Original article

doi: 10.1111/jcal.12168

Journal of Computer Assisted Learning

Three interaction patterns on asynchronous online discussion behaviours: A methodological comparison

I. Jo,* Y. Park† & H. Lee‡

*Institute for Teaching and Learning, Honam University, South Korea
†Department of Educational Technology, Ewha Womans University, South Korea
‡Department of Human Resource Development, Ticket Monster Inc., South Korea

Abstract

An asynchronous online discussion (AOD) is one format of instructional methods that facilitate student-centered learning. In the wealth of AOD research, this study evaluated how students' behavior on AOD influences their academic outcomes. This case study compared the differential analytic methods including web log mining, social network analysis and content analysis which were selected by three interaction patterns: person to system (P2S), person to person (P2P) and person to content (P2C) interaction. Forty-three undergraduate students participated in an online discussion forum for 12 weeks. Multiple regression analyses with the predictor variables from P2S, P2P and P2C and with a criterion variable of a final grade indicated several interesting findings. For P2S analysis, visits on board (VOB) had a significant variable to predict final grades. Also, the result of P2P analysis proved that in-degree and out-degree centrality predicted final grades. The P2C results based on cognitive presence represent that students' messages were mostly affiliated to the exploration and integration levels and also predicted the final grades. This study ultimately demonstrated the effectiveness of using multiple analytic methodologies to address and facilitate students' participation at AOD.

Keywords asynchronous online discussion, content analysis, data-mining, social network analysis (SNA).

Introduction

Within the paradigm of social constructivism, student-centered learning frameworks and practices have been widely adopted and applied into the current education field. In contrast to traditional instructor-led learning, a student-centered approach emphasizes facilitating learners to negotiate multiple perspectives, reconcile conflicting ideas and construct new knowledge (Land, Hannafin, & Oliver, 2012). Instructional methods such as problem-based learning, computer-supported collaborative learning, project-based learning and discussion-based learning are the exampler format. These methods especially with the integration of diverse online environments and information technologies, support positive effects on not only knowledge acquisition (Weinberger & Fischer, 2006) but also an increase in communication, social and collaboration skills, and learning ability, such as metacognitive skill, time management and self-regulation, which can be termed as *hidden curriculum* in distance education (Anderson, 2001; Garrison & Arbaugh, 2007). However, such successful results depend highly on students' active participation and the instructor's proper facilitation, which has been recognized as a challenge (Goodyear, Jones, & Thomson, 2014). Issues, such as lack of participation, time spent resulting from meaningless argumentation,

Accepted: 31 October 2016
Correspondence: Yeonjeong Park, Institute for Teaching and Learning, Honam University, 417 Eodeung-daero, Gwangsan-gu, Gwangju, 62399, South Korea. Email: ypark@honam.ac.kr

feeling of disconnections and high demands of instructors' managerial roles, have emerged as the drawback of constructive instructional methods in the real world (Branon & Essex, 2001).

Asynchronous online discussion (AOD), a major focus of this study, is not the exception. Rather, because its inception of online learning, AOD has been recognized as the most common and widely accepted instructional method. As a long-standing and dominant research strand in the area of online learning and education (Loncar, Barrett, & Liu, 2014), a series of studies on AOD have actively discussed the benefits and limitations, success factors and guidance for effective discussion (Andresen, 2009). While AOD promised pedagogical meaningfulness that includes not only aforementioned advantages but also critical thinking and argumentation skill (Andriessen, Baker, & Suthers, 2013) considered highly important in higher education (Garrison, Anderson, & Archer, 2001), the challenges were also investigated because such skills are only developed throughout consistent discourse practices, regular participation in discussion and deep learning (Andresen, 2009; Garrison, Anderson, & Archer, 1999; Garrison et al., 2001; Kanuka & Anderson, 1998). Other factors such as the instructor's role, discussion strategies, group size and students' peer-facilitations were regarded as important for high quality discussion in an online environment.

A phenomenological review of literature focused on a forum in AODs (Loncar et al., 2014), including 43 journal papers published from 2008 to 2012 in the *Computers & Education, Journal of Computer Assisted Learning and Australasian Journal of Educational Technology*. This review identified the following points.

- Most AOD studies carried out through or on learning management systems (LMSs) in a university setting.
- Unmediated, uncontrolled or un-facilitated forum or AOD use *by students* will likely not result in effective discussion, learning or knowledge construction.
- Instructors need to plan and control the discussion carefully with specific definitions of higher learning as well as elaborated scaffolds, models, protocols and assignment parameters.

In addition to the conclusive messages from the AOD literature, divergent research trends have been also systematically organized. The four categories include: 1) argumentative impulse focusing on models that use a forum to encourage critical thinking, 2) comparative impulse focusing on comparing different forum environments as well as discussion strategies, protocols, and methods, 3) relational impulse focusing on relational aspects of online discussion and 4) analytical impulse focusing on analyzing online interaction. Using this categorization scheme, our study is specifically interested in *analytical impulse*. The studies that fall into the category have employed diverse analysis methods including a *quantitative* approach for measuring students' participation, and a *qualitative* approach for analyzing content in an actual forum or AOD discussion.

Zheng, Yang, and Huang (2012) briefly compared diverse methods, including conversational analysis, social network analysis, content analysis and sequential analysis to analyse interactions in a collaborative learning environment. Among these methods, while content analysis (CA) and social network analysis (SNA) are most often used method, the previous studies have suggested utilizing *multiple methods rather than a single method or analytic tool* (de Laat, Lally, Lipponen, & Simons, 2007; Kim & Lee, 2012). When considering the benefits and limitations of each method, integration of diverse methods became another research trend.

On the other hand, because of the difficulties and time-consuming manual process of analyzing thousands of messages in a forum or discussion board qualitatively, mining the LMS log-data, such as the number of postings, structures of discussion threads and the relationships between students, has been recognized as efficient in the area of educational data mining (EDM) and learning analytics (LA). Furthermore, the approach of LA goes beyond collecting and measuring data toward providing early warning or pedagogical interventions so that students and instructors are able to notice learning progress and to predict learning outcomes at an early stage. In particular, SNA techniques have been welcomed because of their benefits indicating key or peripheral contributors and visualizing discussion interaction 'in a birds' eye view' (Bakharia & Dawson, 2011; Dawson, Bakharia, & Heathcote, 2010) and assessing students' participation in an e-learning environment (Rabbany, Takaffoli, & Zaïane, 2011).

This study takes a standpoint highlighting the integration of multiple methods to assess the quality of AOD. In spite of extensive previous studies on AOD, the remaining important task is to compare the major analytical methodologies. For this, we considered three modes of interactions that occurred in AOD: Person to system (P2S) interaction, person to person (P2P) interactions and person to content (P2C) interaction analysis. The purpose of this study is to assess how students' behavior on AOD influence the academic outcomes. As a case study on AOD, this study analysed 1) individual effects of each of the three modes of interaction on the final grade, and 2) collective effects of the three modes of learner interactions on the final grade. The specific research questions (RQs) are listed below:

RQ1: To what extent does the quantity of students' weblog behavior in AOD predict academic achievement?
RQ2: To what extent does the degree centrality derived from person to person social network analysis predict academic achievement?
RQ3: To what extent does the quality of discussion content predict academic achievement?
RQ4: Does integration of multiple methods predict academic achievement in a higher manner than a single method approach?

