Effects of recommendations on message quality and community formation in online conversations

Evren Eryilmaz¹ · Brian Thoms² · Zafor Ahmed³ · Kuo-Hao Lee⁴

Received: 23 July 2020 / Accepted: 14 October 2020 / Published online: 21 October 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

As asynchronous online discussions gain a broader usage due to the COVID-19 epidemic, the need for understanding of students' ideas from unstructured textual data becomes more pressing. In this experimental study, we examine the effects of recommendations on message quality and community formation from voluminous online discussions. Drawing on literature from group cognition, knowledge building discourse, and learning analytics, we calculate message quasi-quality index (QQI) scores based on message lexical complexity and topic-related keyword usage by participants in explaining their ideas. Furthermore, we examine the empirical evidence on the relationship between QQI scores and participants' interactions. Finally, we visualize network structures via sociograms and hierarchically cluster participants to identify subgroups. Our analysis of 281 messages finds that recommendations helped participants to write more messages that compared alternative viewpoints and refined preliminary ideas with higher QQI scores. Results also show that recommendations cultivated a sense of collective agency to increase the opportunity for creativity and reduce the likelihood that peripheral participants will be dissatisfied and fail to identify with a community. Theoretical contributions and practical implications are discussed.

Keywords Annotation software · Computer-supported collaborative learning · Content analysis · Learning analytics · Recommender systems · Social network analysis

✉ Evren Eryilmaz
evren.eryilmaz@csus.edu

Brian Thoms
brian.thoms@csuci.edu

Zafor Ahmed
ahmedz@uhd.edu

Kuo-Hao Lee
klee@bloomu.edu

Extended author information available on the last page of the article

1 Introduction

Collaboration provides synergistic learning opportunities for students by sharing, reflecting, questioning, answering, and elaborating ideas. Computer-supported collaborative learning (CSCL) environments promote conversational modes of learning. Important components of CSCL environments, asynchronous online discussions (AODs) provide students time to reflect on topics-at-hand and make thoughtful contributions.

However, in a systematic literature review, Qiu et al. (2014) stated that it becomes more difficult to engage in such discussions where groups exceed 30 participants. As reported in extensive research (e.g., Damiani et al. 2015; van Heijst et al. 2019) that has examined AODs from large groups, students rarely are aware of other group members' ideas hindering the synthesis of diverse ideas that could be brought to solve problems or perform the task at hand. For example, Eryilmaz et al. (2019) found a large degree of navigational uncertainty statements that represent students' lack of awareness of interesting and useful ideas based on their preferences due to the sheer volume of discussion posts. This finding infers a limitation of AODs and their inability for linking messages across different threads in order to better synthesize ideas.

Voluminous information, now common in CSCL, can lead to "information overload", which demands more time and effort requirements to focus on relevant topics (for an overview, see Qiu 2019). However, not every student has sufficient capability and/or time to filter information. van Heijst et al. (2019), among others, found that students often become frustrated by spending a lot of time trying to find messages from voluminous AODs. Thus, information overload can lead to a breakdown in collaboration. For example, to cope with information overload, students can engage in "cumulative" or "disputational" talk, which both lack joint reasoning as noted by Mercer (2000). Furthermore, voluminous AODs can overwhelm instructors as they search for messages that require urgent responses to improve learning outcomes and reduce student drop-outs (Qiu 2019).

This study examines the effects of recommendations on message quality and learning community formation from large annotation-based literature discussions. Drawing on learning analytics, we calculate automatically message quasi-quality index (QQI) scores from voluminous discussions based on lexical complexity and students' topic-related keyword usage to explain their ideas. Next, empirical evidence observes the relationship between QQI scores and participants' interactions. These relationships are visualized using social network analysis and subgroups within the communities are identified by hierarchically clustering participants.

In the following section, we review related literature that forms the theoretical lens. We then describe our CSCL environment. Next, we outline the experimental study and present the results. The final sections discuss results along with implications for theory and practice, limitations, future work, and conclusions.

2 Review of related literature

2.1 Learning communities, group cognition, and knowledge building discourse

Hod et al. (2018) define the mutual interdependence of group members as an essential characteristic of a learning community. This definition means that a healthy learning

community capitalizes on its members who share common educational needs and contribute knowledge to the benefit of the community. Furthermore, it emphasizes the importance of connectedness so that a sense of belonging is developed. Accordingly, Hod et al. (2018) states that each individual in a learning community learns by actively supporting the learning of the group as a whole.

Numerous theories explain how learning takes place in a learning community. Our theoretical lens centers on group cognition and knowledge building discourse, which underscore that students' AOD activities and the formation of a learning community influence each other and intertwine. Group cognition posits that new ideas (i.e., knowledge objects or constructed artifacts) which emerge from AODs are not due to an individual alone because students learning effectively in groups encourage each other to reflect on different perspectives, ask questions, and articulate reasonings (Hernández-Lara et al. 2019). Accordingly, group cognition views each student as a node in a learning community (Zheng et al. 2015).

In knowledge building discourse, group cognition presumes that a fruitful collaboration requires diverse ideas based on evidence and reasoning. These ideas establish a basis for intersubjectivity that refers to how group members make sense of different perspectives in order to accomplish common goals (van Heijst et al. 2019). As Borge and Mercier (2019) noted, the purpose of this sense making is not to learn facts, but to build new ideas no single member had prior to collaboration. In sum, group cognition recognizes that knowledge building occurs at the individual, subgroup, and community levels. Drawing on these levels, Stahl (2013) explains that ideas created in a sub-group can impact other levels.

