Computers & Education 157 (2020) 103982

Contents lists available at ScienceDirect

Computers & Education

journal homepage: http://www.elsevier.com/locate/compedu

Fine grained analysis of students' online discussion posts

Mladen Raković [a,*], Zahia Marzouk [b], Amna Liaqat [c], Philip H. Winne [b], John C. Nesbit [b]

[a] University of North Carolina, 100 E Cameron Ave, Chapel Hill, NC, 27514, USA
[b] Simon Fraser University, 8888 University Drive, Burnaby, BC, V5A 1S6, Canada
[c] University of Toronto, 27 King's College Cir, Toronto, ON, M5S, Canada

ARTICLE INFO

Keywords:
Cooperative/collaborative learning
Distance education and online learning
Post-secondary education

ABSTRACT

Collaborative discussions should engage all students, not just a few who dominate ("leaders") while others participate as "followers" (Zhu, 2006). Cunningham (1991) noted that collaborating learners bring, discuss and debate multiple perspectives to develop their own position while acknowledging others' views. Higher levels of knowledge construction emerged when posts stimulated frequent reply by multiple participants (Aviv, Erlich, Ravid, & Geva, 2003) and were strongly content- and task-oriented (Rovai, 2007). So, to help students more actively and productively engage in knowledge-constructing discussions, an instructor needs to detect students' posts that do not stimulate replies, identify content those posts introduce, and guide students to revise posts to encourage peers' responses. However, such monitoring would be very time- and energy-consuming, especially in large-enrolment courses (Hura, 2010). To set a stage for developing a classifier to automate these tasks, we proposed 10 rhetorical moves characteristic of the interactive mode of Chi and Wylie's ICAP framework (2014) and categorized fine-grained content in discussion posts using these moves. We then identified attributes of posts that triggered a greater number of responses. Rhetorical moves of "asking questions," "requesting justification," "building-on," "giving a reason" and "making a claim" triggered more peer responses. Posts with moves of "disagreeing," "comparing" and "making claims" predicted students' achievement on a test and an argumentative writing task. We propose analytics for learners and instructors about forming and revising posts to promote constructive discussions and subsequent achievements.

1. Introduction

Since the 1990s, constructivism has become a preferred learning theory for designing many learning activities, including those in online courses. According to constructivism, learners explore complex topics and construct individual understandings based on experiences they bring to the learning context (Ertmer & Newby, 1993) supplemented by understandings they develop through social negotiation. Cunningham (1991) noted learners who collaborate bring, discuss and debate multiple perspectives about content to "arrive at self-chosen positions to which they can commit themselves, while realizing the basis of other views with which they may disagree" (p. 14).

Collaboration to construct knowledge in discussion groups should engage all students, not just a few who dominate ("leaders") while others participate as "followers" (Zhu, 2006). Collaborators' discussion activities can be illustrated using an undirected graph.

* Corresponding author.
E-mail address: mladen.rakovic@unc.edu (M. Raković).

https://doi.org/10.1016/j.compedu.2020.103982
Received 9 December 2019; Received in revised form 5 July 2020; Accepted 15 July 2020
Available online 17 July 2020

Nodes represent students and edges identify exchanges of information (Rabbany, El Atia, Takaffoli & Zaïane, 2014). Extensive and parity of information exchange would be indicated when each node has a large number of edges (a node's degree) and all nodes have a similar degree.

Characterizing online discussions by triggers and responses, Aviv et al. (2003) reported participants were involved in high levels of knowledge construction when posts were frequently responded to by multiple participants. Rovai (2007) augments mere patterns of information exchange by noting a participant should be aware that constructing content knowledge happens best in discussions that are strongly content- and task-oriented. Therefore, providing both frequent and on topic posts is important to advance knowledge construction.

Based on these findings, more productive discussions would not be characterized by "leaders" and "followers." Rather, an instructor should design the discussion or collaboration task to stimulate multiple exchanges approximately equally among all students. To guide the evolution of such "even-handed" discussions, an instructor needs to detect posts that do not receive responses, identify content those posts introduce, and offer those students' recommendations about how to form posts likely to increase frequency of response by peers. Such monitoring would be very time- and energy-consuming (Hura, 2010), especially in large enrolment courses. Automating learning analytics for both learners and instructors could be one remedy wherein statistical analysis, exploratory and predictive modelling facilitate understanding and recommend appropriate actions (Bichsel, 2012). Learning analytics serving this goal have three main components: (1) collecting learners' data, (2) automated analysis of those data to identify targets for action, and (3) taking action and monitoring results to improve learning iteratively (Gašević, Dawson & Siemens, 2015). One key to this sequence is meaningfully labeling data (Brooks, Greer, & Gutwin, 2014).

