015). Therefore, ensuring an environment for these interactions to happen is of vital importance. One of the negative factors that may harm students' interactions is participation inequality, which refers to the phenomenon that a tiny minority of users accounts for a disproportionately large amount of community content and other activities (Nielsen, 2006). For example, the "free rider" phenomenon (Burdett, 2003) and the perception of an asymmetric collaboration among teammates (Capdeferro & Romero, 2012) were identified as frustrating things by the students who participate in online Computer-Supported Collaborative Learning (CSCL). Egalitarian collaborative systems are the preferred future organizational form (Brafman & Beckstrom, 2006). In line with the Knowledge Building goal of re-creating schools as knowledge creating organizations, symmetric knowledge advancement is highlighted in Knowledge Building (Scardamalia, 2002).

Therefore, it is of great importance to study and measure with appropriate methodologies the manner that students distribute contributions, the extent to which uneven contributions occur in online collaboration, and furthermore, how uneven contributions may influence students' knowledge building. The social entropy index has been proposed as a possible approach for understanding system-level evenness (e.g., Bruno, 2010; Matei et al., 2006; Matei, et al., 2015). Mathematically, the normalized social entropy index is the sum of relative participation proportions of each individual divided by the log of the total number of people. It was derived from Shannon's Theory of Communication (Shannon & Weaver, 1998). The idea is that in the realm of communication all symbols are equally likely to occur if they are only decided by chance, just as all atoms are likely to be in a random state in physical systems (Shannon & Weaver, 1998). The opposite is that the more organized the system is, the lower the entropy will be. Matei et al. (2010) proposed that human affairs can be understood as atoms or symbols, and individual's contributions would not be greater than chance can predict in a purely random and unstructured state. In this situation, the social entropy of this group is maximized. However, when a group is structured (i.e. when members take on some specific tasks, interact with others in a preferential manner, or contribute more or less than chance can predict), the social entropy starts to decrease. Bruno (2010) measured the participation evenness of 25 wiki groups, and the number of characters that students contributed were utilized as the parameter. He found that there is a curvilinear relationship between participation evenness and learning outcomes, and an optimal level of evenness which was close to the high spectrum exists (see figure 4 in Bruno, 2010, p109). Adjusting the entropy to the scale of 0 to 1, it seems the optimal level ranges from 0.85 to 0.95. The results may somehow imply the optimal level evenness of knowledge building community although the study was not directly conducted in knowledge building communities.

Although the quality of students' contributions plays a significant role in a community, it is not considered in the evenness measurement (Biuk, Kelen, & Venkatesan, 2008). In this study, we added the

dimension of the quality of students' ideas to the participation evenness and define the variable as "idea evenness" – the sum of individual's relative apportionment of ideas (both quantity and quality) divided by the log of the total number of people.

In this exploratory study, we aim to investigate the value of idea evenness in five cases of Knowledge Building communities to see how idea evenness reflects the distribution of students' contribution in community level. This study may serve as an initial attempt to understand idea distribution in a community level and to help reveal the relationship between idea evenness and learning performance. In order to evaluate the quality of students' contributions, we employed the "epistemic complexity" and the "scientific sophistication" measures, proposed by Zhang, Scardamalia, Reeve, and Messina (2009). The epistemic complexity of ideas indicates students' efforts to produce theoretical explanations and elaborations of phenomena and the ideas that their community works on. The scientific sophistication dimension assesses to what extent students move from an intuitive to a scientific understanding.

## Methods and data analysis

Secondary data analysis consisted of 1209 notes posted in Knowledge Forum – an online environment supporting Knowledge Building (Scardamalia, 2004). The notes were written by Grade 1 to Grade 5/6 (blended grades) students at a Knowledge Building school in downtown Toronto. In each grade, one class with about 20 students was included in this analysis. Each class worked in a communal Knowledge Forum space which was considered as an online community.

Primary analysis on this dataset was conducted by Resendes (2013) and her colleague using the "ways of contributing" framework (Chuy et al., 2011) which consists of six dimensions and 24 sub-dimensions. The six dimensions are: questioning, theorizing, obtaining information, working with information, synthesizing and making analogies, and supporting discussion. A note might fall into several dimensions that are applicable. Theorizing plays an important role in knowledge advancement (Chen, Resendes, Chai, & Hong, in press) for exhibiting students' attempts to produce original ideas, to produce and improve explanations, and to express alternative directions (Resendes, Chen, Acosta, & Scardamalia, 2013), and for underscoring students' pursuit to construct new knowledge (Carey & Smith, 1993). Therefore, in this study we will focus only on theorizing (i.e. proposing an explanation, supporting an explanation, improving an explanation and seeking an alternative explanation) notes. All the theorizing notes were coded using epistemic complexity scale (1 = unelaborated facts, 2 = elaborated facts, 3 = unelaborated explanations, and 4 = elaborated explanations) and scientific sophistication scale (1 = pre-scientific, 2 = hybrid, 3 = basically scientific, and 4 = scientific). The overall agreement was 81.65% for epistemic complexity and 82% for scientific sophistication. With regard to a note, a composite score by multiplying the epistemic complexity score and the scientific sophistication score was considered as its depth of understanding (Zhang et al., 2009). For each student, the composite scores of his/her notes were added up, and the total score was used as the parameter to calculate the idea evenness of this community.