2.2 Annotation-based literature discussions

Annotations can support intersubjectivity among group members' understandings. As a specialized form of AOD, anchored or annotation-based discussion environments link or "anchor" discussion threads to annotations (i.e., highlighted and numbered passages within a text) to make academic reading more interactive. Prior research has investigated the advantages of such systems. For example, Eryilmaz et al. (2018) summarized the benefits of annotations on various learning activities such as (1) selecting relevant information, (2) articulating ideas by relating academic documents (e.g., journal articles) to prior understanding, (3) identifying comprehension gaps, and (4) revising incomplete and incorrect ideas. Chen et al. (2020) noted that anchored or annotation-based discussion environments improve students' reading comprehension, meta-cognitive skills, and learning outcomes via the immediacy of accessing information from text. Taken together, both reviews underscore three points important for our study. First, anchored or annotation-based discussion environments require a user-friendly interface to offer students an easy way to interact with peers while not hindering the interaction between instructional materials and annotations. Second, the act of annotation causes students to think about the content they are about to annotate in order to ensure the relevance and merit of their ideas. Third, annotations in small groups facilitate browsing and discovery of content segments (e.g., phrases and paragraphs) from text.

Learning communities vary in size. In large settings, Chen et al. (2020) found that students face navigation difficulties in finding and keeping track of content segments

from text due to an overwhelming quantity of messages annotated to the text. Eryilmaz et al. (2019) described this problem as disorientation and demonstrated students' lack of awareness of interesting messages based on their preferences, and their coping strategies with content and heat map analyses. These issues represent a paradox related to group size. While large groups expose students to many different perspectives, these groups demand more time and effort from students to organize construction of knowledge. When students cope information overload with social loafing and free-riding strategies, knowledge building discourse becomes less effective. These issues point to a potential constraint of annotation-based discussion environments when they are implemented in large group settings.

2.3 Message quality and learning analytics

Much research evaluates the success of CSCL based on the quality of students' messages. However, quality is often difficult to measure. Therefore, most CSCL studies employ a systematic message-coding schema to investigate message quality (for an overview, see Clarà and Mauri 2010). For example, Eryilmaz et al. (2018) and Hernández-Lara et al. (2019) showed that messages coded as reflection, clarification, and elaboration heavily impact collaborative knowledge construction in AODs. Furthermore, these studies demonstrated that the process through which students resolve comprehension problems is not linear.

Combining learning analytics with systematic message-coding schemas opens the door for a better understanding of group cognition and knowledge building discourse in learning communities. Support for ideas, uniqueness of ideas, writing styles, and voting have been examined by prior research. Regarding support for ideas, Shu and Gu (2018) identified 27% of 5090 messages from an AOD as mere copies of existing information because they lacked any interpretation and did not offer a venue to resolve existing comprehension problems. Regarding the uniqueness of ideas, Matuk and Linn (2018) showed that enhancing exiting preliminary ideas with more evidence and facts allows students to construct more coherent explanations than generating many preliminary ideas with lower levels of support. Regarding writing styles, Gunawardena et al. (2016) found that the categories of Gunawardena et al.'s (1997) message coding schema or the interaction analysis model do not correlate with message sentiment. Emphasizing gender, Rizvi et al. (2019) found that female students are more inclined to post help seeking messages. Finally, Procaci et al. (2019) integrated a rating mechanism into an AOD to allow community members to collaboratively identify message quality. Drawing on expert analysis, this same study reported that 74% of the ratings were consistent with the experts' decisions.

3 CSCL environment

In this section, we summarize our CSCL environment and discuss the characteristics that turn it into a perfect candidate for creating recommendations. We adopted the anchored or annotation-based AOD system of Eryilmaz et al. (2018) as our CSCL environment. This environment not only displays AODs in the same visual pane as the instructional material, but also clearly links threads to referenced content segments.

Research evaluating these CSCL environments shows that this linking reduces the effort required to clarify the context of a thread, increases the quantity of messages, and focuses discussions on understanding the meaning of a text (for a review, see Eryilmaz et al. 2018). However, as demonstrated in Eryilmaz et al. (2019) and Chen et al. (2020), too many messages annotated to the text can create navigation difficulties in finding and keeping up with the discussions anchored to text via annotations.

3.1 Recommender system

Recommender systems can reduce students' information overload by highlighting relevant information in a personalized way. Najafabadi et al. (2017) summarized the capabilities of recommender systems as (1) providing students learning resources (e.g., books, lecture notes, test items, and assignments) based on their learning styles, (2) assisting students to plan their semester schedule based on their needs and course availability lists, (3) helping students to find messages, which they may not have found themselves, and (4) stimulating students' collaboration by recommending peers who have similar learning interests. In this broad range of capabilities, algorithm optimization attracts many research efforts because recommender systems have to deal with big data. Put differently, recommender systems have high memory and CPU consumption (Damiani et al. 2015; George and Lal 2019a; Najafabadi et al. 2017; Zheng et al. 2015). Moreover, as described in these overviews, this research focus presumes that recommender systems with better information retrieval measures (e.g., accuracy, precision, recall) will improve students' learning performance by reducing their time-consuming search activities. However, as Howell et al. (2018) and Damiani et al. (2015) caution, promoting better awareness of hard to find information is not enough because learning also requires using the recommended information to resolve comprehension gaps. This caution is consistent with Santos et al. (2014) who did not find a correlation between high recommender system accuracy and user satisfaction.

A detailed search of the relevant literature yielded few studies that go beyond information retrieval measures to examine CSCL environments extended by recommender systems. Regarding reading material recommendations, prior research demonstrates that recommendations improve students' learning efficiency and satisfaction with electronic books by aiding them to find contents in an easy and efficient manner (Klašnja-Milićević et al. 2018; Núñez-Valdez et al. 2015). Regarding the formation of learning communities, prior research based on social network analysis shows that students mostly choose to collaborate with those who have similar interests and learning styles (Dascalu et al. 2015; Gašević et al. 2019). Regarding AODs, Damiani et al. (2015) demonstrated that recommendations increase users' communication frequency. The benefits of recommendations on the categorization of AOD messages are described as follow. Eryilmaz et al. (2019) found that recommendations decrease the number of navigational uncertainty markers that represent students' lack of awareness of the information they need from voluminous annotation-based literature discussions. Prior research also demonstrates that recommendations allow students to summarize ideas (George and Lal 2019b) and encourage them to ask questions (Reynolds & Wang, 2014).