In this study, we propose and demonstrate a novel approach to labeling and analyzing content in students' posts. Our context is an undergraduate course in which the instructor assigned students to take a stand pro or con a controversial issue, and justify their stance with evidence from the course textbook. The goal of this online discussion activity was to promote constructing knowledge by negotiating mutual understanding. Students collaboratively discussed in small groups before they drafted a group position statement about the assigned topic. We analyzed fine-grained content in students' posts through using a framework, ICAP, proposed by Chi and Wylie (Chi, 2009; Chi & Wylie, 2014).

As described in more detail later, the ICAP framework categorizes cognitive processes underlying learners' overt behaviors into four modes: interactive, constructive, active, and passive. We segmented information in students' posts and classified each segment as representing one rhetorical move within the interactive mode of ICAP. Our findings inform future development of theoretically supported learning analytics for learners and instructors that guide forming posts to promote discussion and leverage the interactive mode. Because the instructor of the course from which we sampled posts intended students to build domain knowledge, we also investigated characteristics of content in posts that predicted achievement in the two course examinations and two argumentative writing tasks, a group position statement and a final essay.

2. Related work

2.1. ICAP framework

Chi (2009) and Chi and Wylie (2014) proposed a hierarchical framework to classify types of student learning behaviors according to distinguishing cognitive features: interactive, constructive, active and passive. In reverse order, the *passive* mode is defined as "learners receiving information without overtly doing anything related to learning" (Chi & Wylie, 2014, p. 221), e.g., listening to a teacher without taking notes. In the passive mode, learners theoretically use "attending" cognitive processes to store information in episodic form but do not integrate it with prior knowledge or form schemas. In the *active* mode learners are observed to manipulate information, for instance, highlighting text. Theoretical "gap filling" covert cognitive processes activate prior knowledge and assimilate new information into existing schemas. The *constructive* mode is identified when learners are observed to generate information beyond what is presented to them, for instance, when learners draw a concept map. Underlying cognitive processes integrate new information with prior knowledge by "elaborating, comparing and contrasting, generalizing, reflecting on, and explaining how something works" (Chi & Wylie, 2014, p. 228). When a learner exchanges information with peers or a learning system, the *interactive* mode is operational. "Mutually generative" cognitive processes synthesize peers' feedback and new ideas with prior knowledge and new content. The ICAP framework hypothesizes *passive* engagement leads to "minimal understanding," *active* engagement facilitates "shallow understanding," *constructive* engagement promotes "deeper understanding that might transfer" and the *interactive* mode enhances "understanding that might generate novel ideas" (2009). These relations are ordered as *interactive* > *constructive* > *active* > *passive*.

The ICAP model was validated in a group of studies investigating "usual" and collaborative learning in classrooms (see Chi & Wylie, 2014). We posit these findings can be extended to online discussions. For example, if a student reads other students' posts with little cognitive effort applied to understand content, the student's behavior is passive. A student may behave actively, e.g., by copying and pasting text fragments from posts into a draft essay. Moving along the ICAP ordering, a student behaving constructively might be observed to add information to a discussion beyond what has so far been contributed, e.g., providing a new explanation or developing a personal example. At the topmost level of ICAP, the interactive mode, a student may collaborate with a peer in "co-constructing while dialoguing" (Chi, Kang, & Yaghmourian, 2017, p. 12). We ground our analytical approach on the interactive mode of ICAP because it is hypothesized to activate the greatest engagement for learning and provides a solid framework describing learning activities in collaborative learning settings (Chi & Wylie, 2014).

2.2. ICAP in online discussions

A limited number of studies have applied the ICAP framework to investigate online learning. Wang, Yang, Wen, Koedinger, and Rosé (2015) used ICAP to categorize students' contributions in MOOC (massively open online course) discussions. They investigated relationships between types of student's posts and learning gains using the post as the sampling unit and coding each post according to nine categories: active (repeat, paraphrase, note-taking), constructive (ask novel questions, justify or provide reasons, compare or connect) and interactive (acknowledgment of partners' contribution, build on partner's contribution, defend and challenge). In an extension, Wang, Wen, & Rosé, 2016 and Chen (2018) designed and adapted a more detailed coding scheme to analyze students' online discussions. They also used a post as the sampling unit. To investigate the cognitive engagement of participants in MOOCs, Atapattu, Thilakaratne, Vivian, and Falkner (2019) computed an index of semantic similarity between each post and related course materials, and then classified posts into the active or the constructive ICAP mode. Wu and Wu (2018) also drew on the ICAP framework for their design of a supervised machine learning classifier to categorize students' posts as off-task, active question, active answer/-opinion, constructive question, constructive answer/opinion or interactive cognition. Taking a finer-grained approach, Vellukunnel et al. (2017) focused their analysis on questions in posts, classifying each as active, constructive, logistical or content-clarification.

