# Analytics of communities of inquiry: Effects of learning technology use on cognitive presence in asynchronous online discussions

**Citation for published version:**
Kovanovic, V, Gasevic, D, Joksimovic, S, Hatala, M & Adesope, O 2015, 'Analytics of communities of inquiry: Effects of learning technology use on cognitive presence in asynchronous online discussions', *Internet and Higher Education*, vol. 27, pp. 74–89. https://doi.org/10.1016/j.iheduc.2015.06.002

**Digital Object Identifier (DOI):**
10.1016/j.iheduc.2015.06.002



**Published In:**
Internet and Higher Education

# Analytics of communities of inquiry: Effects of learning technology use on cognitive presence in asynchronous online discussions

Vitomir Kovanović[a,*], Dragan Gašević[a], Srećko Joksimović[b], Marek Hatala[b], Olusola Adesope[c]

[a] *University of Edinburgh, Edinburgh EH8 9YL, United Kingdom*
[b] *Simon Fraser University, Vancouver BC V5A 1S6, Canada*
[c] *Washington State University, Pullman WA 99163, United States*

**Abstract**

This paper describes a study that looked at the effects of different technology-use profiles on educational experience within communities of inquiry, and how they are related to the students' levels of cognitive presence in asynchronous online discussions. Through clustering of students (N=81) in a graduate distance education engineering course, we identified six different profiles: 1) task-focused users, 2) content-focused no users, 3) no users, 4) highly intensive users, 5) content-focused intensive users, and 6) socially-focused intensive users. Identified profiles significantly differ in terms of their use of learning platform and their levels of cognitive presence, with large effect sizes of 0.54 and 0.19 multivariate $\eta^2$, respectively. Given that several profiles are associated with higher levels of cognitive presence, our results suggest multiple ways for students to be successful within communities of inquiry. Our results also emphasize a need for a different instructional support and pedagogical interventions for different technology-use profiles.

*Keywords:* Community of Inquiry, Learner Profiles, Clustering, Distance Education, Human Agency

## 1. Introduction

The importance of social interaction for reaching higher levels of learning is widely acknowledged in contemporary education (Anderson and Dron, 2010). Educational research offers many accounts of the benefits of social interaction on the development of skills such as critical thinking, creativity, and argumentation (Sawyer, 2006; Garrison et al., 2010b; Dawson et al., 2011). Affordances of the modern (educational) technology enable for effective social interaction, information seeking, and knowledge building. More importantly, educational research offered approaches that can help design, facilitate, and direct an effective educational experience in communities and/or networks of learners. The Community of Inquiry (CoI) model (Garrison et al., 1999; Garrison, 2011; Garrison and Arbaugh, 2007) is a well-known framework in this context. By using qualitative and quantitative research methods, the research centered around the CoI model offered a remarkable amount of empirical evidence that explain an interplay of teaching, cognition, and socialization in communities of inquiry (Garrison et al., 2010b).

Although heavily dependent on educational technology, our review of the CoI literature revealed rather limited research that studied the relationships between learners' use of educational technology and the dimensions of the CoI model. The only study found in our literature review that focused on this issue was by Rubin et al. (2013), and it investigated the association of learners' perceived value of educational technology affordances and perceived value of the core dimensions of the CoI model. However, the study of Rubin et al. used self-reports to gather students' perceived value of educational technology. In this paper, we propose that *learning analytics* (Buckingham Shum and Ferguson, 2012; Siemens and Gasevic, 2012) can: i) offer methods to advance understanding of the CoI model, especially in relation to learners' knowledge construction process and agency, ii) reveal how learners interact with educational technology in communities of inquiry, and iii) drive the development of new instructional approaches that can enhance educational experience for diverse sub-populations of learners that can emerge in communities of inquiry. More specifically, in this paper we report on the results of a study in which we:

1. Propose a *method* for identification of learner profiles – reflective of learners' agency about decisions making when selecting tools to study – based on trace data about their online learning activities performed in learning management systems.

2. Investigate the *effect* of the identified learner profiles on the development of cognitive presence – one of the three main dimensions of the CoI model – extracted from online discussion transcripts of a community of inquiry.

3. *Interpret* results in relation to instructional practice and

*Corresponding author.

existing theories on metacognition, motivation, and conceptions of and approaches to learning.

## 2. Theoretical Background

### 2.1. The community of inquiry model

Built upon the social constructivist perspective to learning, the Community of Inquiry model is recognized by some as the most important model of e-learning today (Garrison and Arbaugh, 2007). The CoI model defines a community of inquiry as *"a group of individuals who collaboratively engage in purposeful critical discourse and reflection to construct personal meaning and confirm mutual understanding"* (Garrison, 2011, p. 2). The model describes a community of inquiry through the three interdependent dimensions, also known as *presences* (Garrison, 2007; Garrison et al., 2010a; Kanuka, 2011):

1) *Cognitive presence* is a central dimension of the model that describes the learning phases from the initial practical inquiry to the eventual knowledge construction and problem solution (Garrison et al., 2001).

2) *Social presence* explains important social relationships among the members of the learning community and the social climate that contributes to the success of learning and attainment of the learning objectives (Rourke et al., 1999).

3) *Teaching presence* is focused on the role of instructors in course design, organization, and delivery, and instructions that guide social and cognitive processes to desired learning outcomes (Anderson et al., 2001).

This paper focuses on the study of cognitive presence which is defined as *"the extent to which the participants in any particular configuration of a community of inquiry are able to construct meaning through sustained communication"* (Garrison et al., 1999, p. 89). Cognitive presence proved to be a suitable instrument to assess critical thinking (Garrison et al., 2001), given that oral and textual communication (e.g., via discussion forums) have been shown to stimulate development of critical thinking skills. In essence, cognitive presence is a process model describing the development of higher-order thinking rather than individual learning outcomes (Akyol and Garrison, 2011b; Akyol et al., 2009). It is rooted in Dewey's (1910) social-constructivist views of learning and is operationalized through the practical inquiry model (Garrison et al., 2001) that defines four phases of inquiry learning cycle:

1) *Triggering Event*: In this phase, a learning cycle is initiated by a problem or dilemma, which is in the formal educational setting typically introduced by the instructor.

2) *Exploration*: This phase is characterized by exploration, brainstorming, and other activities in which students gather information relevant to the problem or task at hand.

3) *Integration*: In this phase, after gathering an appropriate body of information, students synthesize and integrate different bits of information, while being selective and filtering out all irrelevant information.

4) *Resolution*: The last phase is the resolution of the original problem which is – in the context of formal education – typically achieved through vicarious actions and hypothesis testing. Very often resolution of the original problem initiates a new learning cycle with a new triggering event.

The research methods related to the CoI of inquiry include both: i) *qualitative methods* – primarily based on the use of quantitative content analysis of discussion transcripts and different coding schemes for the assessment of the three dimensions of CoI model (Rourke and Anderson, 2004) and, ii) *quantitative methods* – primarily based on the CoI survey instrument, which was developed for measuring self-reported values of each of the three CoI dimensions (Garrison et al., 2010b). Both coding schemes (i.e., high inter-rater reliability) and the survey instrument (i.e., consistency and factor loadings) have been validated in a number of studies (Arbaugh et al., 2008; Gorsky et al., 2011; Rourke and Anderson, 2004).