Using each student's share of depth of understanding, idea evenness was calculated using the normalized social entropy formula in each class separately. To better understand the essence of the social entropy index concept, we would paraphrase how Matei et al. (2010) discussed it in computer-mediated collaboration environment:

Suppose in an online community space (M), there are m students who contribute n notes in total (a note only belongs to a student). Let C be a set of each student's notes.

$$M = \{M_1, M_2, \dots, M_n\}, C = \{C_1, C_2, \dots, C_m\}, \text{ then } C=M$$

$S_i$  the  $i^{\text{th}}$  student's share (mathematical proportion) of notes in the note space M.

$$S_i = C_i / \sum_{j=1}^m C_j \quad \sum_{i=1}^m S_i = 1$$

If we have only one participant in M, then there is no uncertainty of who posted the notes. But if we have two, a degree of uncertainty of contribution happens. For the perspective of information theory (Shannon & Weaver, 1998), there are two possible answers to the question of who posted a note, which carry 1 bit ( $\log_2 2$ ) of information. If we have m participants, the answer to the above question will have m possibilities, which carry  $\log_2 m$  bit of information.

Mathematically, the social entropy of a random variable X is defined as:

$$H(X) = - \sum_{i=1}^m p(x) \log_2 p(x) \text{ where } p(x) \text{ represents the share of each student's contributions}$$

More participants indicate more diverse participation, the more distributed of the contributions by the students, and the higher social entropy, which may even hide the "lurker" problem. For example, the social entropy for a community with two students who contribute the same (1/2, 1/2) is 1, however, the social entropy for a community with six students who contribute (1/4, 1/4, 1/4, 1/4, 0, 0) is 2. Although the students in the first

community participate more evenly, and there are two lurkers in the second community, the social entropy index for the second group is higher than that of the first group. To compare the evenness of different communities with different numbers of participants and to handle the “lurker” problem, normalization should be obtained by dividing the entropy by its maximum score  $\log_2 m$ :

$$H_0 = H/H_{max} \text{ where } H_{max} = \log_2 m$$

The normalized social entropy index ranges from 0 to 1. “1” means perfect evenness, while “0” denotes total unevenness.

## Results and discussions

Table 1 shows the number of total notes and total theorizing notes in each class, indicating nearly half of the notes were coded for depth of understanding in each grade. Figure 1 shows 20 Grade 2 students’ shares of depth of understanding, while figure 2 shows 20 Grade 3 students’ (not the same 20 students in Grade 2) related shares, indicating grade 2 students theorized in an even manner than that of grade 3.

The results of the social entropy measurement are shown in table 2. Except for Grade 3, the idea evenness values of other grades seem to fit the optimal level according to Bruno (2010), indicating that students’ depth of understanding in theorizing is distributed in an even and desirable manner in these Knowledge Building communities. However, it should be noticed that the optimal entropy level (Bruno, 2010) was achieved through undergraduate students’ wiki participation and only quantitative indicators were used as the parameters. We need to be careful with the generalization of the results.

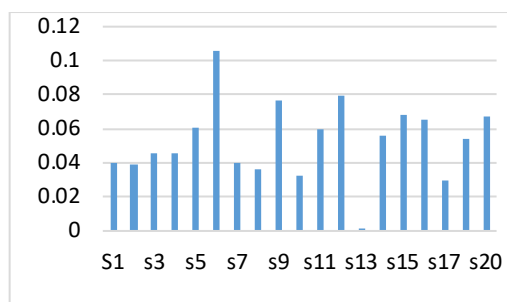
This idea evenness for Grade 2 is much higher than that of Grade 3, which indicates that the depth of understanding of Grade 2 students’ theorizing notes distributed more evenly than that of Grade 3 students’. Also, from figure 1 and figure 2, we noticed that in Grade 3, several students (S5, S11, S17) did not contribute in theorizing, and some students contributed much more than the other students, for example, the depth of understanding in S6’s and S7’s theorizing notes is significantly higher. The idea evenness values of the two classes reflect this kind of different distributions of depth of understanding.

**Table 1: The description of the number of total notes and total theorizing notes in each class**

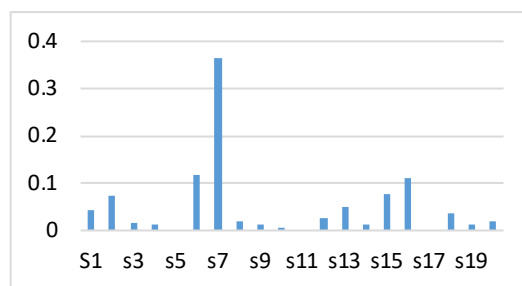
Number	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5/6
No. of notes	370	121	141	272	305
No. of theorizing notes	189	86	66	125	167

**Table 2: The value of idea evenness in each grade**

	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5/6
Idea evenness	0.88	0.95	0.73	0.89	0.88



**Figure 1.** Grade 2 individual share of depth of understanding.



**Figure 2.** Grade 3 individual share of depth of understanding.

## Conclusions and future directions

In this study, we examined students’ idea evenness in knowledge building communities by adopting the normalized social entropy measurement. Each student’s total score of depth of understanding of theorizing notes was used as the parameter. We found that except for Grade 3, the idea evenness values in other grades are relatively high. Also, the idea evenness values reflect the different distributions of students’ depth of understanding in

different grades (e.g., Grade 2 and Grade 3). The results indicate that students participated in theorizing evenly in most knowledge building communities, and the social entropy index can be used as an indicator of the distribution of students' contributions.