We extended our CSCL environment by adopting Eryilmaz et al.'s (2019) collaborative filtering recommender system in order to help students find useful messages,

Fig. 1 A personalized recommendation list

which they may not have found themselves, from voluminous discussions based on their interests. Among possible design approaches, we selected collaborative filtering for four reasons. First, in line with the notion of group cognition, collaborative filtering underscores that it is the nearest neighbors or similar users who are responsible for a recommendation (Hernández-Lara et al. 2019). Second, as Dascalu et al. (2015) and Gašević et al. (2019) demonstrate, students who share similar interests are likely to collaborate with each other. Third, collaborative filtering can deal with deictic references invisible to a keyword metric. Lastly, the objective measurement of undergraduate students' learning interests is difficult (Gašević et al. 2019). However, collaborative-filtering methods can suffer from data sparsity problems, resulting in bad recommendations because collaborative-filtering methods require an existing matrix of initial ratings. Drawing on this limitation, Pan and Wu (2020) found that when data density is around 20% in a rating matrix of users' historical behaviors, collaborative-filtering methods perform well. Consequently, Pan and Wu (2020) noted that a data density greater than 20% effectively solves collaborative-filtering methods' data sparsity problem.

The recommender system employs users' rating data (1–5 stars) to compute k-nearest neighbors (k-NN). Drawing on Eryilmaz et al.'s (2019) experimental results, the similarity is computed by the constrained Pearson correlation coefficient. Furthermore, the number of the nearest neighbors (k) is set as three for decreasing noise. Finally, the system recommends messages these nearest neighbors rated highly, but the target user has not rated. Since users view only the very first recommendations (Najafabadi et al. 2017), the top four messages with highest prediction scores were the output to the users as recommendations. Accordingly, a target user's low rating of a message recommendation can adjust the nearest neighbors and recommendations.

A target user can view personalized recommendations on the left windowpane (see Fig. 1). When this user clicks on a recommendation (box 1), the instructional material on the right windowpane navigates to the referenced position of a similar user's annotation on text (box 2). Furthermore, the system highlights both the clicked recommendation (on the left) and the annotated passage (on the right) in red to assist the target user find useful annotations more efficiently. In Fig. 2, the target user moves

Fig. 2 Screenshot of a discussion thread linked to a similar user's recommended annotation

the cursor over a similar user's annotation on text (box 3). Based on this user input, the left windowpane navigates the discussion to the relevant thread and draws a red border around it (box 4). This design approach allows users quickly find the threads linked to annotations. In response to receiving this user input, the CSCL environment also displays small pop-up boxes adjacent to annotations (box 5). These pop-up boxes display a user's key idea for posting a message, learning community's average rating for that idea, and a 1–5 stars rating schema. Hence, these pop-up boxes afford opportunities for reflection and evaluation.

3.2 Control system

Eryilmaz et al.'s (2018) CSCL environment (see Fig. 3) functions as the control system to answer the research questions below because it isolates the effects of the recommender system.

Fig. 3 Control system

4 Research questions and methodology

To investigate the effects of recommendations on message quality and learning community formation, we address the following research questions:

1. How do recommendations affect students' interactions as defined by the interaction analysis model (IAM) and message quasi-quality index (QQI) scores?
2. How do message QQI scores relate to the IAM?
3. How do recommendations affect learning community formation?

The population for this experiment consisted of 70 sophomore information systems students from two independent sections of an undergraduate system analysis and design course. This population included 34 females and 36 males, with a mean age of 20.3 ($SD = 1.01$). Each section had 35 students. Primary learning objective of the course for the students was to describe different lifecycle models and explain the contributions of system analysis and design within them. To control the influence of the instructor, the same instructor taught both sections. One section was randomly assigned to the recommender system while the other section was assigned to the control system. The instructional topic for the experiment was the “Issues and Challenges of Agile Software Development with Scrum” (Cho 2008).

Participation in the online discussion was required for all students and accounted for 10% of their grade. To stimulate a knowledge building discourse, the students were instructed to reflect on the topic-at-hand and try to help each other develop a better understanding of the instructional topic. Students had two weeks to discuss the instructional topic. At minimum, students were instructed to annotate two passages from the text and evaluate the relevance of two peers' messages by using the 1–5 stars rating schema representing “not valuable” to “very valuable” in the first week. In the second week, at minimum, students were instructed to post two replies.

4.1 Quality measures

We started our investigation by coding messages based on the interaction analysis model (IAM). This coding schema, developed by Gunawardena et al. (1997), aims to serve as a framework to understand efficacious interactions among students. The schema includes five categories: sharing information (i.e. presenting new ideas to others), exploring dissonance (i.e. identifying inconsistencies among ideas), negotiating meaning (i.e. revising ideas based on peers' feedback), testing proposed synthesis (i.e. testing revised ideas against evidence, facts, and personal experience), and agreeing on new knowledge (i.e. approving ideas).

Existing studies consistently found that students' navigation difficulties represented by their uncertainty statements in finding interesting messages based on their preferences can constrain their knowledge building discourse (Chen et al. 2020; Eryilmaz et al. 2019). For example, “*I don't remember well but I have seen some arguments elsewhere about digital divide you brought up in your message.*” or “*I'm glad this was touched on, it was something I kept in mind during the earlier pages. I believe that PAM scores were mentioned somewhere but I can't find it.*”. As shown in Eryilmaz et al. (2019), recommendations can decrease the number of such navigational

uncertainty statements. Thus, recommendations can affect students' interactions as defined by the interaction analysis model (IAM).