Our research goes one step further. Rather than treat each post as an undifferentiated unit, we disassembled posts to capture complexities within them at a finer grain size. We assigned each idea unit in a post to one of 10 information categories characteristic of the interactive mode: *agree, disagree, give reason, request justification, ask question, build-on, share, compare, claim and answer*. We created an eleventh category to code non-contributions to distinguish on-topic idea units for further analysis. On-topic contributions are considered to be critical to advance productive interactions among participants in online discussions (Rovai, 2007).

We use the number of responses a post received to measure interactive engagement. Interactivity is theorized to positively affect information exchange which, in turn, benefits constructing domain knowledge (Chi, 2009). In particular, active participation in discussions that promote constructing knowledge through argumentation promises deeper engagement with course materials, improving learning and achievement of individual participants. Sinha, Rogat, Adams-Wiggins, and Hmelo-Silver (2015) and Slof, van Leeuwen, Janssen, and Kirschner (2020) demonstrated that students boosted acquisition of domain knowledge and performance on subsequent examinations when they actively engaged in a domain-related argumentation through reasoning, and providing evidence or counterarguments.

In this study, we explored how fine-grained characteristics of students' posts related to the number of responses per post. Further, because online discussions we studied required students to employ argumentation techniques – composing claims, supporting arguments, counterarguments and rebuttals – we leverage Chi and Wylie's (2014) interactive mode of the ICAP framework to investigate as well the link between content in students' online posts and their achievement on a midterm and final examination, and two offline argumentative writing tasks.

Accordingly, the following research questions guided our study:

1. What attributes of posts trigger a greater number of responses?
2. Do attributes of posts created by a single student across multiple threads predict the student's achievement on a midterm and final examination, and the course final paper (argumentative essay)?
3. Do attributes of posts within a group's discussion thread predict the group's achievement on the collaborative argumentative writing task?

3. Method

3.1. Data

Our data come from Canvas online discussions among 49 students registered in an online introductory educational psychology course at a university in western Canada. The majority of students were undergraduates preparing for or having declared various majors.

In the beginning of the course, students were randomly assigned to one of five discussion groups (Table 1), each with its own discussion board in Canvas. Across the semester, students engaged in four group discussions, each on a contemporary educational psychology issue, and collaboratively produced a written statement representing the group's position on the discussion topic, details of

Table 1
Number of participants and posts across groups and discussions.

Group ID	Participants	Self-esteem	Inclusion	Memorization	Inquiry
1	10	31	28	43	52
2	9	52	79	43	66
3	10	48	34	63	75
4	10	42	35	43	38
5	10	37	57	61	56
Total	**49**	**210**	**233**	**253**	**287**
M	-	42.0	46.6	50.6	57.4
SD	-	7.51	18.96	9.33	12.58

the four topics are included below. Each group essay had to include a thesis statement, arguments for and at least one counterargument with a rebuttal. A total of 20 group position statements and discussions leading to them formed our data.

Each discussion lasted one week. Students were required to make at least three separate posts on three different days throughout each discussion week. To create a high-quality post, the instructor guided students to directly address the discussion topics, deeply explore/explain presented ideas, extend the existing conversation and support their positions drawing on material in the course textbook. At the outset of discussions 1 and 3, and at the end of discussion 2, a course tutor sent a remainder of these expectations to students.

The four discussion topics were derived from point/counterpoint sections in the course textbook. Each of these sections provides arguments for and arguments against controversial topics in educational psychology. Thus, students had access to samples of arguments for and against topics they discussed online. Topic 1 related to chapter 3 assigned from the course text: "What should schools do to encourage students' self-esteem? Take a stand for or against schools trying to raise students' self-esteem, and justify your choice with evidence from the textbook. Make sure to make a choice for one option or the other in your initial posting." Topic 2 related to chapter 4 of the text: "Is inclusion a reasonable approach to teaching exceptional students? Take a stand for or against inclusive classrooms, and justify your choice with evidence from the textbook. Try to specify a particular context to think about and make sure to make a choice for one option or the other in your initial posting." The prompt for topic 3 was: "What's wrong with memorizing? Thinking about the ideas put forth in Chapter 8 point/counterpoint take a stand on whether memorizing information exactly is an effective learning strategy." Topic 4 addressed this prompt: "Are inquiry and problem-based learning effective teaching approaches? Thinking about the ideas in the Chapter10 point/counterpoint, take a stand on whether inquiry learning methods such as problem-based learning should be used instead of more traditional methods of teaching and learning."

We obtained data from the university learning management system (Canvas) about each post: post ID, student ID, group ID, ID of post being replied to (direction) and the text of the post.

3.2. Achievement

A midterm examination was comprised of 47 4-option multiple-choice items and 6 short-essay items. It counted 24% toward the course grade. The final examination consisted of 49 4-option multiple-choice items and 7 short-essay items. It counted 26% toward the course grade.