Recent studies of the CoI model (Akyol and Garrison, 2011a; Garrison and Akyol, 2013; Shea and Bidjerano, 2010) highlight self-regulated learning (SRL) – a major theory of learning in contemporary educational psychology focusing on the role of metacognition in the learning processes (Bjork et al., 2013) – as central for understanding the CoI model. As cognitive presence includes both self-reflection and collaborative knowledge co-construction (Garrison et al., 2001), *"metacognition mediates between reflection and action"* (Akyol and Garrison, 2011a, p. 186). In order to develop cognitive presence, students need to exercise critical thinking skills, which are primarily meta-cognitive in nature and require communicating one's thinking with others (Akyol and Garrison, 2011a). Garrison and Akyol's research showed that metacognition in a CoI could be characterized as *"complementary self- and co-regulation that integrates individual and shared regulation"* (Garrison and Akyol, 2013, p. 84). That is, participation in a community of inquiry affects their meta-cognitive monitoring and control. This is particularly done through the role of teaching presence whereby instructional design, facilitation, and direct instruction along with peer guidance are intrinsic components of metacognition in a community of inquiry.

### 2.2. Educational technology use and self-regulated learning

One of the central ideas in the modern educational psychology is that learners *do not acquire, but instead construct new knowledge* (Bjork et al., 2013; Winne, 2006; Winne and Hadwin, 1998). One of the major models which conceptually describes this process is self-regulated learning (SRL) (Bjork et al., 2013). It views knowledge construction as being developed through the use of different cognitive, physical, and digital tools to operate on raw materials to create the products of cognition. These products of cognition are evaluated with respect to standards that can be internal (e.g., efforts budgeted to online discussions) and external (e.g., grading policy for online discussions). Moreover, learners are viewed as human *agents* who constantly meta-cognitively: i) control their learning operations by evaluating their study tools, including the decisions if and how to use the tools (Azevedo, 2005), and ii) monitor their learning progress by comparing the products of their learning with the predetermined learning goals.

As suggested by Winne (1982, 2006) and Perkins (1985), the knowledge construction and agency perspectives to learning have several important implications regarding the learners' use of tools. Typically, learning environments are designed to promote personalization and adaptiveness to the learners' needs (Azevedo, 2005). Still, studies indicate that many students do not make use of the available tools and resources in a way which will maximize benefits to the learning (Ellis et al., 2005; Lust et al., 2011, 2013a). Most of the available tools are underused by the majority of the students indicating the lack of awareness, knowledge, or motivation to use the available tools (Lust et al., 2013a). This is shown to be especially important in the complex, fully online environments given the self-directed nature of learning and the limited opportunities for the physical interactions among the students (Shen et al., 2013). Lust et al. (2012) and Clarebout et al. (2013) indicate that for the successful learning in the modern, complex learning environments learners:

1) need to be able to *recognize the opportunities* (e.g., tools or study tactics) that are available in the learning environment. Not all students have the needed meta-cognitively knowledge to recognize the provided learning opportunities (Clarebout et al., 2013) – e.g., the use of asynchronous online discussions for problem solving.

2) need to be able to *draw a connection* between the opportunity and their task at hand – e.g., that participation in the asynchronous online discussions is beneficial for their learning.

3) need to be *meta-cognitively skillful enough* to be able to use the provided opportunity effectively. For example, to find relevant information in the course readings or the Internet, to articulate the information found in a meaningful way, and to finally integrate the information with the currently existing information in the course discussions.

4) need to be *motivated* to invest time and effort in using the opportunity and to meta-cognitively monitor and control the use of the opportunity in relation to their learning task. Likewise, students need to be comfortable with the types and extent of potential risks associated with the offered opportunity; For example, the potential misinterpretation of discussion contributions or the domination by a student or the group of students (Murphy and Coleman, 2004).

### 2.3. Technology-use profiles

In online and blended learning, one important aspect of student self-regulation of learning is the decision on if and if so, how to use the technology offered in a learning environment (Azevedo, 2005). Consistent with the research on self-regulated learning (Zhou and Winne, 2012), the studies of trace data recorded by learning management systems – typically based on a cluster analysis – showed that there were often several types of educational technology users (Lust et al., 2011, 2013a,b; Yen and Lee, 2011; Wise et al., 2013). It was also shown that patterns of educational technology use had different effects on learning outcomes of learners within the same course (Lust et al., 2013b).

In a blended learning environment, the study by Yen and Lee (2011) identified three groups of students based on their technology-use profiles: i) *technology-oriented* students who preferred mobile and web learning, and subsequently exhibited superficial problem solving abilities and general absence of planning and understanding, ii) *efficiency-oriented* students who were characterized by the efficient monitoring of their learning processes and generally better performance than other two groups, and iii) *hybrid-oriented* students who did not have a preference for a particular instructional modality and mostly passively accepted information from the instructors. In a similar manner, the study by Lust et al. (2011) discovered three technology-use profiles in blended environments: i) *no-users* who did not make use of the available face-to-face tools (i.e., use of learning support and feedback sessions) and had a very limited use of the LMS, ii) *intensive* users that frequently used a majority of the tools available in the LMS, and iii) *incoherent* users who used only online tools and did not use the available face-to-face tools. With respect to the academic performance, both intensive and incoherent students had significantly higher academic performances in the course than no-users.

One of the reasons for the observed differences was found to be students' *self-regulation* of the tool use (Lust et al., 2013a). Aligned with the findings in the field of self-regulated learning (Winne, 1982, 2006; Perkins, 1985), the study by Lust et al. showed that majority of the students regulated their learning; however, only three percent of them regulated in accordance with the course objectives. A majority of the students (59%) used a very limited set of available tools – indicating the lack of ability to regulate effectively their learning activities (Perkins, 1985).

Another construct that was explored by Lust et al. (2013b) is the students' achievement *goal-orientation* (Senko et al., 2011), and it was found to be directly related to learning technology-use profiles of students. Generally, mastery goals focus on *gaining* competence while performance goals focus on *demonstrating* performance. In more recent studies, goal orientation is further distinguished along the emotional value that students give to their standards (approach vs. avoidance), resulting in four possible performance goal orientations: i) *mastery-approach*: focus on gaining skills and knowledge, ii) *mastery-avoidance*: focus on avoiding skill decline and learning failures, iii) *performance-approach*: focus on performing better than peers, and iv) *performance-avoidance*: focus on avoiding performing worse than peers (Senko et al., 2011). What is of the direct importance for the current study are the findings of connection between i) mastery goal orientation and active tool use, and ii) performance goal orientation and selective tool use (Lust et al., 2013b). Furthermore, the students did not differ in terms of their perceived value of the provided tools, and in some cases – such as practice quizzes – no-users even had a significantly more positive opinion on the tool usefulness tools despite the fact that they never used them.

The notion of *approaches to learning* (i.e., deep vs. surface) (Trigwell and Prosser, 1991) is another important construct in educational research which was shown to have a significant impact on the learning outcomes, particularly related to student technology use. Research findings indicate connection between deep learning approaches and higher success of learning (Trigwell and Prosser, 1991) and mastery goal orientation (Phan, 2008). With this in mind, student participation in online discussions was analyzed by Wise et al. (2013), and their study revealed that performance-avoidance was directly related to low cognitive engagement. They identified four distinctive profiles of participation different primarily in terms of the *breadth*, *depth*, and *temporal continuity* of discussion participation and the amount of student *reflection* (Wise et al., 2013). Thus, Wise et al. recommend that instructors need to take into the account differences among their students, especially in relation to their goals and approach to learning. Their findings are consistent with the results of Bliuc et al. (2010), who showed the association between surface approaches to learning and fragmented *conception of learning* in discussions (i.e., discussions were considered a way of getting a correct answers fast rather than a way of deepening their broader knowledge), and also cohesive notion of learning in discussions and deep approaches to learning.