Although the qualitative analysis of message content based on IAM is advantageous to count the frequency of different actions (categories), it does not afford analyzing how students explain their ideas in categorized messages. For example, how well do students refer to instructional materials' topic-related keywords to ask questions that represent understanding gaps? As shown in Gunawardena et al. (2016), utilizing topic-related keywords demonstrates students' understanding of instructional materials. Furthermore, are there messages with higher rate of readability and understandability? For example, Thoms et al. (2020) showed that message lexical complexity influences the number of reply sequences in a thread. Neglecting these characteristics of messages may pose severe limitations for the analysis and its generalizability regarding learning. Hence, we implemented Thoms et al.'s (2020) quasi-quality index (QQI) algorithm. This algorithm computed a QQI score for each coded message by the following:

$$QQI = \left(\frac{1}{n} \sum_{i=1}^n X_i \right) + \left(\frac{d}{\frac{1}{n} \sum_{i=1}^n d_i} * \frac{u}{\frac{1}{n} \sum_{i=1}^n u_i} \right)$$

n = Total elements

x = Post readability score

d = Post keyword density u = Total post non-stopwords

$d = \frac{k}{W-S}$

W = Total post words

S = Total post stopwords

k = Total post keywords

The algorithm has two main steps. As detailed in Thoms et al. (2020), it employs multiple readability metrics (e.g., Flesch-Kincaid Readability Test, Automated Readability Index) to calculate the average lexical complexity of a message. Linear mapping normalized these metrics to a 0-to 8-point scale. Furthermore, the algorithm compares and contrasts students' topic-related keywords in their messages with the instructional material's topic related keywords. We identified the keywords via the open-source word counter (<http://www.wordcounter.com/about.html>) program. The experts in the field validated these keywords and included missing ones related to the instructional material. As shown above, for a given message, keyword density is computed by dividing the total number of keywords within a message by the total number of words in a message minus its stop-words. In order to simplify assessment, QQI scores are reported out of 100 in the results section.

In general, empirical studies found that poor idea presentation is negatively associated with message quality (Gunawardena et al. 2016). Thoms et al. (2020) reported that uncertainty statements lower textual readability and decrease message QQI scores because uncertainty statements make ideas difficult to understand and require community members to expend more effort searching for relevant information. Because

recommendations can decrease the number of navigational uncertainty statements (Eryilmaz et al. 2019), they are likely to affect message QQI scores.

Gunawardena et al. (1997) described that the advancement of ideas in AODs can be represented by the IAM's sequential order of the message categories: sharing information (i.e. presenting new ideas to others), exploring dissonance (i.e. identifying inconsistencies among ideas), negotiating meaning (i.e. revising ideas based on peers' feedback), testing proposed synthesis (i.e. testing revised ideas against evidence, facts, and personal experience), and agreeing on new knowledge (i.e. approving ideas). Drawing on the QQI algorithm, when a message's keyword density increases that message's QQI score also increases. Therefore, messages with coherent explanations based on higher levels of support (evidence and facts) can have higher QQI scores than messages with preliminary ideas based on lower levels of support.

4.2 Learning community formation measures

Collaborative-filtering methods leverage on users with similar preferences (i.e., close neighbors) to generate recommendations. Thus, recommendations can resolve the collaboration breakdown among users with similar learning interests caused by information overload. In this way, recommendations can assist users to form subgroups based on their mutual interests within a community.

To gain insight in community formation, we first visualized network structures via sociograms. As a method of exploration, this approach works best because sociograms represent actors (participants) as nodes and their relationships (interaction patterns) as ties (Freeman 2004). Moreover, labeling nodes with their QQI scores explicitly shows how participants' positions on sociograms relate to how they explain their ideas. In line with Gunawardena et al. (2016), we defined collaborative relations as participants' directed responses to their peers' previous contributions. By analyzing AOD logs, we created node and edge tables for each community. Node tables included node id and node label, whereas edge tables included source, target, weight (frequency of communication), and direction of the edge (directed or undirected). We uploaded these tables to Gephi (<http://www.gephi.com>). The Yifan Hu layout mode is adopted to generate the sociograms. In our sociograms, relative node size represents degree centrality or the number of ties a node has. Moreover, connection line (arrow) thickness represents tie strength between two nodes (Freeman 2004).

Second, we computed network degree and closeness centrality for each participant because according to Dawson (2008), these measures predict an individual's perceived sense of community, and they influence how information flows in a learning community (Haythornthwaite 2002). Regarding degree centrality, two measures captured directional interactions: in-degree centrality (number of responses a participant received) and out-degree (number of responses a participant made). The higher the in-degree centrality, the more central a student is to a community. The higher the out-degree centrality, the more active a participant is in disseminating ideas in a community. Regarding closeness centrality, this measure indicates the distance between a participant and all other participants in a community. Accordingly, a participant who has a smaller closeness centrality is relatively closer to others in a community compared to a participant who has a larger closeness centrality.

Finally, based on the Ward method (Seifert 1995), we hierarchically clustered the participants in each community to identify if there were any subgroups within the communities. In line with Dawson (2008) and Haythornthwaite (2002), we employed the degree and closeness centralization indices to identify subgroups of participants in each community. To determine the appropriate number of clusters, we utilized the squared Euclidian distance as proximity measure and the Ward Method's dendrogram.

5 Results

Using a randomized dataset of 50 messages, three coders were trained to perform coding with the IAM. The unit of analysis for coding was determined to be the message because participants' messages were short and they fell in a single category. After training, all coders independently coded the remaining 281 messages. There were 149 messages ($M = 4.26$, $SD = 0.44$) in the recommender system and 132 messages ($M = 4.14$, $SD = 0.40$) in the control software. The intercoder reliability was measured by Krippendorff's alpha. The results of reliability analysis among the coders was 0.76, indicating satisfactory agreement. All disagreements among coders were discussed and resolved. There were 145 ratings ($M = 4.14$, $SD = 0.49$) in the recommender system and 116 ratings ($M = 4.09$, $SD = 0.37$) in the control software.