An assigned argumentative essay was scored out of 20 points. We only used the instructor assigned scores for argumentation (out of 12) in our analysis. Other dimensions of the scoring protocol (essay macrostructure and writing mechanics) do not relate to our research questions. Students were instructed to use an argumentation schema for their essay that included a thesis statement, arguments, counterarguments and rebuttals. Students were instructed to write 1500 to 2000 words of text on a specific issue or topic within the theme of teaching and learning. Links to over 160 educational psychology source articles were posted on the course web site. Students were required to use 4 to 7 of these articles to develop a single, very focused thesis, construct and articulate arguments and evidence that support the thesis, as well as counter arguments.

3.3. Codebook level 1: coding posts to idea units

We adopted Dunlosky, Hartwig, Tawson and Lipko's (2011) operational definition of an idea unit as an intermediate conceptual unit of information neither an atomic proposition (the smallest unit of information, e.g., "blue", "car") nor a set of complex propositions involving multiple atomic propositions in a sentence. To illustrate, slashes bound three idea units in this illustrative post: "/Yes, there are benefits as mentioned in the textbook/such as social skill development for exceptional students/and improved understandings of disabilities for non-exceptional ones./"

The coding protocol included additional conventions:

1. Because a quote or citation from the text or a post was not developed by the student, these were coded as one idea unit even if the quotation or citation contained multiple idea units per our operational definition.
2. Because a dependent clause cannot stand on its own, we coded them and their referent as one idea unit. Examples spanned constructions like "if … then", "not only … but also", "either … or", "neither … nor", "when (conditional)."
3. Text segments with a subject or object separated by a slash, e.g., "teachers/students can …" were coded as two idea units if the constituents were not synonyms, e.g., students/pupils. Synonymous constructions like "teachers/instructors" were coded as 1 idea unit.
4. Idea units having the same meaning within one post were treated as 1 idea unit.

3.3.1. Level 1 – post segmentation

Three coders segmented posts into idea units. To train, all three together segmented 47 randomly selected posts. Then another 25 posts were segmented independently. To examine inter-rater reliability, we followed the recommendation from Strijbos, Martens, Prins, and Jochems (2006) to calculate the proportion agreement from the perspective of each coder relative to other two coders. For example, coder 1 identified 172 idea units. Relative to this set, 143 idea units were identical for coder 2 (83% agreement from the perspective of coder 1). Conversely, coder 2 identified 167 posts of which 140 matched coder 1's segments (84% agreement from the perspective of coder 2). Considering all possible pairings (Table 2), the lower bound of agreement was 74% between coder 3 and coder

2, from the perspective of coder 3; the upper bound was 90% representing agreement between coders 1 and 3 from both perspectives. The three coders discussed points of disagreements until consensus was reached on the subsample of posts. We considered these results acceptable and proceeded to segment remaining posts independently.

3.4. Codebook level 2: operational definition of fine-grained categories based on ICAP's interactive mode

When a learner posts information, the interactive mode of ICAP is identified if the post involved "mutually generative" cognition that integrates information across other posts by introducing new ideas to construct and organize new knowledge. Chi and Wylie (2014) described specific learning activities associated with the interactive mode, including: defending and arguing a position in dyads or small group; asking and answering comprehension questions with a partner; debating with a peer about justifications; and discussing similarities and differences. In this study, we operationalized elements identified in this description as 10 finer-grained categories: agree, disagree, give reason, request justification, ask question, build-on, share, compare, claim and answer. Table 3 shows categories with operational definitions and common linguistic markers. Non-interactive contributions were coded (code 0) to mark idea units that did not substantively advance the discussion, yet which might generate many responses. An example is, "What is the due date for the position statement?" Although "sharing" can be a type of elaboration, we made it a separate category because students in online discussions tend to share artifacts. The code for sharing allowed us to map this behavior differently from the category of build on.

3.4.1. Codebook level 2: inter-rater reliability

In the sample of 47 posts used to train coders, a corpus of 366 idea units were classified according to the 11 coding categories (Table 3). Each coder assigned one code to each idea unit, producing a total of 1098 codes.

Because the distribution of codes across interaction categories was largely concentrated in the build on category, we used the AC1 statistic (Gwet, 2001, 2002) to calculate interrater reliability. AC1 corrects for the probability of chance agreement when raters identically score content in a category and therefore is considered more robust approach than a simple percentage agreement or Cohen's kappa (Haley, Thomas, Petre, & De Roeck, 2008, pp. 1–24). The AC1 interrater reliability for categorizing idea units across the 11 categories among the three coders was 0.81, which can be interpreted as 81% of agreement. We considered this acceptable, so coding of remaining posts was done independently.