Three interaction modes and analytic methods in AOD

This study reconceptualized possible interaction patterns that online learners may experience in an AOD environment within three modes: person to content (P2C), person to person (P2P) and person to system (P2S). This section briefly introduces the concept of each interaction pattern; conventionally these patterns have been described by content analysis, social network analysis and log-data analysis.

Person to content (P2C) interaction

The P2C interaction conducted by content analysis was first analysed more than 200 years ago to analyse textual material from newspaper and magazine articles, advertisements, political speeches, hymns, folktales and riddles (Harwood & Garry, 2003). Many subsequent studies have evaluated the content of data. It has been used as a technique in both qualitative and quantitative studies (Collis & Hussey, 2013; Neundorf, 2002; Robson, 1993) and is usually described as being qualitative in the development stages of research and quantitative where it is applied to determine the frequency of phenomena of interest (Elo & Kyngas, 2007). This enables the identification of data units, which are categorized, recorded, compared and contrasted to reach a conclusion about the content of the communication (Collis & Hussey, 2013; Harris, 1996).

P2C analysis has been recognized as effective when researchers aimed to characterize the type and level of online interaction (Kim & Lee, 2012) and to report very detailed analytical data (de Laat et al., 2007). The major benefit is its unobtrusive, unstructured, context sensitive nature (Krippendorff, 1980). It applies to the research settings not only having flexibility in an analytical practice but also not interrupting its original intent or meaning. On the other hand, there have been several limitations. First, it is a technique that only focuses on analyzing the content. Because P2C does not provide dynamics of students' participation and their interactions (Zheng et al., 2012) other analytic tools need to be combined to analyse discussion activities from various angles. Second, there is a reliability and validity problem between human coders. Although this problem can be improved through diverse efforts, P2C interaction is inevitably biased when investigating the messages qualitatively. Therefore, researchers need a certain framework in order to analyse or categorize substantial messages. Exemplar frameworks developed by several leading scholars (Gunawardena, Lowe, & Anderson, 1997; Henri, 1992; Zhu, 2006) have actively applied in analyzing communications in online learning.

Among the diverse frameworks for online learning, the cognitive presence (CP) of learners rather than the social aspects in messages has been considered as a useful concept to analyse written content containing the authors' critical thoughts and reflecting their higher-order thinking. CP assesses the cognitive level and process of individual participants (Garrison et al., 2001; Garrison & Cleveland-Innes, 2005). In other words, CP measures the extent to which discussion participants are able to construct and confirm meaning through sustained reflection and discourse in a critical community of inquiry (Garrison et al., 1999). Remarkably, CP reflects higher-order knowledge acquisition and application and is associated with research related to critical thinking (Garrison et al., 2001). Because this study aims to understand the meaningful context in AOD, we utilized CP as an appropriate framework for P2C analysis. As shown in Table 1, phases of CP are in hierarchical order, which

Table 1. Phases of Cognitive Presence with Indicators

Levels	Phase of cognitive presence	Indicators
1	Triggering	(1) Recognizing the problem (2) Sense of confusion (3) Take discussion in new direction
2	Exploration	(4) Divergence within the online community (5) Divergence within a single message (6) Information exchange (7) Suggestions for consideration (8) Brainstorming (9) Leaps to conclusions
3	Integration	(10) Convergence among group members (11) Add more idea (12) Convergence within a single message (13) Connecting ideas, synthesis (14) Creating solutions
4	Resolution	(15) Vicarious application to real world (16) Testing solutions (17) Defending solutions

* Adapted from Garrison et al. (2001), American Journal of Distance Education 15(1), 7–23.

indicates that the higher phase is in a more advanced level of cognitive thinking (Garrison et al., 1999, 2001). Therefore, we considered the phases in ordinal manner of the thinking process, but each phase does not represent hierarchies of individual indicators.

Person to person (P2P) interaction

Analysing P2P interaction is to find out dynamic communication among participants in online learning. SNA is a well-known method to analyse phenomena with quantifiable data, and it enables researchers to examine the relationships among actors from a macro perspective (Cho, Lee, Stefanone, & Gay, 2005; Enriquez, 2008; Jun, 2004). In the teaching and learning context, it has been recognized as a useful method to assess students' knowledge construction (Dawson et al., 2010; Enriquez, 2008) and to observe their interaction patterns on the discussion boards (de Laat et al., 2007; Palazuelos, García-Saiz, & Zorrilla, 2013), which have turned out to be one of the most influential factors affecting learning outcomes on computer-supported collaborative learning (Liccardi et al., 2007; Vrasidas & McIsaac, 1999).

The unit of analysis for P2P interaction is a 'link' generated as an interaction occurs between two persons. Complexity of the links in a network is measured by degree centrality, in-degree, out-degree centrality, betweenness centrality and closeness centrality. Also, the graphics that are called a sociogram enable us to visualize the network patterns at a glance. Here, we review the levels of P2P analysis, visualization effect and prediction from the use of P2P variables.

Two levels of P2P analysis

P2P analysis involves the individual level and group level. The individual level analyses the individual position/network-attribute embedded in the whole network. Group level analysis provides a measure of cohesion/dispersion based on the overall network. The key indicator of individual level is centrality. *Centrality* measures the behavior of individual nodes within a whole network, and indicates the degree to which individual nodes interact with other participants in the network (Wasserman & Faust, 1997). It provides researchers with information about who is a central participant of a network. Based on a link between nodes, in-degree, out-degree and betweenness degree measures can be generated. *In-degree centrality* is a form of centrality that represents the total number of links that a central participant has received from other group members (Tsai, 2001). *Out-degree centrality* counts the number of links a person sent to other nodes of the network. Furthermore, *betweenness centrality* is calculated as the number of shortest paths that have to pass to reach a given node. As another measurement, *prestige* is a more sophisticated network variable, which gives information about who is the recipient of extensive links. Both centrality and prestige measure the prominent factor in a given network (Aggarwal, 2011).

At the *group* level, density and centralization are reported as key indicators of the network that affect group learning. *Density* describes the general level of cohesion/interaction among the actors in a network. It is defined as the number of links between nodes in a network divided by the maximum number of possible links (Scott, 1991). This varies between 0 and 100%, as the number of links varies between 0 and equal the maximum number of possible links. The more nodes connected to one another, the density of the network will be higher (Borgatti, Everett, & Freeman, 2002; Scott, 1991). *Centralization* describes how tightly interaction within a network is organized around its most focal points. Thus, a high level of centralization means that the network is formed around a particular focal participant.