5.1 Interaction analysis model results

Based on the categories of the IAM, we created six message scores for each participant. We calculated the message scores as the proportion of participants' messages in each category. For example, if a participant posted 6 messages and 4 of those messages were coded as exploring dissonance, the exploring dissonance message score for the participant was 4/6 or 0.67. Table 1 summarizes the descriptive statistics and t-tests used to analyze the data. Table 1 shows that participants who received recommendations posted more messages in exploring dissonance and negotiating meaning, but fewer messages in the sharing information category than control group participants.

Table 1 IAM results

Category	Control Software ($n = 35$)		Recommender System ($n = 35$)		Test Statistics	
	M	SD	M	SD	p	Cohen's d
Sharing information	0.43	0.22	0.23	0.18	<0.001	−0.99
Exploring dissonance	0.24	0.13	0.32	0.14	0.02	0.60
Negotiating meaning	0.14	0.13	0.21	0.12	0.03	0.56
Testing proposed synthesis	0.07	0.11	0.10	0.13	0.33	0.25
Agreeing on new knowledge	0.06	0.12	0.08	0.13	0.54	0.16
Off-topic messages	0.07	0.11	0.05	0.09	0.41	−0.20

Note: Cohen (1988) defined effect sizes as small ($d = 0.2$), medium ($d = 0.5$), and large ($d > 0.8$)

Table 2 QQI score results

	Control Software (<i>n</i> = 35)		Recommender System (<i>n</i> = 35)		Test Statistics	
	M	SD	M	SD	<i>p</i>	Cohen's <i>d</i>
QQI scores	67.84	8.87	73.12	7.97	0.01	0.63

5.2 Message QQI score results

Table 2 summarizes the participant level descriptive statistics and the t-test used to analyze the data. Based on Table 2, participants who received recommendations posted messages with higher QQI scores than control group participants.

5.3 Relationship between message QQI scores and IAM

Table 3 summarizes the descriptive statistics and t-tests used to analyze the data. Results reveal that participants who received recommendations posted messages in exploring dissonance and negotiating meaning categories with higher QQI scores than control group participants.

Table 3 Relationship between message QQI scores and IAM

QQI score for category	Control Software (<i>n</i> = 35)		Recommender System (<i>n</i> = 26)		Test Statistics	
	M	SD	M	SD	<i>p</i>	Cohen's <i>d</i>
Sharing information	67.10	11.90	68.83	15.07	0.62	0.13
Exploring dissonance	Control Software (<i>n</i> = 30)		Recommender System (<i>n</i> = 33)		Test Statistics	
	M	SD	M	SD	<i>p</i> value	Cohen's <i>d</i>
	67.13	12.58	74.35	13.89	0.04	0.55
Negotiating meaning	Control Software (<i>n</i> = 21)		Recommender System (<i>n</i> = 28)		Test Statistics	
	M	SD	M	SD	<i>p</i> value	Cohen's <i>d</i>
	70.38	12.85	80.27	14.24	0.02	0.72
Testing proposed synthesis	Control Software (<i>n</i> = 10)		Recommender System (<i>n</i> = 13)		Test Statistics	
	M	SD	M	SD	<i>p</i> value	Cohen's <i>d</i>
	80.30	4.19	81.73	7.42	0.59	0.24
Agreeing on new knowledge	Control Software (<i>n</i> = 11)		Recommender System (<i>n</i> = 11)		Test Statistics	
	M	SD	M	SD	<i>p</i> value	Cohen's <i>d</i>
	63.54	13.19	68.18	16.79	0.48	0.31

Fig. 4 Sociogram of the learning community from the control software

5.4 Learning community formation results

The sociograms below depict community formation by using the investigated messages reported above. Illustrated in Figs. 4 and 5, each node represents a participant in the

Fig. 5 Sociogram of the learning community from the recommender system

Table 4 Degree and closeness centralization indices results

	Control Software (<i>n</i> = 35)		Recommender System (<i>n</i> = 35)	
	M	SD	M	SD
In-degree	2.20	3.07	2.89	2.60
Out-degree	2.20	1.18	2.89	0.99
Closeness	0.42	0.18	0.35	0.07

AODs. We labeled each node with a participant's average QQI score to portray how participants' positions on the sociograms relate to the quality of their messages. Table 4 shows the descriptive statistics for degree and closeness centralization indices per participant in each group.

Next, we hierarchically clustered participants in each group via the ward method. The inspection of the dendrogram diagrams allowed us to determine empirically the following clusters that correspond to a learning community: central, intermediate, and peripheral. Table 5 shows the number of participants in clusters and z-test results that were used to compare differences between communities.

Table 5 Cluster analysis results

	Control Software (<i>n</i> = 35)		Recommender System (<i>n</i> = 35)		Test Statistics	
	Frequency	Proportion	Frequency	Proportion	<i>z</i>	<i>p</i>
Number of central participants	4	0.114	8	0.229	1.27	0.20
Number of intermediate participants	12	0.343	17	0.486	1.21	0.23
Number of peripheral participants	19	0.543	10	0.286	−2.18	0.03

Table 6 presents statistically significant differences in clusters between the two communities with respect to the IAM.

Table 6 Differences in clusters between the two communities with respect to the IAM

Peripheral cluster	Control Software (<i>n</i> = 19)		Recommender System (<i>n</i> = 10)		Test Statistics	
	M	SD	M	SD	<i>p</i>	Cohen's <i>d</i>
Sharing information	0.41	0.18	0.14	0.12	<0.001	−1.77
Negotiating meaning	0.14	0.12	0.24	0.10	0.03	0.90
Intermediate cluster	Control Software (<i>n</i> = 12)		Recommender System (<i>n</i> = 17)		Test Statistics	
	M	SD	M	SD	<i>p</i>	Cohen's <i>d</i>
Sharing information	0.43	0.23	0.23	0.17	0.01	−0.99

Table 7 QQI scores organized by clusters within communities

	Central Cluster			Intermediate Cluster			Peripheral Cluster			ANOVA Test Results
	n	M	SD	n	M	SD	n	M	SD	
Control software	4	79.50	4.42	12	70.07	10.17	19	64.08	5.89	$F(2, 32) = 7.66, p = 0.002, \eta_p^2 = 0.69$
Recommender system	8	83.61	2.44	17	73.61	3.22	10	63.70	4.58	$F(2, 32) = 71.52, p < 0.001, \eta_p^2 = 2.11$

There was no statistically significant difference in QQI scores among respective clusters between the communities. Table 7 summarizes the descriptive statistics of each cluster's QQI score within the communities and ANOVA test results.