From a more holistic perspective, Valle and Duffy (2009) consider the very own idea of *student freedom* in distance education as a challenge for many students. As a result, a key to success is seen in the effective managing of the learning demands associated with the freedom of distance education (Valle and Duffy, 2009). Valle and Duffy in their study found similar three types of technology-use profiles: i) mastery-oriented, who were shown to posses the greatest amount of relevant background experience and were willing to put substantial effort into learning, ii) task-oriented, who had the overall lower levels of effort and were spending the minimal required time, and iii) minimalist, who also put less effort like the task-oriented students, and were also found to prefer working in groups rather than self-paced – indicating a need for more motivation-related support. Their results suggest that the ways in which learners work might be a good indicator of their commitment to learning. Still, in terms of their success, both Wise et al. (2013) and Valle and Duffy (2009) did not found any significant difference among the students which suggest that many different approaches might be successful in terms of academic performance.

### *2.4. Educational technology use within the communities of inquiry*

Educational technology is the major enabler of communities of inquiry. While the effects of technology use on learning and factors of technology acceptance received considerable research attention (McGill and Klobas, 2009), research of the effects of educational technology affordances on the three dimensions of the CoI model is hardly reported in the literature published to date. To our knowledge, a recent paper by Rubin et al. (2013) reports the first study that tried to shed some light on this important issue. By using the three dimensions of the CoI model and self-reports of the use of a selected set of LMSs, they found several patterns of how educational technology relates to the CoI model. Of direct importance for the study reported in this paper is the finding that the level of self-reported cognitive presence is predicted by the self-reported ease of communication provided by the LMS and the self-reported amount of online reading materials, while the self-reported ease of finding information was marginally significant. These findings were expected, as for effective participation in a community of inquiry (e.g., to achieve a high level of cognitive presence), not only did learners participate in online discussions, but rather their learning involved a number of activities such as seeking and reading learning materials, modeling their knowledge through quizzes, and completing course assignments.

Existing research on technology use and self-regulated learning offers a number of theoretical, methodological, and empirical accounts warranting future research on the relationships between educational technology use and the CoI model. It is widely accepted that external and internal conditions play an important role in regulating students' approach to study (Winne and Hadwin, 1998). Studies by Garrison and Cleveland-Innes (2005) and Akyol and Garrison (2011b) looked at learning approaches within communities of inquiry and showed that specific forms of teaching presence – such as instructional leadership in facilitation, direct instruction, and appropriate course structure – already have positive effect on the promotion of deep approaches to learning, and thus, establishing and sustaining of high levels of cognitive presence.

With respect to the adopted methodological approaches, one of the primary means of studying technology use, student agency, and self-regulation of learning is through the self-reported data (Winne and Jamieson-Noel, 2002). The previously mentioned studies by Lust et al. (2013b), Bliuc et al. (2010), and Valle and Duffy (2009) are some of the examples. However, self-reports are not the most reliable instrumentation to study the effects of education technology on learning processes and outcomes, given the biases and inaccuracy associated with understanding metacognition. For example, Winne and Jamieson-Noel showed that learners tended to overestimate their use of study tools in a learning software. Although self-reports offer valuable insights in learners' perception of learning, previous research found that learners often self-report *"biased information arising from incomplete and reconstructed memories plus subjective and implicit theories of the mental processes involved"* (Zhou and Winne, 2012, p. 414). In order to overcome some of the challenges of self-reported measures, the use of more objective measures – such as students' trace data (Zhou and Winne, 2012, p. 414) – is often recommended (Gonyea, 2005).

## 3. Research questions

From the studies presented in Section 2.3 we can see a strong evidence supporting the difference among students in terms of their technology use, and the importance of students' goal-orientation, self-regulation, and approaches to learning on shaping technology-use profiles. From the existing research studies (Lust et al., 2011, 2013a,b; Wise et al., 2013; Valle and Duffy, 2009; Yen and Lee, 2011; Bliuc et al., 2010), several reoccurring technology-use profiles can be seen: i) a group of students with lower activity levels (i.e., no users and minimalist users) typically associated with lower levels of meta-cognitive capabilities for self-regulation, surface approaches to learning, and performance-goal orientation, ii) a group with very high levels of activity (i.e., intensive and mastery oriented-users) who had deep approaches to learning and mastery-goal orientation, and iii) a group of selective users (i.e., incoherent, selective, limited, and efficiency-oriented users) that typically exhibited performance goal-orientation and higher of learning self-regulation – although for most of the time not in a desirable way.

Our hypothesis is that we will find the same or very similar technology-use profiles within the communities of inquiry. Still, given the social-constructivist view of learning in communities of inquiry, we are interested in how this particular context affect the hypothesized profiles. Thus, our first research question is:

> RESEARCH QUESTION 1:
> *What are the main technology-use profiles within communities of inquiry? How does the collaborative nature of learning within communities of inquiry affects the theorized technology use profiles? Are there any CoI-specific learning technology-use profiles not previously identified, and if so, how they can be explained in terms of the students self-regulation of learning, goal-orientation and approaches to learning?*

In this paper, we build on the existing research of the effects of technology-use (Rubin et al., 2013) and self-regulation (Garrison and Akyol, 2013; Akyol and Garrison, 2011a) on the learning success within communities of inquiry. More precisely, we explore effects of individual, internal regulation – as evident through the different profiles of technology use – on the development of the cognitive presence. Building on the suggested relationship between approaches to learning and technology-use profiles, we investigate the relationship between the CoI model and student approaches to learning, as indicated by the observed technology-use profiles. Thus, our second research question is:

> RESEARCH QUESTION 2:
> *How are the profiles of learning technology use related to the development of cognitive presence in communities of inquiry and which profiles have the strongest effect on cognitive presence?*

Given the existing evidence of the effects of approaches to learning and goal orientation on the development of deep critical thinking skills (Trigwell and Prosser, 1991; Phan, 2008; Entwistle, 2009; Bliuc et al., 2010; Wise et al., 2013; Lust et al., 2013b), we expect to find differences in terms of the students development of cognitive presence and ultimately on the success of learning. In this study, we focus on the cognitive presence; however in the future studies we will also examine the effects on the academic performance – as operationalized through the final course grades. From the practical perspective, this research question seeks to provide: i) insights which can be potentially used for adaptation of the provided feedback to learners based on their technology-use profiles, and, ii) a guide for instructors that can help them to define specialized and eventually more effective instructional interventions targeting students with specific styles of educational technology use.

## 4. Methods and materials

### *4.1. Course*

#### *4.1.1. Course Organization*

The data for this study originated from a thirteen weeks long, masters level course offered through a fully online instructional condition at a Canadian public university. The course is research intensive and focuses on understanding of the current research trends and challenges in the area of software engineering field. To successfully finish the course, students were expected to complete several activities including four tutor marked assignments (TMAs):

- **TMA1** (15% of the final grade, submitted during weeks 3-5): The students are expected to: i) select and read a peer-reviewed paper on a course topic, ii) prepare a short video presentation that summarizes information presented in the paper and provides a critical review of the paper, iii) initiate a new discussion about the paper with other students. This assignment is primarily factual, and focus on presenting particular challenges in a software engineering field.