Use of P2P analysis in AOD

One merit of SNA is network visualization (sociograms), which allow an instructor to find who is a central or outlying actor that can reflect student participation and decide whether interventions are needed. Instructors can identify student's participation by using individual measures such as centrality and prestige, and can also assess the quality of AOD by using network measures, such as density and centralization. Successful online forum requires high participation (Webb, Jones, Barker, & van Schaik, 2004), high level of interconnectedness between learners (Dron & Anderson, 2007; Zhu, 2006), participation of various learners rather than a few star participant (Dawson et al., 2010), student-centered rather than instructor-facilitated interaction (Light, Nesbitt, Light, & White, 2000; Nickel, 2002; Schrire, 2004) and diverse exchanges among participants (Dawson et al., 2010). If we transfer these conditions to network characteristics, the discussion network will be depicted as a higher rate of active participants and few or no outliers, high level of density (as density describes the general level of cohesion/interaction among the actors in a network), low level of centralization (as centralization assesses whether a network is organized around its most central points), student centered community rather than a facilitator centric pattern or 'wagon-wheel pedagogy', which is shown as an interconnected shape rather than a star shape, and diverse participation regardless of discussion topics, that can be analysed by two-mode network analysis.

Predicting by P2P variables

Predicting students' learning performance by using interaction variables is one of the oldest and most useful applications. Numerous studies have attempted to estimate the unknown value of students' learning performance, participation and interaction since the introduction of SNA. A few previous studies have examined the direct effect of social network variables on individual performance. Baldwin, Bedell, and Johnson (1997) revealed that network centrality among MBA students significantly influenced students' performance, team outcomes and satisfaction with the learning program. Cho, Gay, Davidson, and Ingraffea (2005) empirically investigated the relationship between communication style, social network and performance in an online learning environment. They suggested that students' communication styles and pre-existing friendship networks significantly influenced social network variables, which significantly affected students' performance.

Recently, the introduction of online communication resources, such as chat logs, discussion forums, blog posts and comments, a core component of distance learning supplementing face-to-face teaching practices, enables social network researchers to easily access detailed event logs listing all online peer interactions (Dawson et al., 2010). This opportunity allows SNA researchers to investigate students' communication with unpolluted and actual data based on their real behavior. Heo, Lim, and Kim (2010) investigated the patterns of online interaction among participants analysed by SNA with logs recorded on a discussion forum. The study revealed that each team has a different frequency of read and response networks and unique patterns of interaction between learners. Another study confirmed that the high interaction among students affects communication, mutual support and group cohesion (Guzzo & Dickson, 1996). However, based on the unexpected result that the group with the highest interaction received the lowest score, previous studies suggested that combining content analysis with SNA is required. Tirado, Hernando, and Aguaded (2012) investigated the relationship between network structure and knowledge construction in AOD, and also demonstrated the potential of a mixed method, the combination of content analysis and SNA, to enable a better understanding of knowledge construction in AOD.

The mixed method proposed by researchers is not limited to combining P2C and P2P. There are various factors or characteristics that can affect student learning performance: demographic, socio-economic status, educational, cultural, high school GPA, SAT scores and the interaction between student and faculty (Araque, Roldan, & Salguero, 2009; Campbell & Oblinger, 2007). As a result, different techniques have been applied to predict student learning. Romero, Lopez, Luna, and Ventura (2013) suggested the combined use of quantitative, qualitative and network information for predicting students' performance in a course by application of classification algorithms and classification via clustering algorithms. They found the most powerful set of variables to predict final students' performance: two quantitative variables (the number of sending message and the number of written words), together with qualitative variable (the average score obtain in messages) and the two social network variables (the degree centrality and prestige).

Person to system (P2S) interaction

Person to system (P2S) interaction analysis is a means of understanding a behavioral pattern of learners within the online system and interaction with it. P2S analysis follows the same key context with log-data analysis. P2S analysis focuses on the treatment to enhance students' academic achievement by presenting useful information extracted from large data sets (Elias, 2011; Jo, Yoon, & Ha, 2013). This large data set saved as unstructured data is called log-data. The unit of analysis of this analytic method is each 'log' activity. Based on this log-data from a P2S analysis, dozens of variables can be derived and created by researchers in accordance with research problems (Jo, Yu, Lee, & Kim, 2014b).

Total log-in time and time spent on board

Time investment is fundamental for a good quality learning activity because learners need a certain amount of time to construct and arrange their knowledge (Beaudoin, 2002). Mason (2011) accentuated that if learners have enough time to be absorbed in learning in order to become familiar with subjects, a meaningful discussion would be possible. In prior study (Jo, Kim, & Yoon, 2014a), time investment was calculated as a total log-in time within the learning management system (LMS), obtained by adding up each time spent per visit. Likewise, time spent on board was suggested on the basis of an assumption that students who spend more time within the discussion board per access can be regarded as an active learner participating deeply in discussion. This includes time spent both on reading others postings and writing their own words (Wise, Speer, Marbouti, & Hsiao, 2013) and was calculated using the same procedure of total log-in time but focused on the duration of the discussion board only.

Total log-in frequency and visits to the board

The number of log-ins into the LMS indicated an active participation in an online learning environment (Lara, Lizcano, Martínez, Pazos, & Riera, 2014). The positive correlations between active participation and the attendance rate with regards to the online discussion forum have also been interpreted in the online learning environment (Webb et al., 2004). This variable was calculated using the total number of visits to the LMS. The visit to the board is similar to the total log-in frequency.

Log-in regularity and regularity in discussion

Students can improve their learning experience through regularly participating in discussions (Andresen, 2009), and this inclination is closely related to self-regulated learning, especially time management. 'Connected' efforts to intensively and actively participate in discussion are critical when constructing, transferring and applying knowledge (Mason, 2011). Log-in regularity means that the average log-in interval with regards to the LMS and regularity in discussion represents how students regularly access the discussion board.

Number of postings

The participation in the discussion forum is a significant condition for productive discussion (Wise et al., 2013). The number of postings determines the interest in a particular topic and is an important indicator of students' active participation in discussion because this variable is closely related to students' performance (Hung & Zhang, 2008; Topcu & Ubuz, 2008; Webb et al., 2004). In other words, this shows those who uploaded more postings than others can be considered as an active learner. This variable was calculated by counting the number of postings.

Average word count
The average length of writing can be connected to the idea that students who invest enough time in constructing their own thought will write more to concretely state certain opinions (Hewitt, Brett, & Peters, 2007). This variable was calculated by counting the total number of words used for all postings a student made, divided by each student's total number of postings in a discussion board. The strength of using P2S analysis is that online behavior patterns of students are based on the objective, accurate data that is free from responder bias. Because P2S data that comes from an online environment are recorded simultaneously with students' online activities, less errors occur while collecting and calculating data. In addition, using P2S data is superior in time and cost-efficiency because it is relatively convenient to process big amounts of data. On the other hand, in terms of understanding the depth of discussion content, P2S analysis has shown significant limitations.

Methodology

Research background

The participants of this study were 43 students and one instructor who were participating in the blended learning course entitled 'Administration, Legislation, and Politics' during the second semester in 2013 in a university of South Korea. This class was selected for several reasons. The course is part of a three-credit core course for students majoring in public administration, and was a representative formal course amenable to design as online-based discussion learning. Second, the instructor incorporated a unique approach to facilitate student's discussion activity. Consequently, students' participation was extensive, which allowed us to analyse the diverse interaction types of P2P, P2S and P2C. Finally, the individual evaluation in this course was measured quantitatively and qualitatively in accordance with the course objectives. Fifteen per cent of the final grade was assigned to online discussion activities, and the essay writing was 20% of the final grades. Mid-term and final-exam scores contributed 30% of the final score, respectively, and the participation score was only 5%.