Regarding control group participants, post hoc comparisons using the Tukey HSD test indicated that the central cluster ($M = 79.50, SD = 4.42$) had significantly higher QQI scores than the peripheral cluster ($M = 64.08, SD = 5.89$) ($p < 0.001$). There were no other significant differences in pairwise comparisons. Regarding participants who received recommendations, post hoc comparisons using the Tukey HSD test indicated that the central cluster ($M = 83.61, SD = 2.44$) had significantly higher QQI scores than the intermediate cluster ($M = 73.61, SD = 3.22$) ($p < 0.001$) and the peripheral cluster ($M = 63.70, SD = 4.58$) ($p < 0.001$). Finally, the intermedia cluster ($M = 73.61, SD = 3.22$) had higher QQI scores than the peripheral cluster ($M = 63.70, SD = 4.58$) ($p < 0.001$) in the recommender system group.

6 Discussion

The paper has examined the effects of recommendations on message quality and learning community formation. We now turn to a discussion of our important findings.

6.1 Effects of recommendations on the interaction analysis model

To begin with, Table 1 reveals that participants who received recommendations posted more messages in exploring dissonance and negotiating meaning categories, but fewer messages in the sharing information category than control group participants. With respect to knowledge building discourse (van Heijst et al. 2019), messages in exploring dissonance and negotiating meaning categories are pivotal for refining ambiguous, figurative, and partial understandings. In this way, participants assigned to the recommender system sustained a productive line inquiry to gradually refine and strengthen their preliminary ideas that can be incomplete or incorrect and thus reducing group performance and learning outcomes.

This new insight is interesting and important because, as Matuk and Linn (2018) noted, students who enhance existing preliminary ideas with higher levels of support (evidence and facts) construct more coherent explanations than those who generate

many preliminary ideas with lower levels of support. With respect to the literature on collaborative filtering recommender systems, this finding supports Procaci et al. (2019) who found that 74% of ratings in a community accurately identified message quality in AODs. Moreover, it enriches prior research, which has shown that recommendations (i.e., rating-based information) allow students to summarize ideas George and Lal (2019a) and encourage them to ask questions (Reynolds & Wang, 2014). Our interpretation of this finding is that recommendations afforded participants to keep track of what is going on in the discussion based on their mutual interests.

In contrast, control group participants' messages in the sharing information category implies fewer threads achieved intensive knowledge building discourse in the control group. Accordingly, it is possible that finding relevant messages within a large quantity of messages ($n = 132$) was burdensome for these participants. Due to this disorientation problem, as defined in Eryilmaz et al. (2019), participants assigned to the control software might have scanned for points where they can most easily contribute their existing individual ideas, rather than collaboratively refining and constructing new conceptual understandings. It is noteworthy that this problem is persistent and widespread in large AODs (Qiu 2019). Accordingly, our finding supports Chen et al.'s (2015) remark that when students are unaware of interesting messages from AODs based on their preferences, they are hindered in their knowledge building discourse.

6.2 Effects of recommendations on message quasi-quality index scores

Second, Table 2 shows that participants who received recommendations posted messages with higher QQI scores than the control group participants. With respect to the QQI algorithm, this result underscores that these participants wrote better explanations because their messages were highly readable, and they integrated more topic-related keywords to fill understanding gaps. Given that messages can hinder or support collaboration in AODs, we can deduct that understanding and relating their ideas to those of others did not hinder these participants as compared to the ones assigned to the control group.

6.3 Relationship between the interaction analysis model and message quasi-quality index scores

In the same vein, we found that participants who received recommendations posted messages in exploring dissonance and negotiating meaning categories with higher QQI scores than control group participants. This relationship demonstrates a deliberate effort to improve community knowledge based on mutual interests. Accordingly, recommendations helped participants to refine their preliminary ideas when an overwhelming quantity of messages annotated to the text. Collectively, the quality cues provided by the QQI algorithm on coded messages overcome the limitations of prior research that employed systematic coding schemas without learning analytics tools (Gunawardena et al. 2016). Moreover, these findings advance the literature on the use of learning analytics tools that neglected content analysis to analyze knowledge embedded in AOD messages (Hernández-Lara et al. 2019). With respect to the literature on large groups, these findings enrich Qiu (2019) who found that group size does not correlate with the average number of threads by a student.

6.4 Effects of recommendations on learning community formation

Third, Table 5 shows that recommendations reduced the number of participants peripheral to a learning community. Furthermore, we found that participants in the recommender system group's peripheral cluster posted more messages coded as negotiating meaning, which represents adjustments of preliminary ideas by comparing alternative viewpoints or adopting evidence provided by others. From these results, we can deduct that peripheral participants were aware of what was going on in the discussion based on their interests. This is an important contribution because collaborative learning not only involves knowledge co-construction, but also developing a feeling of belonging to a community. In terms of group cognition (Zheng et al. 2015), these findings indicate that recommendations cultivated a sense of collective agency to increase the opportunity for creativity and reduce the likelihood that peripheral participants will be dissatisfied and fail to identify with the community.

In contrast, control group's peripheral cluster had difficulty refining preliminary ideas that might have led to information loss or misunderstandings. Regarding a potential information loss, control group's sociogram (Fig. 4) clearly shows that two participants with average QQI scores of 84.00 and 82.50 in the peripheral cluster established very few relations with the other participants. Given that these participants' average QQI scores are as high as the central cluster's average QQI score (83.61) in the recommender system group, we can consider these two participants' messages as buried treasures due to information overload in the control group's knowledge building discourse.