3.5. Coding group's position statements

Two authors jointly rated each group's position essay for the presence and quality of thesis, argument, counterargument and rebuttal components per the scoring rubric for this assignment that was distributed by the instructor. One point was assigned for the presence of each instance of an identified component. For example, if a position statement contained two arguments, two points were assigned for the argument component. Quality of each argument component was judged using a 3-point scale (0, 0.5, 1). For example, if the quality of one argument was 0.5 and another was 1, total quality of the argument component is 1.5. When coding was completed, each position statement was given a total score equal to the sum of component scores.

3.6. Analytical plan

Simultaneous-entry multivariate regressions were computed to identify attributes of students' posts that predict number of responses (RQ1), examination and essay scores (RQ2) and group position statement scores (RQ3). To compute content overlap between discussion posts and exam questions, as needed for RQ2, we applied topic modelling as described later.

4. Results

4.1. Research question 1: WWhat are attributes of posts that trigger more responses?

Distributions of number of replies per post in each of the five discussion groups were examined for normality. Numbers of replies per post in groups 3 and 5 were skewed (kurtosis was 4.857 for group 3 and 5.856 for group 5). Therefore, we used a Kruskal-Wallis test (Kruskal & Wallis, 1952) to examine differences between groups. No statistically detectable difference between groups was observed with regard to number of replies per post, $H(4) = 7.046$ ($p = 0.133$).

Table 2
Inter-rater agreement for idea units segmentation for each pair of coders.

Coder		1		2		3	
	Total Created	Identical Idea Units	Proportion Agreement	Identical Idea Units	Proportion Agreement	Identical Idea Units	Proportion Agreement
1	172	-	-	143	83%	155	90%
2	167	140	84%	-	-	135	81%
3	185	166	90%	137	74%	-	-

Table 3
Categories and operational definitions.

ICAP category and code	Operational definition and common markers	ICAP category and code	Operational definition and common markers
Non-Contribution (0)	Unrelated to discussion topic *"Hi", "Hello", "Good point",* posts about wiki page, questions about assignments	Build on (6)	Adding details to reasons, answers, agreement, disagreement, claims *"For example", "In addition", "Moreover", "Further", "Furthermore",* Rephrasing idea units Providing summaries of posts for wiki page
Agree (1)	Affirmation of another student's claim *"I agree", "Yes"* + repeat an idea from thread, *"Of course", "You took the words right out of my mouth", "Right", "You 're spot on"*	Share (7)	Sharing a link or article (a learning artefact)
Disagree (2)	Negation of another student's claim *"I disagree", "No"* + express disagreement with prompt or another student's post, *"Though", "But", "However"* *"On the other hand", "Although", "In spite", "Despite", "I don't think", "I don't agree"*	Compare (8)	Comparing/contrasting similarities and differences Comparing personal experiences *"Similar", "Like", "While", "Compared to", "Rather than"*
Give reason (3)	Provide justification *"Because", "For", "In order to", "As", "Since", "Due to", "Owing to", "On account of", "To", "Having", "Given", "So that"* Idea units coming immediately before *"therefore", "hence", "which in turn", "so", "that's why"* Note: must have explicit marker	Make a claim (9)	A specific statement differing from previous claims that requires further support *"I believe", "I feel", "I think"*
Request justification (4)	Questions for information/clarification about a previous post, often a request for reason	Answer (10)	Answering a question from a previous post
Ask a question (5)	New question related to discussion, rhetorical questions *"Do you think?", "Who", "What", "Where", "When", "Why", "How", "I am curious about"*		

Since no difference in number of replies per post was detected between groups, data were pooled and a within-groups simultaneous-entry regression was computed to investigate whether categories comprising the interaction category of ICAP predicted the number of responses per post. The model with 10 predictors (non-contributions excluded) was statistically detectable, R^2=0.053, $F(9)$=5.108, $p < 0.001$. Statistical attributes of posts triggering peers' replies identified at the conventional threshold for type I error (p<.05) were: asking a question, β=1.617 (p=0.043); requesting justification, β=1.518 ($p \leq$.003); and building on, β=1.067 (p=0.017). Relaxing type I error rate, we also identified giving a reason β=0.885 (p=.094) and making a claim β=0.876 (p=0.052) as factors contributing to frequency of response to a post. Other characteristics did not predict responses to posts, all $p \geq 0.183$.