Instructional strategies for AOD

The online forum activity in this case study had some unique instructional strategies. First, the instructor uploaded sample essays that could facilitate online discussions among students about their essay assignments in each week's topic in the forum. Second, greetings between students, simple expressions on agreements or disagreements were not allowed by the instructor. Students had to write an essay focusing only on the discussion topic of the week. In addition, the learners were encouraged to write an essay of more than 60 words, and if the writing was less than the 60 words, these messages were not included for discussion in the forums. The course had some strict rules for online discussion activities. The final distinctive instructional strategy was that students could freely choose discussion topics of their own from 12 different discussion forums. The participation completely depends on each student's desire to participate. Students can choose to write on the topics as many as they want to, or not to write any comment at all.

Research procedures

P2S analysis
The entire log-data was collected by means of the Moodle-based Lab Management System (LMS), and it was computed by an automatic data collection module embedded in the LMS. Basic content analysis was first performed to recognize what the major discussion topics were, and to what extent the students participated in each discussion topic. Next, researchers arranged plenty of data containing each variable's information in one large table to implement P2S analysis.

Correlational analysis for selecting what variables to analyse for P2S was proceeded with an SPSS statistic package. Eight input variables were processed: total log-in time (TLT), time spent on board (TSB), total log-in frequency (TLF), visits on board (VOB), log-in regularity (LR), regularity in discussion (RID), number of postings (NOP) and average word count (AWC). TLT and TSB had a high correlation ($r = .961$), as did TLF and VOB ($r = .705$) and LR and RID ($r = .913$). Because this research focused on discussion activity in particular, TSB instead of TLT, VOB instead of TLF, and RID instead of LR were selected. Those were the variables calculated in the discussion-board-specific environment. The NOP variable in P2S had almost the same meaning as the out-degree centrality variable in P2P analysis because out-degree centrality was calculated based on the number of postings. Correlation analysis supports such fact (r = .997). Therefore, NOP variable was deleted and only out-degree centralities for P2P analysis were used. In addition, AWC was also not included for the final analysis because the target course required a minimum number of words as a strict grading policy, which was regarded that AWC would not reflect natural variations. Finally, three variables were selected for P2S analysis: VOB, TSB and RID.

P2P analysis

Participants' online log-data were extracted from the LMS as done with P2S analysis. These log-data included the 'time' when the messages were left, the 'person' who wrote the messages and the 'link' where we can see the actual discussion forum stages on Cyber Campus with temporary identification in order to observe the personal information protection principle. The UCINET and NetDraw program were employed to analyse P2P patterns. UCINET is a software for the analysis of social network data, and it comes with the NetDraw. NetDraw is for network visualization. For this process, some matrices which include interaction data were produced by a researcher with a background in educational technology. With these matrices, degree, out-degree, in-degree centrality, density and sociogram patterns of participants were derived.

P2C analysis

For the P2C, three researchers coded the message data. The inter-rater reliability by Cohen's Kappa test was .561. The discussion forum left a total of 1373 messages. However, the total included meaningless messages, so some messages were deleted in the P2C process. A total of 1273 messages were analysed and coded based on Garrison's cognitive presence frame as shown in Table 1. The four phases of cognitive presence during an online discussion included triggering, exploration, integration and resolution. As the phases proceeded, more complicated and high-quality discussion occurred. The triggering phase included discussion behaviors, such as presenting background information helpful to facilitate the discussion and asking questions to create discussion subjects. The exploration phase included behaviors, such as depicting contradictions on prior ideas, representing various ideas or themes in one message, describing personal narratives or facts related to the discussion subject, expressing thoughts roughly and offering unsupported opinions. The integration phase included behaviors, such as referring previous discussion content to support one's own opinion, adding new ideas and representing logical content to prove one's own opinion. Finally, the resolution phase included behaviors supporting a summative conclusion of discussion and creating logical solutions as a result. Furthermore, it included discussion to test the solution and defend its effect. If the content involved these contexts in a message, researchers coded the unit into each phase as an indicator of cognitive presence. For the stochastic procedures, four phases were taken as a predictor variable to predict academic achievement in an ordinal scale.

Comparison of each method

As introduced above, all data collection methodologies for each analysis proceeded in different manners. At the final stage, derived data were put in a multiple regression model with independent variables of which there were three types of analysis data and a dependent variable of the end-of-the-semester final academic achievement score. SPSS 18.0 was used to clarify the statistical relationship between research variables.

Results

Person to system (P2S) analysis

Overall analysis

In this case study, the discussion forums had a total of 1373 messages. For 12 weeks, 98 discussions proceeded in the 12 forums with different topics. Table 2 shows the major themes and quantitative characteristics of the 12 forums. Twenty nine out of the 43 students participated in each of the discussion forums. Themes of the forum were chosen and structured by the instructor's initiation. However, students were allowed to decide the number of discussion topics to participate in of their own. Consequently, the number of messages and the number of participants indicate the students' attention and interest in the discussion topics and themes of the forum.

Table 2. Themes and Quantitative Characteristics of the 12 Forums

Forum	Number of discussion topics	Number of messages	Number of participants
1	8	109	30
2	6	77	29
3	6	86	28
4	11	135	32
5	9	100	30
6	5	70	29
7	9	130	31
8	5	79	27
9	8	119	28
10	11	168	29
11	16	233	32
12	4	67	24
Total (Sum /Average)	98/8	1373/114	349/29

Descriptive statistics of log variables

Table 3 below illustrates the descriptive statistics (i.e., minimums, maximums, means and standard deviations) for the four independent variables and the dependent variable in this study. The four independent variables are the log results related to the AOD activities because these include the behavior on online discussion board. From many variables in LMS, researchers judged that these four variables were relevant to the AOD interactions. Therefore, the variables were derived in this research.

There were 537.35 student visits to the discussion board in the LMS throughout the semester (Table 3). The average time spent on the board was 79 558 s, indicating that students spent more than 22 h on the online forum board throughout the semester. However, the comparatively high value of the standard deviation suggests that board log-in and participation differ from the individual student very much. To measure the extent to which students consistently make efforts to participate in an online discussion, the regularity in discussion was calculated using a standard deviation of the login intervals. Therefore, this variable technically means the 'irregularity of access interval'; the average value was 60.08. The students' total scores were distributed between 14.11 and 70.00, and the mean was 44.54.

Person to person (P2P) analysis

Density of students' online interaction

In order to examine the changes of discussion patterns over time, the discussion period of the course in this case study was operationally divided into three stages: early (T1: 1–4 weeks), middle (T2: 5–8 weeks) and late stage (T3: 9–12 weeks). As seen in Table 4, the network patterns appreciably changed over time, with a dramatic increase in student interaction noticed by density. In an early stage, the participants' interaction rate was only 0.034 and almost half of the students did not attend to the forum. However, at the end of the stage, the density nearly increased 10-fold and the value was 0.334. This means the student interaction became more active with time.