Finally, in both groups, we found that participants who were in the core of a community posted messages with higher QQI scores than peripheral members. Extending Gunawardena et al. (2016), who demonstrated that social presence in a learning community is not associated with the categories of the IAM, our findings underscore that writing highly readable messages with topic-related keywords to fill understanding gaps acts as a catalyst to foster information brokers in a community.

6.5 Limitations

Our study has several limitations, which can be addressed by future research. First, we reported findings from 281 messages because coding and counting messages with respect to IAM categories is demanding in time and effort. While 281 is a reasonably good sample size, it would be better to have more messages. Second, participation in the online discussion was required for all students to stimulate a vibrant interchange of ideas. Interestingly and importantly, these limitations make medium and large effects reported in our tables even more impressive. Future research can enrich our contribution by assessing students' voluntary use of an AOD over a longer duration. Third, two domain experts validated all topic-related keywords used by the QQI algorithm. This manual process can be automated (Zhu et al. 2014). Fourth, the overlaps among recommender system group students' 145 ratings ($M = 4.14$, $SD = 0.49$) for 149 messages ($M = 4.26$, $SD = 0.44$) suggest that the recommender system did not suffer from the data sparsity problem. Although this supposition needs to be tested more thoroughly, it is consistent with Pan and Wu (2020)'s finding that collaborative-filtering performs well when a rating matrix's data density is around 20%. In future work, we

plan to conduct an in-depth investigation to identify what topic-related keywords students use, do not use, and what alternatives they find are truly effective in their explanations. Through the design science research methodology, we will employ these findings to convert our collaborative filtering recommender system into a hybrid one.

7 Conclusion

Our findings make important contributions to research and practice in CSCL. The fact that students can use recommendations intentionally to engage in co-creating knowledge relevant to themselves and their community extends the literature on learning analytics to foster group cognition (Borge and Mercier 2019). Furthermore, this study showed how to go beyond IAM's message categorization for a better understanding of the collaborative knowledge construction process. At a higher level, our proposed and employed analytical approach overcomes the limitations of the past research that investigated learning analytics tools predominantly from a quantitative standpoint and neglected the qualitative analysis of messages based on systematic coding schemas (Hernández-Lara et al. 2019). This evidence suggest that the QQI algorithm can be used in other CSCL environments, and as a supplement to face-to-face collaboration. Furthermore, our findings extend the literature that examines recommender systems integrated into AOD tools via information retrieval measures (e.g., Damiani et al. 2015; George and Lal 2019b; Najafabadi et al. 2017; Zheng et al. 2015).

In practice, findings demonstrate to instructors who build their pedagogy around knowledge building discourse the value of forming cohesive subgroups based on mutual interests when there is an already existing and established stock of knowledge to be improved or revised. As this study showed, these subgroups can enhance collaboration opportunities in a shared knowledge space. For CSCL system developers, our findings showed how recommendations affect discussion thread development. This contribution can serve as base for CSCL system developers to organize or group similar posts or inquires together because recommendations can influence which posts the students read, in what order, and where they make replies. We hope that our theoretical lens and empirical findings will set the stage for the future research aimed at understanding and optimizing learning and the environments in which learning occurs.

References

- Borge, M., & Mercier, E. (2019). Towards a micro-ecological approach to CSCL. *International Journal of Computer-Supported Collaborative Learning*, 1-17.
- Chen, C. M., Li, M. C., & Chen, T. C. (2020). A web-based collaborative reading annotation system with gamification mechanisms to improve reading performance. *Computers & Education*, 144, 103697.
- Chen, B., Scardamalia, M., & Bereiter, C. (2015). Advancing knowledge-building discourse through judgments of promising ideas. *International Journal of Computer-Supported Collaborative Learning*, 10(4), 345–366.
- Cho, J. (2008). Issues and challenges of agile software development with SCRUM. *Issues in Information Systems*, 9(2), 188–195.
- Clarà, M., & Mauri, T. (2010). Toward a dialectic relation between the results in CSCL: Three critical methodological aspects of content analysis schemes. *International Journal of Computer-Supported Collaborative Learning*, 5(1), 117–136.