4.2. *Research question 2: Do attributes of posts created by a single student across multiple threads predict the student's achievement on a midterm and final examination, and the course final paper (argumentative essay)?*

For each student a total number of contributions per category was computed (e.g., total number of idea units categorized as claims, as answers, etc.). These variables were correlated with scores on the midterm (M = 16.68, SD = 3.21) and the final examinations (M = 17.23, SD = 4.32), and the argumentative essay (M = 10.31, SD = 1.91). A simultaneous-entry regression model with 10 predictors of midterm examination scores was not statistically detectable (R^2=0.295, $F(9)$=1.521, p=0.194). Two models using 10 predictors of final examination scores and the score on the argumentative essay were statistically detectable: R^2=0.450, $F(9)$= 2.948, p=0.008, and R^2=0.608, $F(9)$=5.588, $p < 0.001$), respectively.

Posts including rhetorical moves of disagree, compare, and make a claim predicted achievement on the final exam, respectively: β=-1.030 (p = 0.018), β=-0.659 ($p < 0.001$), and β=0.080 (p = 0.025). The same set of predictors – disagree, compare, and make a claim – were statistically detectable predictors of achievement on the essay, respectively: β=-0.509 (p = 0.003), β=-0.375 ($p < 0.001$), and β=0.026 (p = 0.055).

To analyze the overlap of topics covered in the final examination and online posts, we first identified a list of the 10 most prominent topics in examinations then calculated the overlap using latent Dirichlet allocation (Blei, Ng, & Jordan, 2003) based on topic modeling. Results revealed 52% of topics were common to the final examination and discussions.

4.3. *Research question 3: Do attributes of posts within a group's discussion thread predict the group's achievement on the collaborative argumentative writing task?*

The simultaneous-entry regression model with 10 predictors of the total score assigned to a group's position statement (M = 5.92, SD = 4.91) was not statistically detectable (R^2= 0.281, $F(9)$= 0.714, p= 0.236). The predictors in this model could not account for variance in scores assigned to the group's position statement.

5. Discussion, implications and future research

We developed a novel two-level coding scheme to analyze content in online discussion posts. Our methodology segmented posts into idea units, then categorized each idea unit according to fine-grained rhetorical moves characteristic of ICAP's interactive mode. This approach captures multidimensional attributes of single posts, a methodological advance in research on collaborative learning. Future research might refine this coding scheme to break down the "build-on" category to represent attributes such as examples, restatements and summaries.

Five attributes of posts predicted the volume of responses in our corpus: asking a question, requesting justification, building on, giving a reason and making a claim. These results resonate with previous studies showing that posts containing on-topic questions, requests, examples of course concepts and/or opinions are more likely to receive a response than the posts not including these rhetorical features (Arguelo et al., 2006; Jeong, 2003; Nandi, 2013; Nandi, Hamilton, & Harland, 2012). Our results indicate giving a reason is another rhetorical move that may boost interaction. In our study, the likelihood of receiving a response increased when students included reasoning in their posts, as in this example, *"Putting students in collaborative groups is beneficial because teachers/-schools should encourage all students to develop at their own pace and be proud of their accomplishments."* We speculate instructions guiding the writing task might have prompted students to read and respond to posts having reasons because the goal of discussions was to collaborate on writing a position paper with well-supported arguments, counterarguments and rebuttals. We recommend future research investigate whether our profile of response-stimulating posts replicates across topics and instructional contexts.

Findings relating to research question 2 reveal students who created more claims in their posts across four weeks of discussion were more likely to receive higher scores on both the final exam and the argumentative essay. Two findings from learning science may explain this relationship. Consider these example posts: *"I feel segmented inclusion is more beneficial to students." "I think the example by William James (pg 186) is possibly the worst way of applying memorization in teaching." "Problem solving skills would be transferrable into the group work and collaborative skills that the text talks about."* To develop these posts, students may have revisited content studied earlier in preparation for their writing assignment. This operationalizes the spacing effect which is well-documented to benefit learning (Carpenter 2020). The posts quoted also implicitly refer to or include information approaching an explanation for the claim. A recent meta-analysis demonstrates gains in achievement when learners self-explain (Bisra, Liu, Nesbit, Salimi, & Winne, 2018). Future research can directly test these hypotheses if students work on their assignments using software like nStudy n nStudy(Winne et al., 2019). nStudynStudy traces learning by logging time-stamped events that record information on which students operate (e.g., text selected for highlighting, content entered in a text field labeled "reason" in a structured note form, search queries) and operations students apply (e.g., re-viewing selections, generating information entered in the field labeled "reason" in a note template labeled "Explain note").

Unexpectedly, disagreeing and comparing related inversely to scores on the final exam and the argumentative essay. It may be important to note we operationalized disagreeing as simple negation of another student's claim (e.g., "*I don't agree*", "*I disagree*" …) without considering whether there was elaboration of the disagreement. Further analysis might investigate disagreement in combination with other categories, such as giving reasons, making claims, and requesting justifications. Regarding idea units we categorized as compare, these included comparing or contrasting personal experiences as well as empirical/logical information. The grading scheme for essays penalized students for using personal experience to support arguments rather than information from empirical sources. As well, students who compared/contrasted personal experiences may not have engaged thoroughly with textbook content. Together, these might negatively affect learning represented by scores on the achievement test. In future research codes distinguishing comparing/contrasting personal experiences versus empirical/logical information might be investigated.