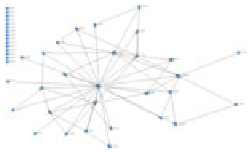
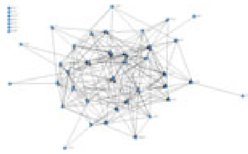
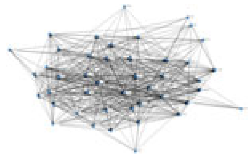
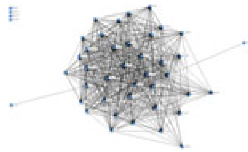
Centralization of students' online interaction

The P2P not only allows examination of overall figures of the networks through NetDraw, but also provides numeric measures of the networks, such as the degree centralization, out-degree and in-degree centralization using UCINET. Degree centralization looks at the extent to which one actor in a network holds all of the ties in that network. Degree centralization consists of two types: out-degree centralization and in-degree centralization. They are differentiated by the direction of ties.

Table 3. Descriptive Statistics of Students' Log Variables

	Minimum	Maximum	Mean	SD
PV1. Visits on board	40	1157	537.35	324.07
PV2. Time spent on board(s)	2288	344 815	79 558.02	65 405.05
PV3. Regularity in discussion[1]	20.58	280.94	60.08	45.89
CV. Total score	14.11	70.00	44.54	10.88
$n = 43$				

Table 4. Interaction Patterns of the Stages

Stage	Early (1–4 weeks)	Middle (5–8 weeks)	Late (9–12 weeks)	Total (All 12 weeks)
Socio-gram				
Density	0.034	0.116	0.334	0.401

As shown in Table 5, the degree centralization score decreased from the early to late stage. The overall degree centralization at the late stage was 8.53%, which corresponded to about one-third of the score at the early stage. High degree centralization at the early stage means that only a few students led the whole discussion interaction. However, those leading positions were spread to many other students because the discussion became more active and interactive over time. The out-degree centralization at the early stage was 20.88%, but dropped to 8.26% at the late stage. This signifies the centrality of message senders was decreased all the way down. Thus, it shows that the more students participated in sending messages to interact actively. Finally, the in-degree network centralization scores at the middle and late stages (12.35% and 11.60%) were almost double that of the early stage (5.81%). This indicated that almost everyone received messages equally at first. However, with time, some outstanding students received other students' responses intensively. In other words, those students had authority, and it created a centralization phenomenon in the social network.

Person to content (P2C) analysis

Content analysis on all discussion messages has resulted in an exploration type of message that comprised the greatest portion among all (around 60%). As shown in Table 6, integration (around 30%) followed. The two types took more than 90% of all messages, so most of the discussions were in an exploration and integration level. The third-ranked type was triggering, in which the instructor mainly suggested the discussion topic and stimulated student participation. Most of the messages in this type belonged to the instructor. However, no messages developed into the resolution phase. Only three messages reached the state of considering real world adaption. No discussion messages suggested actual resolution to test or defend. Specifically, brainstorming type of messages comprised the majority of the exploration level. A quarter of all messages illustrated the student's own opinion and sharing of it.

There are some meaningful implications in this result. First, the asynchronous discussion in online forum was quite well triggered and facilitated by the instructor. Therefore, participants interacted and explored actively. Furthermore, they could integrate their opinions or even propose a suggestion about all the diverse discuss topics. Even though these AOD activities could not arrive at a resolution level, we could verify the effect of AOD activities by seeing the results of the other three stages. Convergence among group members or discussion messages, brainstorming and exchanging information were actively progressed. Second, P2C analysis enabled the identification of the pattern and the depth of content in AOD, which would be hard with only one single method, P2P. Through P2P analysis, researchers can get the centralization and density scores but the content

Table 5. Changes of Network Centralization

Network centralization	Early (1 ~ 4 weeks)	Middle (5 ~ 8 weeks)	Late (9 ~ 12 weeks)	Total (All 12 weeks)
Degree	22.67%	12.33%	8.53%	13.57%
Out-degree	20.88%	14.73%	8.26%	16.53%
In-degree	5.81%	12.35%	11.60%	14.39%

Table 6. Frequency Results of Each Sub-Category Indicator of Cognitive Presence

Cognitive presence level	Level frequency	Sub-category indicator	Indicator frequency
Triggering	95 (7.47%)	Recognizing the problem	37 (2.91%)
		Sense of puzzlement	6 (0.47%)
		Take discussion in new direction	52 (4.08%)
Exploration	763 (59.94%)	Divergence—within the online community	45 (3.53%)
		Divergence—within a single message	112 (8.80%)
		Information exchange	133 (10.45%)
		Suggestions for consideration	107 (8.41%)
		Brainstorming	293 (23.02%)
		Leaps to conclusions	73 (5.73%)
Integration	412 (32.36%)	Convergence—among group members	73 (5.73%)
		Add more idea	94 (7.38%)
		Convergence—within a single message	101 (7.93%)
		Connecting ideas or Synthesis	44 (3.46%)
		Creating solutions	100 (7.86%)
Resolution	3 (0.24%)	Vicarious application to real world	3 (0.24%)
		Testing solutions	0 (0%)
		Defending solutions	0 (0%)
Total	1273 (100%)		1273 (100%)

that the students wrote about topics did not appear in that value. The combination of two methodologies, P2P and P2C, is complementary and enabled us to interpret and assess AOD from the perspective of both social interaction and content aspects with abundant implications.

Prediction of academic achievement

Regression result of person to system analysis (RQ 1)

Regression analysis was done to clarify whether an online learning pattern based on the log data of learners anticipates their academic achievements. Independent variables were VOB, TSB and RID. A dependent variable was the learners' final score on the course which can be interpreted as their academic achievement. Thirty-three students were analysed. The adjusted R squared value from the regression model was 0.603, which indicates that the variables explain approximately 60% of student's academic achievement. The regression result is presented in Table 7.

VOB ($t = 3.94$, $p < .05$) appeared to be a meaningful variable to explain academic achievements. TSB and RID were not valid.

Regression result of person to person analysis (RQ 2)

In this study, the centrality value of individual actors or participants was calculated based on the matrix created by the actors' communication data. However, these alternative actions (uploading students' discussion essays) were converted into the original students' actions. Using the converted data, the multiple regression result indicated that the in-degree and out-degree centrality predicted final scores with an explanation of around 68.9% ($F = 47.631$, $p = .000$) (See Table 8).

Regression result of person to content analysis (RQ 3)

To find the prediction rate for P2C variables for academic achievement, the cognitive presence value was transformed into scores. Because hierarchy only existed among levels and not among sub-indicators, weight was given to the score. Triggering was weighted as 1 point, exploration as 2 points, integration as 3 points and resolution as 4 points. Because the number of postings by individual students differed, the sum of each student's cognitive presence scores was divided into the student's number of postings. As shown in Table 9, the calculated cognitive presence score was a meaningful value in predicting academic achievement. The CP variable solely showed 33.3% of the R squared value, which suggests that the CP is one of the major variables to explain academic achievement.

Regression result of multiple methodologies (RQ 4)

To reveal the level of predicting academic achievement of representative predictors of P2S, P2P and P2C, all variables were put into a multiple regression model.