- **TMA2** (25% of the final grade, submitted at end of week 6): The students were required to write a literature review paper (5-6 pages in the ACM proceedings format) on a selected topic in software engineering. The marking scheme for this assignment was as follows: i) 80% of the grade was given based on two double blind peer reviews (35% of the grade each) and instructor review (30% of the paper grade), and ii) 20% was given by the instructor based on the quality of provided peer-review comments. This assignment has strong focus on building conceptual understanding of a particular research problem in software engineering field.

- **TMA3** (15% of the final grade, submitted at end of week 9): Students were required to answer six questions (400-500 words per each question) related to course readings that were designed to demonstrate critical thinking and synthesis skills. The focus of this assignment is also on conceptual knowledge and analysis and evaluation of the existing solutions of a given research problem.
- **TMA4** (30% of the final grade, submitted at end of the course): In the final assignment, students worked in small groups (2-3 students) on a selected software engineering topic. The main outcome was a project report and a all developed software artifacts (e.g., models and source code) that were then marked by the instructor. This assignment has a particularly procedural focus on building practical skills related to selected research topic and evaluation of proposed solutions.
- **Course Participation** (15% of the final grade): The course had a particular focus on stimulating productive online discussions and the students were expected to actively participate in course discussions.

#### *4.1.2. Dataset*

The data consisted from the 6 offerings of the described course (Winter 2008, Fall 2008, Summer 2009, Fall 2009, Winter 2010, Winter 2011) with the total of 81 students with the average cohort size of 13.5 students (SD=5.1). The slightly larger variations in cohort sizes were due to the course under study not being a mandatory course, but a part of the group of 11 'core' courses, and university regulations required students to complete three core course in their masters degree programs.

The course was offered through the Moodle LMS[1], which hosted all the readings, assignments and student discussion boards. The trace data was obtained by automated extraction process from the Moodle's PostgreSQL database and consisted of almost 200,000 log records of the different student activities. In these six offerings, the students posted 1,747 messages in total which – together with the LMS trace data – represented the main data source for this study. The numbers of students and messages in each course are shown in Table 1.

In order to measure the levels of cognitive presence, all 1747 messages from online discussion forums were coded using the CoI coding instrument described in Garrison et al. (2001). All messages were coded by two human coders and they achieved an excellent coding agreement (Cohen's Kappa=0.97), disagreeing in less than 2% of the messages (i.e., total of 32 messages). In those cases, the disagreements were resolved through the discussion between the coders. The results of the coding are shown on Table 2.

[1]http://moodle.org

**Table 1**
Course offering statistics

| | Student count | Message count |
|---|---|---|
| Winter 2008 | 15 | 212 |
| Fall 2008 | 22 | 633 |
| Summer 2009 | 10 | 243 |
| Fall 2009 | 7 | 63 |
| Winter 2010 | 14 | 359 |
| Winter 2011 | 13 | 237 |
| Average (SD) | 13.5 (5.1) | 291.2 (192.4) |
| Total | 81 | 1747 |

**Table 2**
Cognitive Presence Coding

| ID | Phase | Messages | (%) |
|---|---|---|---|
| 0 | Other | 140 | 8.01% |
| 1 | Triggering Event | 308 | 17.63% |
| 2 | Exploration | 684 | 39.17% |
| 3 | Integration | 508 | 29.08% |
| 4 | Resolution | 107 | 6.12% |
| | All phases | 1747 | 100% |

### *4.2. Measurement instrument*

In order to identify technology-use profiles, thirteen variables based on students' use of LMS were extracted (Table 3), similarly to the work of Lust et al. (2011, 2013a,b) and Valle and Duffy (2009). We extracted count and time-on-task variables, focusing only on LMS activities that students were expected to use given the particular course design. For most of the activities, both counts and time-on-task were extracted, while for some activities only count measures were extracted, as the notion of time-on-task was not meaningful (e.g., searching discussion boards). Table 3 shows that the extracted variables can be divided into two groups: variables related to the static course content (reading resources and assignments) and variables related to online discussions.

With respect to the outcome variables, we used the counts of messages in the phases of cognitive presence that were collected through a quantitative content analysis using CoI's cognitive presence coding scheme (Garrison et al., 2001) which is described in detail in Section 4.1.1. Hence, for each student five outcome measures were extracted, four corresponding to the four phases of cognitive presence and one corresponding to the messages without traces of cognitive presence (coded as other). Typically, non-cognitive (i.e., other) messages included messages serving purely social purposes, such as acknowledging someone else's message.

**Table 3**
Extracted Features

| # | Type | Code | Name | Description |
|---|---|---|---|---|
| 1 | Clustering Variables | ULC | UserLoginCount | Total number of times student logged into the system. |
| 2 | (Content) | CVC | CourseViewCount | Total number of times student viewed general course information. |
| 3 | | AVT | AssignmentViewTime | Total time spent on all course assignments. |
| 4 | | AVC | AssignmentViewCount | Total number of times student opened one of the course assignments. |
| 5 | | RVT | ResourceViewTime | Total time spent on reading the course resources. |
| 6 | | RVC | ResourceViewCount | Total number of times student opened one of the course resource materials. |
| 7 | Clustering Variables | FSC | ForumSearchCount | Total number of times student used search function on the discussion boards. |
| 8 | (Discussions) | DVT | DiscussionViewTime | Total time spent on viewing course's online discussions. |
| 9 | | DVC | DiscussionViewCount | Total number of time student opened one of the course's online discussions. |
| 10 | | APT | AddPostTime | Total time spent on posting discussion board messages. |
| 11 | | APC | AddPostCount | Total number of the discussion board messages posted by the student. |
| 12 | | UPT | UpdatePostTime | Total time spent on updating one of his discussion board messages. |
| 13 | | UPC | UpdatePostCount | Total number of times student updated one of his discussion board messages. |
| 1 | Outcome Variables | TEC | TriggeringEventCount | Number of posted *triggering event* messages. |
| 2 | | EC | ExplorationCount | Number of posted *exploration* messages. |
| 3 | | IC | IntegrationCount | Number of posted *integration* messages. |
| 4 | | RC | ResolutionCount | Number of posted *resolution* messages. |
| 5 | | OC | OtherCount | Number of posted non-cognitive (*other*) messages. |

### 4.3. Pre-processing clickstream data

As recorded trace data is mainly a stream of actions together with occurrence timestamps, the first step in our analysis was to pre-process trace data to extract count and time-on-task variables. Count measures were extracted by simply counting for each action the number of times it was performed by each student, while time-on-task variables were calculated from the time differences between the logged actions. This is the typical approach that has been extensively used in the similar studies (Lust et al., 2013b,a, 2011; Wise et al., 2013; Valle and Duffy, 2009), as well as in many Learning Analytics and Educational Data Mining studies (Macfadyen and Dawson, 2010; Morris et al., 2005; Romero et al., 2008). The primary assumption – which is commonly done in time-on-task estimation (Valle and Duffy, 2009) – is that time between two logged events is spent on a particular learning activity.

One particular challenge of this approach that has been already identified by Wise et al. (2013) and Valle and Duffy (2009) is the detection of time when a user has left the system. Even though LMSs have a logout button, a great majority of students do not use it and simply close their web browser window. Therefore, to prevent from severely overestimating time-on-task measures, durations of of last activities for each study session (i.e., activities that were followed by a login action) were estimated as the student's average time for that particular activity. Finally, as sometimes students would just leave the browser window open for an extended period of times (and thus their next study session does not start with the login action), an upper limit of the duration of each activity was set to one hour, similarly to the work of Valle and Duffy.