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale: Erlbaum.
- Damiani, E., Ceravolo, P., Frati, F., Bellandi, V., Maier, R., Seeber, I., & Waldhart, G. (2015). Applying recommender systems in collaboration environments. *Computers in Human Behavior*, 51, 1124–1133.
- Dascalu, M. I., Bodea, C. N., Moldoveanu, A., Mohora, A., Lytras, M., & de Pablos, P. O. (2015). A recommender agent based on learning styles for better virtual collaborative learning experiences. *Computers in Human Behavior*, 45, 243–253.
- Dawson, S. (2008). A study of the relationship between student social networks and sense of community. *Journal of Educational Technology & Society*, 11(3), 224–238.
- Eryilmaz, E., Thoms, B., Ahmed, Z., & Lee, K. H. (2019). Affordances of recommender Systems for Disorientation in large online conversations. *Journal of Computer Information Systems*, 1–11.
- Eryilmaz, E., Thoms, B., & Canelon, J. (2018). How design science research helps improve learning efficiency in online conversations. *Communications of the Association for Information Systems*, 42(1), 21.
- Freeman, L. (2004). *The development of social network analysis*. Vancouver: Empirical Press.
- Gašević, D., Joksimović, S., Eagan, B. R., & Shaffer, D. W. (2019). SENS: Network analytics to combine social and cognitive perspectives of collaborative learning. *Computers in Human Behavior*, 92, 562–577.
- George, G., & Lal, A. M. (2019a). Review of ontology-based recommender systems in e-learning. *Computers & Education*, 142, 103642.
- Gunawardena, C. N., Flor, N. V., Gómez, D., & Sánchez, D. (2016). Analyzing social construction of knowledge online by employing interaction analysis, learning analytics, and social network analysis. *Quarterly Review of Distance Education*, 17(3), 35.
- Gunawardena, C. N., Lowe, C. A., & Anderson, T. (1997). Analysis of a global online debate and the development of an interaction analysis model for examining social construction of knowledge in computer conferencing. *Journal of Educational Computing Research*, 17(4), 397–431.
- Haythornthwaite, C. (2002). Building social networks via computer networks: Creating and sustaining distributed learning communities. *Building virtual communities: Learning and change in cyberspace*, 159–190.
- Hernández-Lara, A. B., Perera-Lluna, A., & Serradell-López, E. (2019). Applying learning analytics to students' interaction in business simulation games. The usefulness of learning analytics to know what students really learn. *Computers in Human Behavior*, 92, 600–612.
- Hod, Y., Bielaczyc, K., & Ben-Zvi, D. (2018). Revisiting learning communities: Innovations in theory and practice. *Instructional Science*, 46(4), 489–506.
- Howell, J. A., Roberts, L. D., & Mancini, V. O. (2018). Learning analytics messages: Impact of grade, sender, comparative information and message style on student affect and academic resilience. *Computers in Human Behavior*, 89, 8–15.
- George, G., & Lal, A. M. (2019b). Review of ontology-based recommender systems in e-learning. *Computers & Education*, 142, 103642.
- Klašnja-Miličević, A., Vesin, B., & Ivanović, M. (2018). Social tagging strategy for enhancing e-learning experience. *Computers & Education*, 118, 166–181.
- Matuk, C., & Linn, M. C. (2018). Why and how do middle school students exchange ideas during science inquiry? *International Journal of Computer-Supported Collaborative Learning*, 13(3), 263–299.
- Mercer, N. (2000). *Words and minds: How we use language to think together*. New York, NY: Routledge.
- Najafabadi, M. K., Mahrin, M. N. R., Chuprat, S., & Sarkan, H. M. (2017). Improving the accuracy of collaborative filtering recommendations using clustering and association rules mining on implicit data. *Computers in Human Behavior*, 67, 113–128.
- Núñez-Valdez, E. R., Lovelle, J. M. C., Hernández, G. I., Fuente, A. J., & Labra-Gayo, J. E. (2015). Creating recommendations on electronic books: A collaborative learning implicit approach. *Computers in Human Behavior*, 51, 1320–1330.
- Pan, Y., & Wu, D. (2020). A novel recommendation model for online-to-offline service based on the customer network and service location. *Journal of Management Information Systems*, 37(2), 563–593.
- Procaci, T. B., Siqueira, S. W. M., Nunes, B. P., & Nurmikko-Fuller, T. (2019). Experts and likely to be closed discussions in question and answer communities: An analytical overview. *Computers in Human Behavior*, 92, 519–535.
- Qiu, M. (2019). Fostering both extensive and intensive threaded discourse—Discussion thread development and class size. *Computers & Education*, 139, 81–101.
- Qiu, M., Hewitt, J., & Brett, C. (2014). Influence of group configuration on online discourse writing. *Computers & Education*, 71, 289–302.
- Reynolds, B. L., & Wang, S. L. (2014). An investigation of the role of article commendation and criticism in Taiwanese university students' heavy BBS usage. *Computers & Education*, 78, 210–226.

- Rizvi, S., Rienties, B., & Khoja, S. A. (2019). The role of demographics in online learning; a decision tree based approach. *Computers & Education*, 137, 32–47.
- Santos, O. C., Boticario, J. G., & Pérez-Marín, D. (2014). Extending web-based educational systems with personalised support through user Centred designed recommendations along the e-learning life cycle. *Science of Computer Programming*, 88, 92–109.
- Shu, H., & Gu, X. (2018). Determining the differences between online and face-to-face student–group interactions in a blended learning course. *The Internet and Higher Education*, 39, 13–21.
- Seifert, T. L. (1995). Characteristics of ego-and task-oriented students: A comparison of two methodologies. *British Journal of Educational Psychology*, 65(1), 125–138.
- Stahl, G. (2013). *Theories of collaborative cognition: Foundations for CSCL and CSCW together*, In *Computer-supported collaborative learning at the workplace* (pp. 43–63). Boston: Springer.
- Thoms, B., Eryilmaz, E., Dubin, N., Hernandez, R., & Colon-Cerezo, S. (2020). Real-time visualization to improve quality in computer mediated communication. In *Web Intelligence* (no. preprint, pp. 1–13). IOS press.
- SoLAR. (2011). Open learning analytics: An Integrated & Modularized Platform. WhitePaper. Society for Learning Analytics Research. Retrieved from <https://solaresearch.org/wp-content/uploads/2011/12/OpenLearningAnalytics.pdf>
- van Heijst, H., de Jong, F. P., van Aalst, J., de Hoog, N., & Kirschner, P. A. (2019). Socio-cognitive openness in online knowledge building discourse: Does openness keep conversations going?. *International Journal of Computer-Supported Collaborative Learning*, 1–20.
- Zheng, X. L., Chen, C. C., Hung, J. L., He, W., Hong, F. X., & Lin, Z. (2015). A hybrid trust-based recommender system for online communities of practice. *IEEE Transactions on Learning Technologies*, 8(4), 345–356.
- Zhu, D., Fukazawa, Y., Karapetsas, E., & Ota, J. (2014). Activity-based topic discovery. *Web Intelligence and Agent Systems: An International Journal*, 12(2), 193–209.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Evren Eryilmaz¹ · Brian Thoms² · Zafor Ahmed³ · Kuo-Hao Lee⁴

¹ California State University Sacramento, Sacramento, CA 95819, USA

² California State University Channel Islands, Camarillo, CA 93012, USA

³ University of Houston-Downtown, Houston, TX 77002, USA

⁴ Bloomsburg University, Bloomsburg, PA 17815, USA