In this exploratory study, we only considered students' theorizing notes, given the importance of theorizing in knowledge advancement (Resendes et al., 2013). The next steps will be taking all notes posted online and face to face into consideration, and adding more qualitative indicators, for example, ways of contributing (e.g., original idea creation, connecting ideas, and critical appraisal), collective responsibility (at an individual level, intergroup level, and intragroup level) and so forth to the idea evenness measurement. Another issues worth studying is that right now, if one student contributes more notes, it may compromise the undesirable quality of his/her notes. Moreover, the relationship between entropy and students' learning performance needs to be studied specifically with rich knowledge building communities, and the results may inform a teacher how will the level of entropy in his/her class influence students' knowledge building and if he/she should take actions to address the idea unevenness issue, if exists.

## References

- Brafman, O., & Beckstrom, R. (2006). *The starfish and the spider: The unstoppable power of leaderless organizations*. New York: Portfolio.
- Broadbent, S., & Gallotti, M. (2015). *Collective intelligence, how does it emerge*. Retrieved from [https://www.nesta.org.uk/sites/default/files/collective\\_intelligence.pdf](https://www.nesta.org.uk/sites/default/files/collective_intelligence.pdf)
- Bruno, R. J. (2010). *Social differentiation, participation inequality and optimal collaborative learning online* (Unpublished doctoral thesis). Purdue university, West Lafayette, USA.
- Burdett, J. (2003). Making groups work: University students' perceptions. *International Education Journal*, 4(3), 177-191.
- Capdeferro, N., & Romero, M. (2012). Are online learners frustrated with collaborative learning experiences? *The International Review of Research in Open and Distributed Learning*, 13(2), 26-44.
- Carey, S., & Smith, C. (1993). On understanding the nature of scientific knowledge. *Educational Psychologist*, 28(3), 235-251.
- Chen, B., Resendes, M., Chai, C. S., Hong, H.-Y. (in press) Two tales of time: Uncovering the significance of sequential patterns among contribution types in knowledge-building discourse. *Interactive Learning Environments*.
- Chuy, M., Resendes, M., Tarchi, C., Chen, B., Scardamalia, M., & Bereiter, C. (2011). Ways of contributing to an explanation-seeking dialogue in science and history. *QWERTY - Interdisciplinary Journal of Technology, Culture and Education*, 6(2), 242-260.
- Matei, S., Oh, K., & Bruno, R. (2006). Collaboration and communication: A social entropy approach. In *Proceedings of National Communication Association Annual Convention*, San Antonio.
- Matei, S. A., Oh, K., & Bruno, R. (2010). Collaboration and communication in online environments: A social entropy approach. *Comunicare și comportament organizațional*, 83-99.
- Matei, S. A., Bruno, R., & Morris, P. (2015). Visible effort: Visualizing and measuring group structuration through social entropy. In S. A. Matei, M. Russell, & E. Bertino (Eds.), *Transparency in social media—Tools, methods and algorithms for mediating online interactions*. New York: Springer.
- Nielsen, J. (2006, October). Participation inequality: Encouraging more users to contribute [Web log message]. Retrieved from [http://www.useit.com/alertbox/participation\\_inequality.html](http://www.useit.com/alertbox/participation_inequality.html).
- Resendes, M. (2013). *Enhancing knowledge building discourse in early primary education: Effects of formative feedback* (PhD thesis). University of Toronto.
- Resendes, M., Chen, B., Acosta, A., & Scardamalia, M. (2013). The effect of formative feedback on vocabulary use and distribution of vocabulary knowledge in a grade two knowledge building class. In *To see the world and a grain of sand: Learning across levels of space, time, and scale: CSCL 2013 Conference Proceedings* (Vol. 1, pp. 391-398).
- Shannon, C. E., & Weaver, W. (1998). *The mathematical theory of communication*. Urbana: University of Illinois Press.
- Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Ed.), *Liberal education in a knowledge society* (pp. 67-98). Chicago, IL: Open Court.
- Scardamalia, M. (2004). CSILE/Knowledge Forum®. In *Education and technology: An encyclopedia* (pp. 183-192). Santa Barbara: ABC-CLIO.
- Zhang, J., Scardamalia, M., Reeve, R., & Messina, R. (2009). Designs for collective cognitive responsibility in knowledge-building communities. *The Journal of the Learning Sciences*, 18(1), 7-44.