### 4.4. Clustering

For the discovery of students' technology-use profiles cluster analysis techniques and the popular agglomerative hierarchical clustering algorithm were adopted (Hastie et al., 2013). Much like Wise et al. (2013) and Valle and Duffy (2009), we used Ward's merging procedure and Euclidean distance measure (Hastie et al., 2013). As some of the variables are counts and some are time durations, similarly to the work of Valle and Duffy all variables were first standardized in order to enable their equal weighting. Finally, each cluster was summarized by calculating cluster's *centroid*, which represented the mean values of all cluster members across all clustering variables.

### 4.5. Statistical analysis

For assessing the difference between student clusters a multivariate analysis of variance (MANOVA) (Tabachnick and Fidell, 2007) was used. To validate the difference between the discovered clusters a MANOVA model with cluster assignment as a single independent variable and thirteen clustering variables (Table 3) as the dependent measures was constructed, similarly to the work of Lust et al. (2011, 2013a,b). To check for the difference in terms of students' cognitive presence, we constructed a MANOVA model with cluster assignment as a single independent variable and five dependent variables: four measures of cognitive presence (i.e., the number of messages in four phases of cognitive presence) and the number of non-cognitive messages (coded as other).

Before running MANOVAs, similarly to the work of Lust et al. (2011, 2013a,b), the homogeneity of covariances assumption was checked using Box's M test and homogeneity of variances using Levine's test. To protect from the assumption violations, we log-transformed the data and used the Pillai's trace statistic which is considered to be a robust against assumption violations (Field et al., 2012). As a final protection measure, obtained MANOVA results were compared with the results of the robust rank-based variation of the MANOVA analysis using the approach by Nath and Pavur (1985).

In the case of significant MANOVA, a follow-up univariate one-way analyses of variance (ANOVA) were conducted

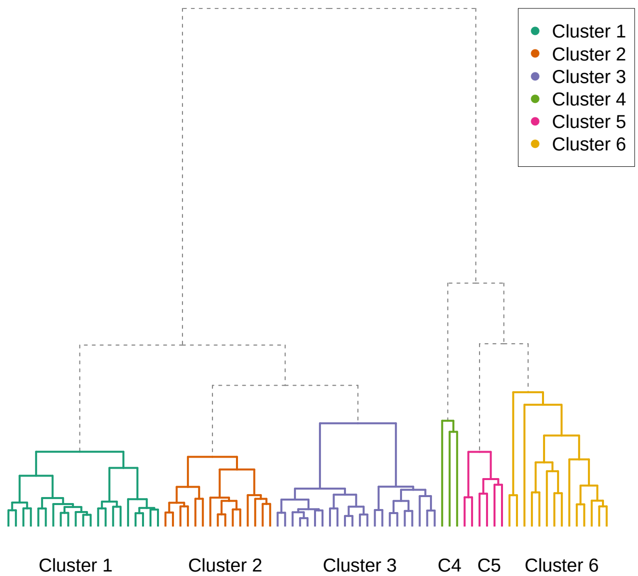



**Fig. 1** Dendogram of student clustering.

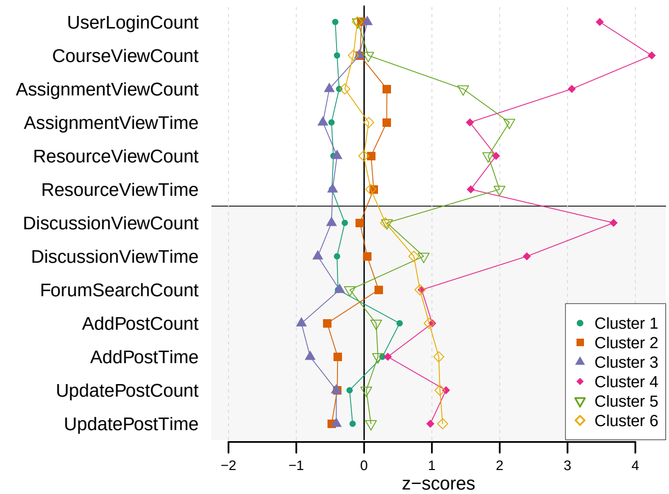



**Fig. 2** Clustering results

on each of the dependent variables. This use of the univariate follow-ups after a significant multivariate analysis is often considered as a "protection" from the Type I errors arising from the direct use of multiple ANOVAs (Bock, 1985). However, this approach only protects against Type I error inflation for those dependent variables for which a significant multivariate effect was found (Bray and Maxwell, 1985). Thus, in order to further control for the Type-I error rate inflation due to the multiple comparisons, the very conservative Bonferroni correction was adopted. Before running ANOVAs, the homogeneity of variance was checked using Levine's test, and when it was violated, non-parametric Kruskal-Wallis test was used. Significant Kruskal-Wallis tests were followed up by a pairwise comparisons also using Bonferroni correction. Finally, after significant ANOVAs, Tukey's honest significant difference (HSD) test was used to check for the differences among the individual pairs of clusters.

Given that univariate follow-up analysis does not examine multivariate differences among different conditions, we used the discriminatory factor analysis (DFA), which is very commonly used to assess the multivariate effects of the significant MANOVA analyses (Field et al., 2012). Two approaches together (i.e., ANOVAs and DFA) are considered to provide a complete picture of the multivariate differences among the different groups (Field et al., 2012).

## 5. Results

### *5.1. Clustering results*

#### *5.1.1. Selecting number of clusters*

Fig. 1 shows the dendogram tree of the student clustering by using the agglomerative hierarchical clustering algorithm. The length of the vertical connecting lines in the dendogram tree indicates the difference between two merged clusters. Starting with the two-cluster solution, more detailed cluster solutions were evaluated, and as a final clustering solution selected the solution with six clusters. Clustering solutions with more than six clusters were only different by having additional clusters with either one or two students, which was indicative that the appropriate number of clusters was selected. Fig. 2 shows the difference between the centers of all the six final clusters, while Table 4 shows the raw scores of clustering variables for each cluster. The systematic relationship between individual course offerings and identified clusters (Fig. 3) was checked, and no clear pattern was observed (Pearson's correlation $r = -0.159$, $p = 0.156$). Finally, to make the reporting of the results easier, each cluster was assigned a label (Table 5) based on our analysis and interpretation of the observed cluster differences. Section 6 provides an in-depth discussion of the clustering results.

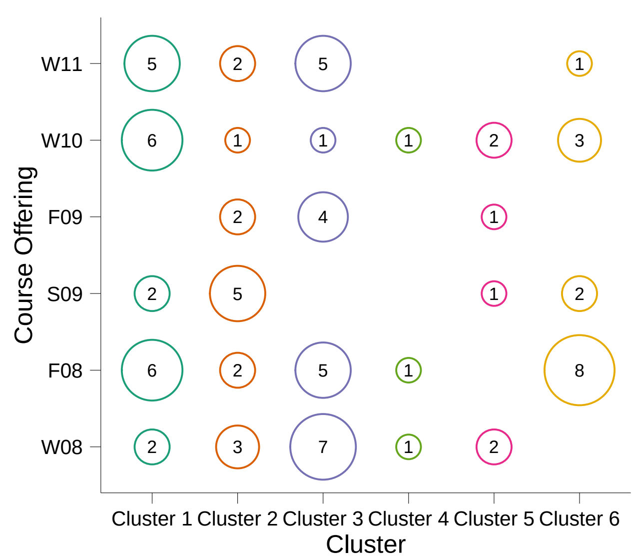



**Fig. 3** Distribution of clusters across course offerings.

**Table 4**
Descriptive statistics of cluster centers (raw scores). Time measures are shown in hours.
1: Task-focused users, 2: Content-focused no users, 3: No users, 4: Highly intensive users, 5: Content-focused intensive users, 6: Socially-focused intensive users