Table 7. Regression Result for Students AOD Activity and Academic Achievement

	Unstandardized coefficient		Standardized coefficient	t	Sig.
	B	Std. error	Beta		
(constant)	32.50	3.80		8.56	.000
Visits on board	.02	.01	.648	3.94*	.000
Time spent on board	.00	.00	.062	.41	.684
Regularity in discussion	−.04	.03	−.150	−1.25	.219

$n = 43$.
Dependent variable: total score
R^2(adj.R^2) = .631(.603), $F = 22.258$, $p = .000$.
$*p < .05$.

Table 8. Regression Result for Actors' Centrality and Academic Achievement

	Unstandardized coefficient		Standardized coefficient		
Model	B	Std. error	Beta	t	Sig.
(constant)	29.20	1.81		16.15	.000
In-degree centrality	.23	.05	.44	4.84*	.000
Out-degree centrality	.24	.04	.58	6.38*	.000

$n = 43$.
Dependent variable: total score.
R^2(adj.R^2) = .704 (.689), $F = 47.631$, $p = .000$.
$*p < .05$.

Table 9. Regression Result for Cognitive Presence and Academic Achievement

	Unstandardized coefficient		Standardized coefficient		
Model	B	Std. error	Beta	t	Sig.
(constant)	23.30	4.51		5.16	.000
Cognitive Presence	.38	.082	.59	4.69*	.000

$n = 43$.
Dependent variable: total score.
R^2(adj.R^2) = .349 (.333), $F = 21.957$, $p = .000$.
$*p < .05$.

Coefficient of determination appeared to be comparatively high as 74.4%, which shows that the variables are highly suitable to predict academic achievement (See Table 10).

As a result, RID($t = -2.70$, $p < .05$) in P2S, in-degree centrality ($t = 2.74$, $p < .05$) and out-degree centrality ($t = 3.74$, $p < .001$) in P2P, and the cognitive presence score ($t = 2.48$, $p < .05$) in P2C were revealed to be meaningful variables within the model. VOB and time spent on board appeared to be insignificant in predicting academic achievement.

Discussion and conclusion

The purpose of this study was to assess students' online behavior, specifically in an AOD context, and to analyse the influences on the academic outcomes. For this, we present the effectiveness of integrating multiple methodologies to analyse the quality of an AOD. The results of P2S, P2P and P2C analysis provide the following methodological implications in regard to the evaluation of AOD.

First, based on the P2S analysis, only the visit onboard variable was determined to be a significant variable for

Table 10. Regression Result for Three Methodologies and Academic Achievement

Model		Unstandardized coefficient B	Std. error	Standardized coefficient Beta	t	Significance
(constant)		30.52	3.73		8.19	.000
P2S	Visits on board	−.01	.01	−.18	−.80	.431
	Time spent on board	.00	.00	.01	.05	.957
	Regularity in discussion	−.07	.03	−.27	−2.70*	.011
P2P	In-degree centrality	.17	.06	.33	2.74*	.010
	Out-degree centrality	.24	.06	.57	3.74*	.001
P2C	Cognitive presence	.15	.06	.24	2.48*	.018

$n = 43$.
Dependent variable: total score.
R^2(adj.R^2) = .780(.744), $F = 23.310$, $p = .000$.
*$p < .05$.

students' academic achievement in AOD. However, time spent onboard and regularity in discussion were invalid as prediction variables. This is because students proceed primary work in the local computer environment rather than directly within the discussion board. MS Office or text pad is an effective tool to revise; therefore, pre-work for writing a discussion message might be proceeded on a local computer and uploaded onto a discussion board through copy-and-paste. As a result, how much time was spent or how regularly students were logged onto the discussion board appeared to be an unimportant issue to interpret.

Second, through P2P analysis, we found out students' interaction patterns, such as frequency and visualization of learners' network centralities, as meaningful indicators to predict academic achievement. Furthermore, we verified that out-degree centralization did decrease but the density level increased as time went by. These imply that more and more students participated in discussion activities and it became increasingly active over time. Unlike the early period of the semester, more students participated in the discussion in the late period.

Third, the P2C results showed that most of the discussion contents belong to the exploration and integration phase. The instructor of the target course tried to stimulate students' interest and motivation as a provoker, but this did not play a role in summarizing the discussed opinions. It is definitely not an easy process to provide resolution based on the summary of the whole discussion. This was indicated by the result that the resolution phase did not appear in any discussion message. Therefore, it is recommended that the instructor's appropriate intervention is surely needed to conclude or synthesize the discussion.

In this research, three approaches including P2S, P2P and P2C analysis were attempted to compare and to integrate. Prior to the research conclusion, Table 11 is a brief comparison of three methodologies. Each methodology had different data-collection approaches (qualitative vs. quantitative), focus, unit of analysis, indicators, limitations, affordances and involvement of manual work for interpreting the discussion activity. We do consider that the understanding of each methodology's characteristics is essential to adapting each methodology to various circumstances.

Each methodology had strengths in analyzing students' discussion patterns. First, P2S analysis supported overviewing overall discussion activity between students because it calculates how many times they visited the discussion board and how much they participated in discussion. Such data can be interpreted as a meaningful indicator for analyzing students' discussion patterns. Second, P2P analyses students' interaction more deeply through calculating their centrality and density within the discussion network based on the data traced through the computer. Because discussion activity is a collaborative process involving not a single actor but actors in a whole network, P2S is a powerful tool to analyse an AOD course. However, such quantitative approaches were limited in their ability to reflect specific content. Cognitive presence in P2C is a source for evaluating students' discussion activity, specifically looking at whether their discussion content involved a high-level cognitive process. Especially, rubrics to analyse cognitive presence play an important role to measure depth in cognitive thinking. Its analysis of messages and themes follows a qualitative approach in general, but an automated system for analysis, such

Table 11. Comparison of Different Interaction Analysis Methods

Analysis methods	Person to system (P2S) analysis	Person to person (P2P) analysis	Person to content (P2C) analysis
Data-collection approaches (qualitative / quantitative)	Quantitative	Quantitative	Qualitative/quantitative
Focus	• Students' current states• Online behavior	• Interaction patterns• Frequency of interaction • Visualization of interaction	• Psychological states• Identification of characteristics • Description of communication patterns
Unit of analysis	Each log	Link	Sentences, messages, themes
Indicators	• Visits on board• Time spent on board • Regularity in discussion, etc	• Centrality (individual)• Density (group) • Centralization (group)	• Phases of cognitive presence
Limitations	Ignoring quality of interaction and content between students	Ignoring quality of interaction and content between students	Ignoring relations between students
Affordances	Both individual and group patterns	Both individual and group patterns	Group patterns
Involvement of manual work	Few	Fair	A lot

as text-mining, meaning that this system of analysis could be used as a quantitative approach.

Manual work for each methodology is contrasting. The automated computer process provides online data for P2S analysis. On the other hand, P2P and P2C analysis involves very labor-intensive work from researchers while processing and arranging a dataset. Particularly for P2C analysis, researchers should generally read up and code all messages, and this process requires a lot of time and effort. Despite its hardship, P2C analysis is necessary for discussion activity analysis for meticulous judgment of students' discussion capacity.