Results relating to research question 3 indicate the rhetorical moves used by groups in our sample did not predict scores on a collaborative writing task. Our data do not permit deeper analysis of negotiation and contribution that underscored the group's writing process. Capturing and analyzing such data in future research may shed light on our finding.

Research documents instructors are challenged to provide timely and clear feedback to students participating in online discussions (Hew & Cheung, 2014; Khalil & Ebner, 2014, pp. 1305–1313). Without guidance from instructors, students may not succeed in analyzing why posts do not receive responses, hindering productive interaction among participants. The five attributes of posts we identified lay a foundation for developing theoretically informed learning analytics about posts (Marzouk et al., 2016). Such analytics can notify students about the attributes their posts omit and guide them to create new posts that trigger more peer responses. Revising posts to include these five rhetorical attributes also may promote critical thinking that enriches the learning process and improves achievement (Macfadyen, Dawson, Pardo, & Gašević, 2014). Future research is needed to investigate whether and to what extent scaffolding students to include these attributes in their posts stimulates productive interaction in online threaded discussions.

6. Limitations

As might be expected, some idea units did not clearly fit any category in our classification system. One occasional case was sharing personal stories or anecdotes that were not explicitly tied to a point we could recognize. We classified these as build-on although they do not represent in a "strong" way adding more information to flesh out a claim, reason, answer, agree or disagree.

We acknowledge potential debate about our operationalization of an idea unit. A post such as "Forcing students to memorize information inhibits the development of creative thought" would be identified as one idea unit while "Self-esteem programs should involve teachers, students and parents" would be coded as three idea units owing to our rule to acknowledge the verb has three objects. Resolving challenges like these should receive further attention.

Funding

This work was supported in part by grants to Philip H. Winne from the Social Sciences and Humanities Research Council of Canada #435-2016-0379 and Simon Fraser University.

CRediT authorship contribution statement

Mladen Raković: Conceptualization, Methodology, Formal analysis, Writing - original draft, Data curation. **Zahia Marzouk:** Conceptualization, Methodology, Data curation, Writing - original draft. **Amna Liaqat:** Methodology, Data curation, Writing - original draft. **Philip H. Winne:** Supervision, Conceptualization, Writing - original draft. **John C. Nesbit:** Conceptualization, Resources, Writing - original draft.

References

Arguello, J., Butler, B. S., Joyce, E., Kraut, R., Ling, K. S., Rosé, C., et al. (2006). Talk to me: Foundations for successful individual-group interactions in online communities. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 959–968).

Atapattu, T., Thilakaratne, M., Vivian, R., & Falkner, K. (2019). Detecting cognitive engagement using word embeddings within an online teacher professional development community. *Computers & Education, 140*, 103594.

Aviv, R., Erlich, Z., Ravid, G., & Geva, A. (2003). Network analysis of knowledge construction in asynchronous learning networks. *Journal of Asynchronous Learning Networks, 7*(3), 1–23.

Bichsel, J. (2012). *Analytics in higher education: Benefits, barriers, progress, and recommendations*. Louisville, CO: EDUCAUSE Center for Applied Research.

Bisra, K., Liu, Q., Nesbit, J. C., Salimi, F., & Winne, P. H. (2018). Inducing self-explanation: A meta-analysis. *Educational Psychology Review, 30*, 703–725.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research, 3*(Jan), 993–1022.

Brooks, C., Greer, J., & Gutwin, C. (2014). The data-assisted approach to building intelligent technology-enhanced learning environments. In *Learning analytics* (pp. 123–156). New York, NY: Springer.

Chen, Y. (2018). *Cognitive presence in peer facilitated asynchronous online discussion: The patterns and how to facilitate (doctoral dissertation*. USA: Syracuse University. Retrieved from https://surface.syr.edu/etd/912.

Chi, M. T. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science, 1*(1), 73–105.

Chi, M. T., Kang, S., & Yaghmourian, D. L. (2017). Why students learn more from dialogue-than monologue-videos: Analyses of peer interactions. *The Journal of the Learning Sciences, 26*(1), 10–50.

Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 219–243.

Cunningham, D. J. (1991). Assessing constructions and constructing assessments: A dialogue. *Educational Technology, 31*(5), 13–17.

Dunlosky, J., Hartwig, M. K., Rawson, K. A., & Lipko, A. R. (2010). Improving college students' evaluation of text learning using idea-unit standards. *The Quarterly Journal of Experimental Psychology, 64*(3), 467–484.