| | | Cluster 1 (N=21) | | Cluster 2 (N=15) | | Cluster 3 (N=22) | | Cluster 4 (N=3) | | Cluster 5 (N=6) | | Cluster 6 (N=14) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| # | Variable | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| 1 | `UserLoginCount` | 285 | 129 | 420 | 169 | 450 | 297 | 1650 | 938 | 401 | 101 | 399 | 183 |
| 2 | `CourseViewCount` | 403 | 176 | 598 | 242 | 592 | 343 | 3080 | 1030 | 668 | 182 | 541 | 294 |
| 3 | `AssignmentViewCount` | 63.9 | 21.5 | 87.1 | 15 | 59 | 16.1 | 177 | 32.7 | 124 | 20.8 | 66.6 | 25 |
| 4 | `AssignmentViewTime` | 6.42 | 2.55 | 11.5 | 4.13 | 5.63 | 3.4 | 19.2 | 6.81 | 22.8 | 6.95 | 9.85 | 4.27 |
| 5 | `ResourceViewCount` | 31.6 | 17 | 50.8 | 24.3 | 33.4 | 15.4 | 115 | 53.6 | 111 | 48.4 | 47.1 | 30.6 |
| 6 | `ResourceViewTime` | 4.07 | 2.37 | 8 | 4.26 | 4.13 | 2.46 | 17.1 | 12.6 | 19.8 | 8.31 | 7.72 | 5.64 |
| 7 | `DiscussionViewCount` | 166 | 39.2 | 216 | 91.1 | 123 | 57.5 | 1050 | 706 | 304 | 105 | 300 | 134 |
| 8 | `DiscussionViewTime` | 10.8 | 4.56 | 15.8 | 7.81 | 7.68 | 5.58 | 42 | 15.2 | 25.1 | 9.74 | 23.5 | 11.5 |
| 9 | `ForumSearchCount` | 0 | 0 | 1.13 | 1.46 | 0.0455 | 0.213 | 2.33 | 2.08 | 0.333 | 0.816 | 2.29 | 3.69 |
| 10 | `AddPostCount` | 33.1 | 6.57 | 20.9 | 5.51 | 16.5 | 6.14 | 38.7 | 18.5 | 29.2 | 10 | 38.1 | 10.1 |
| 11 | `AddPostTime` | 8.17 | 3.62 | 5.34 | 2.58 | 3.6 | 2.04 | 8.51 | 5.29 | 7.89 | 3.9 | 11.7 | 4.33 |
| 12 | `UpdatePostCount` | 5.43 | 6.3 | 2.67 | 3.98 | 2.45 | 2.32 | 27 | 29.7 | 9.17 | 8.75 | 25.6 | 25.5 |
| 13 | `UpdatePostTime` | 0.164 | 0.202 | 0.0517 | 0.0855 | 0.0762 | 0.0824 | 0.579 | 0.603 | 0.261 | 0.237 | 0.645 | 0.593 |

### *5.1.2. Description of identified clusters*

In Fig. 2, we can see that students in cluster one (task-focused users) are mostly below the average mean value for all the clustering variables, except those that were related to posting discussion messages. With respect to cluster two (content-focused no users), it was around average for most variables, except for the those variables related to online discussions. The mean values of Cluster three (no users) are below the means values for all clustering variables, except for the number of logins into the LMS which was around the average. In a complete contrast to cluster three (no users), the students from cluster four (highly intensive users) had all of their mean values above the overall mean with some of them being several standard deviations larger than the average (e.g., the number of logins into the system and number of course, assignment, and discussion views). Students in cluster five (content-focused intensive users) show more moderate values, with the focus on "non-social" aspects of LMS, while students in cluster six (socially-focused intensive users) show the opposite trend, primarily focusing on online discussion participation.

**Table 5**
Summary of Cluster Differences

| # | Size | Label | Characteristics |
|---|---|---|---|
| 1 | 21 | Task-focused users | Overall below average activity, Above average message posting activity |
| 2 | 15 | Content-focused no users | Below average discussions-related activity, Average content-related activity, emphasis on assignments |
| 3 | 22 | No users | Overall below average activity, slightly bigger in discussion-related activities |
| 4 | 3 | Highly intensive users | Significantly most active students, especially in content-related activities |
| 5 | 6 | Content-focused intensive users | Above average content-related activity, Average discussion-related activity |
| 6 | 14 | Socially-focused intensive users | Above average discussion-related activity, Average content-related activity |

### *5.2. Analysis of cluster differences*

In order to check for the statistical significance between the discovered clusters, a one-way multivariate analysis of variance (MANOVA) is conducted with the students' cluster assignment as the single independent variable and the thirteen technology use measures as the dependent variables. As MANOVA requires more data points in each group than the number of dependent variables (Tabachnick and Fidell, 2007), clusters four (highly intensive users) and five (content-focused intensive users) were removed from the analysis, given that they have only three and six students, respectively. The assumption of homogeneity of covariances was tested using Box's M test (Field et al., 2012) which was not accepted. Thus, Pillai's trace statistic was used, as it is more robust to the assumption violations (Field et al., 2012) together with the Bonferroni correction method. A statistically significant MANOVA effect was obtained, Pillai's Trace = 1.62, $F$(39, 174) = 5.28, $p < 10^{-14}$. The multivariate effect size was estimated at multivariate $\eta^2 = .54$, which implies that 54% of the variance in the canonically derived dependent variable was accounted for by the differences in the student cluster assignment, which is according to Cohen (1988) and Miles (2001) considered a large effect size. Finally, our findings were further confirmed with robust rank-based MANOVA for which the significant results were also obtained (Wilks $\Lambda_{rank} = 0.06$, $p < 10^{-15}$).

As a follow-up, a series of one-way ANOVAs with Bonferroni corrections for each of the dependent variables was conducted. The assumption of homogeneity of variance was tested using Levine's F test and it revealed that assumption of homogeneity of variance was satisfied for all but two variables (`UpdatePostTime` and `ForumSearchCount`) for which Kruskal-Wallis tests were conducted. All but three ANOVA models (for `UserLoginCount`, `CourseViewCount` and `ResourceViewCount`) were statistically significant (Table 6a), as well as both Kruskal-Wallis tests (Table 6b). The obtained effect sizes were in the range of $\eta^2 = 0.19$ for the

**Table 6**
Cluster comparison results.
1: Task-focused users, 2: Content-focused no users, 3: No users, 4: Highly intensive users, 5: Content-focused intensive users, 6: Socially-focused intensive users

**(a)**
ANOVA results. Significance level $\alpha = 0.0038\ (0.05/13)$

| Variable | Levene's | | ANOVAs | | |
|---|---|---|---|---|---|
| | $F(3,68)$ | $p$ | $F(3,68)$ | $p$ | $\eta^2$ |
| `UserLoginCount` | 2.23 | .09 | 2.05 | .12 | .08 |
| `CourseViewCount` | 0.77 | .51 | 2.04 | .12 | .08 |
| **`AssignmentViewTime`** | 1.62 | .19 | **8.18** | **.0001** | **.27** |
| **`AssignmentViewCount`** | 2.41 | .07 | **5.23** | **.003** | **.19** |
| `ResourceViewTime` | 0.31 | .82 | 4.29 | .008 | .16 |
| `ResourceViewCount` | 0.69 | .56 | 2.09 | .11 | .08 |
| **`DiscussionViewTime`** | 1.02 | .39 | **13.98** | **<.0001** | **.38** |
| **`DiscussionViewCount`** | 2.52 | .07 | **19.41** | **<.0001** | **.46** |
| **`AddPostTime`** | 0.27 | .85 | **16.77** | **<.0001** | **.43** |
| **`AddPostCount`** | 2.71 | .052 | **38.41** | **<.0001** | **.62** |
| **`UpdatePostCount`** | 1.06 | .37 | **14.15** | **<.0001** | **.38** |