To summarize, it is not adequate to use a single methodology for the analysis for AOD evaluation. This study has identified three implications. First, convergence among P2S, P2P, and P2C has a synergetic effect of interaction based on each methodology's strength on the AOD evaluation approach. Comparing each methodology's coefficient of determination and that of convergence supports the notion; the *R* squared value of P2S, P2P and P2C were 60.3%, 68.9% and 33.3%, respectively, while that of integrated methodologies in multiple regression was 74.4%, which is higher than when using an individual methodology. The result of proceeding multiple regression analysis excluding an invalid variable, that is, TSB showed a slightly higher coefficient of determination to predict academic achievement. Second, a comparison of the *R* square value of each methodology indicated that the P2P variables were the highest. This implies that simultaneous consideration of P2P analysis is a recommended way of evaluating AOD rather than solely measuring a simple participation rate through P2S analysis with a log variable. Third, P2C analysis involves the issue of labor-intensive work for a human coder or grader along with the problem of the consistency of it. P2C analysis is an effective variable to comprehend discussion context and the characteristics of the content if the weakness of the inefficiency of the calculation is overcome. Thus, additive means of substituting labor-intensive work, such as automated text-mining, should be researched and developed. Fine-worked analysis on AOD will lead to the development of an instructional strategy for it. This ultimately contributes to facilitate more active student participation and to improve their learning achievements in student-centered learning environments.

Finally, we believe that this study has a meaningfulness beyond the methodological comparison among three types of interaction. In considering the increased open online learning environment, such as massive open online courses (MOOC), where hundreds of millions of learners are leaving their posts in the forum, this study can contribute to understanding the learning process and outcomes involving MOOC's forum activity. Previous studies have indicated that forum participation is an important indicator of active learners in MOOC environment who become certified registrants or community contributors (Ho et al., 2014; Koller, Ng, Do, & Chen, 2013). Because forum participation itself does not contribute to assess the quality of learning, despite the usefulness to predict their retention, further study calls for systematic research to identify the variables that explain the quality of forum discussion activity and compare the relative importance and usefulness of discussion-related variables.

Acknowledgements

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2015S1A5B6036244).

Notes

[1]Regularity in the discussion was calculated using a standard deviation of the login intervals in the discussion board to measure the extent to which students consistently make efforts to participate in an online discussion; therefore, this variable technically means the 'irregularity of access interval in the discussion board.'

References

Aggarwal, C. C. (2011). An introduction to social network data analytics. In *Social network data analytics* (pp. 1–15). Springer US.

Araque, F., Roldan, C., & Salguero, A. (2009). Factors influencing university drop out rates. *Computer & Education*, *53*, 563–574.

Anderson, T. (2001). The hidden curriculum in distance education: an updated view. *Change: the magazine of higher learning*, *33*(6), 28–35.

Andresen, M. A. (2009). Asynchronous discussion forums: success factors, outcomes, assessments, and limitations. *Journal of Educational Technology & Society*, *12*(1), 249–257.

Andriessen, J., Baker, M., & Suthers, D. (Eds.). (2013). *Arguing to learn: confronting cognitions in computer-supported collaborative learning environments* (Vol. *1*). Springer Science & Business Media.

Bakharia, A., & Dawson, S. (2011). SNAPP: a bird's-eye view of temporal participant interaction. *Paper presented at the Proceedings of the 1st international conference on learning analytics and knowledge.*

Baldwin, T. T., Bedell, M. D., & Johnson, J. L. (1997). The social fabric of a team-based M.B.A. program: Network effects on student satisfaction and performance. *Academy of Management Journal*, *40*(6), 1369–1397.

Beaudoin, M. F. (2002). Learning or lurking? Tracking the "invisible" online student. *The Internet and Higher Education*, *5*, 147–155.

Borgatti, S. P., Everett, M.G. & Freeman, L. C. (2002). *Ucinet for Windows: Software for Social Network Analysis*. Harvard, MA: Analytic Technologies.

Branon, R. F., & Essex, C. (2001). Synchronous and asynchronous communication tools in distance education. *TechTrends*, *45*(1), 36–36.

Campbell, J., & Oblinger, D. (2007). *Academic analytics*. Washington, DC: EDUCAUSE Center for Applied Research.

Cho, H., Gay, G., Davidson, B., & Ingraffea, A. (2005). Social networks, communication styles, and learning performance in a CSCL community. *Computers & Education*, *49*(2), 309–329.

Cho, H., Lee, J. S., Stefanone, M., & Gay, G. (2005). Development of computer-supported collaborative social networks in a distributed learning community. *Behaviour & Information Technology*, *24*(6), 435–447.

Collis, J., & Hussey, R. (2013). *Business research: A practical guide for undergraduate and postgraduate students*. New York: Palgrave macmillan.

Dawson, S., Bakharia, A., & Heathcote, E. (2010). SNAPP: Realising the affordances of real-time SNA within networked learning environments. *Proceedings of the 7th International Conference on Networked Learning*, 125–133.

de Laat, M., Lally, V., Lipponen, L., & Simons, R.-J. (2007). Investigating patterns of interaction in networked learning and computer-supported collaborative learning: A role for Social Network Analysis. *International Journal of Computer-Supported Collaborative Learning*, *2*, 87–103.

Dron, J., & Anderson, T. (2007). Collectives, networks and groups in social software for e-learning. In *Proceedings of world conference on e-learning in corporate, government, healthcare, and higher education* (pp. 2460–2467). Chesapeake, VA: AACE.

Elias, T. (2011). Learning analytics: definitions, processes and potential. *Learning*, *23*, 134–148.

Elo, S., & Kyngas, H. (2007). The qualitative content analysis process. *Journal of Advanced Nursing*, *62*, 107–115.

Enriquez, J. G. (2008). Translating networked learning: un-tying relational ties. *Journal of Computer Assisted Learning*, *24*(2), 116–127.

Garrison, D. R., & Arbaugh, J. B. (2007). Researching the community of inquiry framework: review, issues, and future directions. *The Internet and Higher Education*, *10*, 157–172.

Garrison, D. R., & Cleveland-Innes, M. (2005). Facilitating cognitive presence in online learning: interaction is not enough. *The American Journal of Distance Education*, *19*, 133–148.

Garrison, D. R., Anderson, T., & Archer, W. (1999). Critical inquiry in a text-based environment: computer conferencing in higher education. *The Internet and Higher Education*, *2*, 87–105.

Garrison, D. R., Anderson, T., & Archer, W. (2001). Critical thinking, cognitive presence, and computer conferencing in distance education. *American Journal of Distance Education*, *15*(1), 7–23.

Goodyear, P., Jones, C., & Thomson, K. (2014). Computer-supported collaborative learning: instructional approaches, group processes and educational designs. In J. M. Spector, M. D. Merrill, J. Elen & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology* (4th ed., pp. 439–451). New York: Springer.

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*, 397–431.

Guzzo, R. A., & Dickson, M. W. (1996). Teams in organizations: Recent research on performance and effectiveness. *Annual Review of Psychology*, *47*, 307–338.