Ertmer, P. A., & Newby, T. J. (1993). Behaviorism, cognitivism, constructivism: Comparing critical features from an instructional design perspective. *Performance Improvement Quarterly, 6*(4), 50–72.

Gwet, K. (2001). *Handbook of inter-rater reliability: How to estimate the level of agreement between two or multiple raters*. Gaithersburg, MD: STATAXIS Publishing Company.

Gwet, K. (2002). Kappa statistic is not satisfactory for assessing the extent of agreement between raters. *Statistical Methods forInter-rater Reliability Assessment, 1*(6), 1–6.

Haley, D., Thomas, P., Petre, M., & De Roeck, A. (2008). *Using a new inter-rater reliability statistic System*.

Hew, K. F., & Cheung, W. S. (2014). Students' and instructors' use of massive open online courses (MOOCs): Motivations and challenges. *Educational Research Review, 12*, 45–58.

Hura, G. (2010). A student perspective on how online discussions should Be graded. *Journal of Educational Technology Systems, 39*(2), 163–172.

Jeong, A. C. (2003). The sequential analysis of group interaction and critical thinking in online. *American Journal of Distance Education, 17*(1), 25–43.

Khalil, H., & Ebner, M. (2014). MOOCs completion rates and possible methods to improve retention-A literature review. *EdMedia+ innovate learning*. Association for the Advancement of Computing in Education (AACE).

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association, 47*(260), 583–621.

Macfadyen, L. P., Dawson, S., Pardo, A., & Gaševic, D. (2014). Embracing big data in complex educational systems: The learning analytics imperative and the policy challenge. *Research & Practice in Assessment, 9*, 17–28.

Marzouk, Z., Rakovic, M., Liaqat, A., Vytasek, J., Samadi, D., Stewart-Alonso, J., & Nesbit, J. C. (2016). What if learning analytics were based on learning science? *Australasian Journal of Educational Technology, 32*(6).

Nandi, D. (2013). *A comprehensive framework with design principles for supporting interaction in fully online courses (Doctoral dissertation)*. Royal Melbourne Institute of Technology University.

Nandi, D., Hamilton, M., & Harland, J. (2012). Evaluating the quality of interaction in asynchronous discussion forums in fully online courses. *Distance Education, 33*(1), 5–30.

Rabbany, R., ElAtia, S., Takaffoli, M., & Zaïane, O. R. (2014). Collaborative learning of students in online discussion forums: A social network analysis perspective. In *Educational data mining* (pp. 441–466). Cham: Springer.

Rovai, A. P. (2007). Facilitating online discussions effectively. *The Internet and Higher Education, 10*(1), 77–88.

Sinha, S., Rogat, T. K., Adams-Wiggins, K. R., & Hmelo-Silver, C. E. (2015). Collaborative group engagement in a computer-supported inquiry learning environment. *International Journal of Computer-Supported Collaborative Learning, 10*(3), 273–307.

Slof, B., van Leeuwen, A., Janssen, J., & Kirschner, P. A. (2020). Mine, ours, and yours: Whose engagement and prior knowledge affects individual achievement from online collaborative learning? *Journal of Computer Assisted Learning*, 1–12. https://doi.org/10.1111/jcal.12466.

Strijbos, J. W., Martens, R. L., Prins, F. J., & Jochems, W. M. (2006). Content analysis: What are they talking about? *Computers & Education, 46*(1), 29–48.

Vellukunnel, M., Buffum, P., Boyer, K. E., Forbes, J., Heckman, S., & Mayer-Patel, K. (2017). Deconstructing the discussion forum: Student questions and computer science learning. In *Proceedings of the 2017 ACM SIGCSE technical symposium on computer science education* (pp. 603–608).

Wang, X., Wen, M., & Rosé, C. P. (2016). Towards triggering higher-order thinking behaviors in MOOCs. In *Proceedings of the sixth international conference on Learning Analytics & Knowledge* (pp. 398–407). ACM.

Wang, X., Yang, D., Wen, M., Koedinger, K., & Rosé, C. P. (2015). *Investigating how student's cognitive behavior in MOOC discussion Forums affect learning gains*. International Educational Data Mining Society.

Wu, Y., & Wu, W. (2018). A learning analytics system for cognition analysis in online learning community. In *Asia-pacific web (APWeb) and web-age information management (WAIM) joint international conference on web and big data* (pp. 243–258). Cham: Springer.

Winne, P. H., Teng, K., Chang, D., Lin, M. P-C., Marzouk, Z., Nesbit, J. C., … Vytasek, J. (2019). nStudy: Software for learning analytics about processes for self-regulated learning. *Journal of Learning Analytics, 6*, 95–106.

Zhu, E. (2006). Interaction and cognitive engagement: An analysis of four asynchronous online discussions. *Instructional Science, 34*(6), 451.