**(b)**
Kruskal-Wallis results and posthoc analysis results.

| Variable | $H(3)$ | $p$ | Cluster Pair |
|---|---|---|---|
| **`ForumSearchCount`** | **22.38** | $<10^{-4}$ | 6 - 1 |
| | | | 6 - 3 |
| | | | 2 - 1 |
| | | | 2 - 3 |
| **`UpdatePostTime`** | **24.16** | $<10^{-4}$ | 6 - 1 |
| | | | 6 - 2 |
| | | | 6 - 3 |

**(c)**
ANOVA posthoc analysis results.

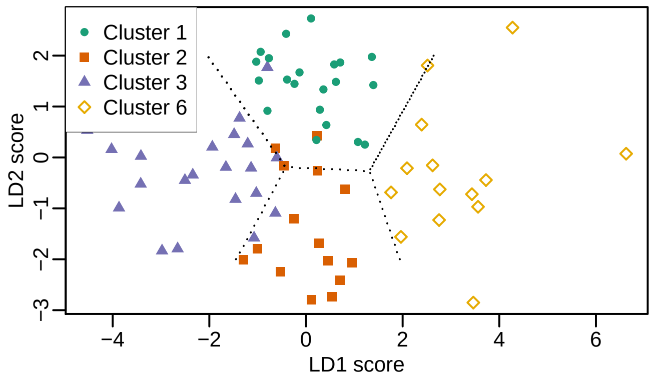
| Variable | Cluster Pair | Difference | P adjusted |
|---|---|---|---|
| `AssignmentViewTime` | 2 - 1 | 0.755 | 0.012 |
| | 2 - 3 | 1.033 | 0 |
| | 6 - 3 | 0.825 | 0.006 |
| `AssignmentViewCount` | 2 - 1 | 0.504 | 0.013 |
| | 2 - 3 | 0.597 | 0.002 |
| | 2 - 6 | 0.464 | 0.049 |
| `DiscussionViewTime` | 6 - 1 | 1.050 | 0.001 |
| | 2 - 3 | 1.094 | 0 |
| | 6 - 3 | 1.628 | 0 |
| `DiscussionViewCount` | 1 - 3 | 0.538 | 0.007 |
| | 6 - 1 | 0.783 | 0 |
| | 2 - 3 | 0.855 | 0 |
| | 6 - 3 | 1.321 | 0 |
| `AddPostTime` | 1 - 3 | 1.000 | 0 |
| | 6 - 2 | 1.059 | 0 |
| | 6 - 3 | 1.501 | 0 |
| `AddPostCount` | 1 - 2 | 0.661 | 0 |
| | 1 - 3 | 1.024 | 0 |
| | 2 - 3 | 0.363 | 0.036 |
| | 6 - 2 | 0.841 | 0 |
| | 6 - 3 | 1.204 | 0 |
| `UpdatePostCount` | 6 - 1 | 2.107 | 0 |
| | 6 - 2 | 2.768 | 0 |
| | 6 - 3 | 2.692 | 0 |

number of assignment views to $\eta^2 = 0.62$ for the total number of posted messages. All these are all considered large effect sizes (Cohen, 1988; Miles, 2001). Significant ANOVA models were followed by Tukey's HSD analysis for the pairwise comparison between clusters (Table 6c), and significant Kruskal-Wallis tests were followed by a pairwise Kruskal-Wallis test with Bonferroni correction (Table 6b).

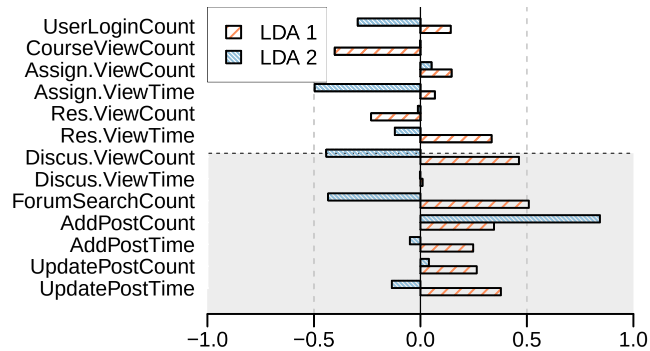
In addition to the univariate analyses, a discriminatory factor analysis (DFA) was conducted to check for the multivariate differences among the clusters in terms of students' technology use. As the four clusters were selected for the analysis, the DFA produced three discriminant functions (LD1-3) whose standardized loadings are shown in Table 7 and Fig. 4b. As Fig. 4a shows, four clusters are reasonably well separated based on the first two discriminant functions, which accounted for 69% and 22% of the variability in students' cluster assignments, respectively. Coefficients for the first discriminant function (LD1) are mostly positive, with only negative coefficients for the number of course and resource views. In terms of LD1, clusters three (no users) and six (socially-focused intensive users) represented two extremes, while clusters one (task-focused users) and two (content-focused no users) were in the middle. On the other hand, coefficients for the second discriminant function (LD2) were mostly negative, except the positive coefficients for total time spent on assignments, posting messages and updating messages. Here a opposite trend was observed, with students from clusters one (task-focused users) and two (content-focused no users) being the extremes, and students in clusters three (no users) and six (socially-focused intensive users) being in the middle.

**Table 7**
Standardized coefficients of discriminant functions

| Variable | LD1 | LD2 | LD3 |
|---|---|---|---|
| `UserLoginCount` | 0.14 | -0.3 | 0.24 |
| `CourseViewCount` | -0.4 | -0.001 | 0.14 |
| `AssignmentViewCount` | 0.15 | 0.052 | -0.74 |
| `AssignmentViewTime` | 0.068 | -0.5 | 0.08 |
| `ResourceViewCount` | -0.23 | -0.012 | 0.48 |
| `ResourceViewTime` | 0.33 | -0.12 | -0.42 |
| `DiscussionViewCount` | 0.46 | -0.44 | -0.26 |
| `DiscussionViewTime` | 0.01 | -0.002 | 0.07 |
| `ForumSearchCount` | 0.51 | -0.43 | -0.019 |
| `AddPostCount` | 0.35 | 0.84 | -0.44 |
| `AddPostTime` | 0.25 | -0.05 | 0.28 |
| `UpdatePostCount` | 0.26 | 0.04 | 0.32 |
| `UpdatePostTime` | 0.38 | -0.14 | 0.44 |
| Variance Explained | 0.69 | 0.22 | 0.091 |




**(a)** Scores for the first two discriminant functions. The DFA decision boundaries are shown with the dotted lines.




**(b)** Loadings of the clustering variables on the first two discriminant functions.

**Fig. 4** Results of the discriminant function analysis for the multivariate differences between clusters in terms of the technology use.
1: Task-focused users, 2: Content-focused no users, 3: No users, 4: Highly intensive users, 5: Content-focused intensive users, 6: Socially-focused intensive users

### *5.3. Analysis of cognitive presence*

In order to check for the cluster differences in the levels of cognitive presence, a one-way multivariate analysis of variance (MANOVA) was conducted. The independent variable (IV) was the students' cluster assignment and dependent variables (DVs) were the counts of messages in the four different phases of cognitive presence (i.e., triggering event, exploration, integration, and resolution) and the count of messages without traces of cognitive presence (coded as "other"). The descriptive statistics for each of the dependent variables are shown in Table 8.