Harris, K. L. (1996). Content analysis in negotiation research: a review and guide. *Behavior Research Methods, Instruments and Computer*, *28*, 458–467.

Harwood, T. G., & Garry, T. (2003). An overview of content analysis. *The Marketing Review*, *3*, 479–498.

Henri, F. (1992). Computer conferencing and content analysis. *Collaborative learning through computer conferencing* (pp. 117–136): Springer.

Heo, H., Lim, K. Y., & Kim, Y. (2010). Exploratory study on the patterns of online interaction knowledge co-construction in project-based learning. *Computers & Education*, *55*, 1383–1392.

Hewitt, J., Brett, C., & Peters, V. (2007). Scan rate: a new metric for the analysis of reading behaviors in asynchronous computer conferencing environments. *American Journal of Distance Education*, *21*, 215–231.

Ho, A. D., Reich, B. J. F., Nesterko, S. O., Seaton, D. T., Mullaney, T. P., Waldo, J. H., & Chuang, I. (2014). HarvardX and MITx: the first year of open online courses, Fall 2012–Summer 2013. *HarvardX and MITx Working Paper*, *1*, 1–33.

Hung, J. L., & Zhang, K. (2008). Revealing online learning behaviors and activity patterns and making predictions with data mining techniques in online teaching. *MERLOT Journal of Online Learning and Teaching*, *4*, 426–437.

Jo, I., Yoon, M., & Ha, K. (2013). Analysis of relations between learners time management strategy, regularity of learning, and learning performance: a learning analytics approach. Paper presented at the e-learning korea 2013, Seoul, COEX.

Jo, I., Kim, D., & Yoon, M. (2014a). Analyzing the log patterns of adult learners in LMS using learning analytics. Paper presented at the Proceedings of the Fourth International Conference on Learning Analytics and Knowledge.

Jo, I., Yu, T., Lee, H., & Kim, Y. (2014b). Relations between student online learning behavior and academic achievement in higher education: a learning analytics approach. Paper presented at the International Conference of Smart Learning Environment, Hongkong.

Jun, L. (2004). *An introduction to social network analysis*, (pp. 145–160). Beijing: *Social Sciences Academic Press*.

Kanuka, H., & Anderson, T. (1998). Online social interchange, discord, and knowledge construction. *Journal of distance education*, *13*(1), 57–74.

Kim, M., & Lee, E. (2012). A multidimensional analysis tool for visualizing online interactions. *Journal of Educational Technology & Society*, *15*(3), 89–102.

Koller, D., Ng, A., Do, C., & Chen, Z. (2013). Retention and intention in massive open online courses: in depth. *Educause Review*, *48*, 62–63.

Krippendorff, K. (1980). *Content analysis: an introduction to its methodology*. Newbury Park: Sage Publications.

Land, S. M., Hannafin, M. J., & Oliver, K. (2012). Student-centered learning enviornments: foundations, assumptions and design. In D. Jonassen & S. Land (Eds.), *Theoretical foundations of learning environment* (pp. 3–25). New York and London: Routledge.

Lara, J. A., Lizcano, D., Martínez, M. A., Pazos, J., & Riera, T. (2014). A system for knowledge discovery in e-learning environments within the European Higher Education Area—application to student data from Open University of Madrid, UDIMA. *Computers & Education*, *72*, 23–36.

Liccardi, I., Ounnas, A., Pau, R., Massey, E., Kinnunen, P., Lewthwaite, S., … & Sarkar, C. (2007). The role of social networks in students' learning experiences. *In ACM SIGCSE Bulletin* (Vol. *39*, No. 4, pp. 224–237). ACM.

Light, P., Nesbitt, E., Light, V., & White S. (2000). Variety is the spice of life: Student use of CMC in the context of campus based study. *Computers and Education*, *34*, 257–267.

Loncar, M., Barrett, N. E., & Liu, G.-Z. (2014). Towards the refinement of forum and asynchronous online discussion in educational contexts worldwide: trends and investigative approaches within a dominant research paradigm. *Computers & Education*, *73*, 93–110.

Mason, R. B. (2011). Student engagement with, and participation in, an e-Forum. *Journal of Educational Technology & Society*, *14*(2), 258–268.

Neundorf, K. (2002). *The content analysis guidebook*. Thousand Oaks, CA: Sage Publications Inc.

Nickel, T. B. (2002). Student-to-student interaction in online discussions: the role of moderator status. *Abstract in Dissertation Abstracts International*, *63*, 152.

Palazuelos, C., García-Saiz, D., & Zorrilla, M. (2013). Social network analysis and data mining: an application to the e-learning context. In C. Bădică, N. Nguyen & M. Brezovan (Eds.), *Computational collective intelligence. Technologies and applications* (Vol. *8083*, pp. 651–660). Heidelberg: Springer Berlin.

Rabbany, R., Takaffoli, M., & Zaïane, O. R. (2011). Analyzing participation of students in online courses using social network analysis techniques. Paper presented at the Proceedings of educational data mining.

Robson, C. (1993). *Real world research. a resource for social scientists and practitioner-researchers*. Oxford: Blackwell Publishers.

Romero, C., Lopez, M., Luna, J., & Ventura, S. (2013). Predicting students' final performance from participation in online discussion forums. *Computers & Education*, *68*, 458–472.

Schrire, S. (2004). Interaction and cognition in asynchronous computer conferencing. *Instructional Science*, *32*(6), 475–502.

Scott, J. (1991). *Social network analysis*. London: Sage.

Tirado, R., Hernando, A., & Aguaded, J. (2012). The effect of centralization and cohesion on the social construction of knowledge in discussion forums. *Interactive Learning Environments*, 1–24.

Topcu, A., & Ubuz, B. (2008). The effects of metacognitive knowledge on the pre-service teachers' participation in the asynchronous online forum. *Educational Technology & Society*, *11*(3), 1–12.

Tsai, W. (2001). Knowledge transfer in intraorganizational networks: Effects of network position and absorptive capacity on business unit innovation and performance. *Academy of management journal*, *44*(5), 996–1004.

Vrasidas, C., & McIsaac, M. S. (1999). Factors influencing interaction in an online course. *American Journal of Distance Education*, *13*(3), 22–36.

Wasserman, S., & Faust, K. (1997). *Social network analysis: Methods and applications*. Cambridge: Cambridge University Press.

Webb, E., Jones, A., Barker, P., & van Schaik, P. (2004). Using e-learning dialogues in higher education. *Innovations in Education and Teaching International*, *41*(1), 93.

Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education*, *46*, 71–95.

Wise, A. F., Speer, J., Marbouti, F., & Hsiao, Y.-T. (2013). Broadening the notion of participation in online discussions: examining patterns in learners' online listening behaviors. *Instructional Science*, *41*, 323–343.

Zheng, L., Yang, K., & Huang, R. (2012). Analyzing interactions by an IIS-map-based method in face-to-face collaborative learning: an empirical study. *Journal of Educational Technology & Society*, *15*(3), 116–132.

Zhu, E. (2006). Interaction and cognitive engagement: an analysis of four asynchronous online discussions. *Instructional Science*, *34*, 451–480.