Before running the MANOVA, cluster four (highly intensive users) with only three students was removed from our analysis, as the MANOVA requires that each condition has more subjects than the dependent variables (Tabachnick and Fidell, 2007) – in this case more than five students as there are five measures of cognitive presence. After the log-transformation of the data, Box's M test for the homogeneity of covariances using the suggested significance level $\alpha = 0.001$ (Tabachnick and Fidell, 2007) indicated that covariance matrices were not significantly different, Box's $M = 114.1,\ p = 0.005$.

A one-way MANOVA was performed to test the difference between the clusters with respect to the number of messages in the four phases of cognitive presence, as well as the number of non-cognitive messages. A statistically significant MANOVA effect was obtained, Pillai's Trace = .76, $F(20, 288) = 3.35$, $p < 10^{-5}$. Similar statistically significant results were confirmed by using a robust rank-based MANOVA, Wilks $\Lambda_{rank} = 0.34,\ p < 10^{-7}$. A multivariate effect size was estimated at multivariate $\eta^2 = .19$, which implies that 19% of the variance in the canonically derived dependent variable was accounted for by the differences in the student cluster assignment, and is considered a large effect size (Cohen, 1988; Miles, 2001).

Before conducting a series of follow-up ANOVAs, the assumption of homogeneity of variance was tested for all five dependent measures using Levene's F test, which was found not significant at $p = 0.05$. A statistically significant difference among student clusters was observed in terms of the number of exploration, integration, and non-cognitive messages, and marginal significance for the number of triggering events (Table 9a). The multivariate $\eta^2$ effect sizes were in the range from 0.27 to 0.32 which are considered large effect sizes (Cohen, 1988; Miles, 2001). Following the significant results of the ANOVA analyses, a series of post-hoc analyses using Tukey's HSD test was performed and Table 9b shows the pairs of clusters where statistically significant differences were observed. In terms of the number of exploration and non-cognitive messages, the students from clusters one (task-focused users) and six (socially-focused intensive users) had a significantly higher number of messages posted than the students from clusters two (content-focused no users) and three (no users). Finally, with respect to the number of integration messages, students from cluster three had a significantly fewer integration messages posted than the students from clusters one (task-focused users), two (content-focused no users), and six (socially-focused intensive users).

A discriminatory factor analysis (DFA) was also performed to assess the multivariate differences between the student clusters. Given that our independent variable was associated with the five levels (due to the elimination of cluster four (highly intensive users) from the analysis), four discriminant functions were discovered accounting for 90%, 8.6%, 0.8% and 0.5%

**Table 10**
Standardized coefficients of discriminant functions

| Variable | LD1 | LD2 | LD3 | LD4 |
|---|---|---|---|---|
| `TriggeringEventCount` | 0.1 | 0.44 | 0.57 | -0.87 |
| `ExplorationCount` | 0.72 | -0.13 | -0.9 | 0.11 |
| `IntegrationCount` | 0.67 | -0.65 | 0.21 | 0.013 |
| `ResolutionCount` | 0.16 | 0.034 | 0.35 | -0.39 |
| `OtherCount` | 0.49 | 0.62 | 0.3 | 0.54 |
| Variance Explained | 0.9 | 0.086 | 0.008 | 0.005 |

**Table 8**
Descriptive statistics of the dependent variable raw scores: Median (Mdn), 25th (Q1) and 75th(Q3) percentiles
1: Task-focused users, 2: Content-focused no users, 3: No users, 4: Highly intensive users, 5: Content-focused intensive users, 6: Socially-focused intensive users

| | | Cluster 1 (N=21) | | Cluster 2 (N=15) | | Cluster 3 (N=22) | | Cluster 4 (N=3) | | Cluster 5 (N=6) | | Cluster 6 (N=14) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| # | Variable | *Mdn* | *(Q1, Q3)* | *Mdn* | *(Q1, Q3)* | *Mdn* | *(Q1, Q3)* | *Mdn* | *(Q1, Q3)* | *Mdn* | *(Q1, Q3)* | *Mdn* | *(Q1, Q3)* |
| 1 | `TriggeringEventCount` | 3 | (1, 5) | 2 | (1, 2) | 2 | (1, 3.8) | 2 | (1.5, 9.5) | 3.5 | (2, 6.5) | 4.5 | (2.2, 5) |
| 2 | `ExplorationCount` | 10 | (6, 11) | 5 | (3.5, 8) | 4.5 | (3, 6) | 10 | (6.5, 12) | 9 | (4.8, 12) | 11.5 | (8.5, 18) |
| 3 | `IntegrationCount` | 8 | (4, 11) | 6 | (4.5, 8) | 3 | (2, 3) | 6 | (6, 9.5) | 4 | (4, 8.5) | 8 | (5.2, 13) |
| 4 | `ResolutionCount` | 1 | (0, 2) | 1 | (0, 1.5) | 0 | (0, 1) | 1 | (0.5, 2) | 1 | (0.25, 2.5) | 1.5 | (0, 2.8) |
| 5 | `OtherCount` | 9 | (5, 11) | 4 | (3, 6) | 4.5 | (3.2, 7.5) | 15 | (10, 16) | 9 | (9, 9.8) | 9.5 | (5.2, 14) |

**Table 9**
Cognitive presence analysis results.
1: Task-focused users, 2: Content-focused no users, 3: No users, 4: Highly intensive users, 5: Content-focused intensive users, 6: Socially-focused intensive users

**(a)**
ANOVA Results. Significance level $\alpha = 0.01$ $(0.05/5)$

| Variable | Levene's | | ANOVAs | | |
|---|---|---|---|---|---|
| | $F(4,73)$ | $p$ | $F(4,73)$ | $p$ | $\eta^2$ |
| `Trig. Event Count` | 1.85 | .13 | 2.69 | 0.038 | .13 |
| **`Exploration Count`** | 1.14 | .34 | **7.71** | **<.0001** | **.30** |
| **`Integration Count`** | 0.66 | .62 | **8.88** | **<.00001** | **.32** |
| `Resolution Count` | 1.09 | .37 | 1.57 | .19 | .08 |
| **`Other Count`** | 1.00 | .41 | **6.79** | **<.001** | **.27** |

**(b)**
Significant pairwise comparisons of cluster centers.

| Variable | Cluster Pair | Difference | P adjusted |
|---|---|---|---|
| `ExplorationCount` | 1 - 2 | 0.745 | 0.029 |
| | 1 - 3 | 0.889 | 0.002 |
| | 6 - 2 | 1.012 | 0.004 |
| | 6 - 3 | 1.156 | 0 |
| `IntegrationCount` | 1 - 3 | 1.194 | 0 |
| | 2 - 3 | 0.959 | 0.004 |
| | 6 - 3 | 1.288 | 0 |
| `OtherCount` | 1 - 2 | 0.843 | 0.007 |
| | 1 - 3 | 0.763 | 0.007 |
| | 6 - 2 | 0.940 | 0.006 |
| | 6 - 3 | 0.861 | 0.007 |

of the variation in students' cluster assignments, respectively. Their coefficients are shown in Table 10 and in Fig. 5b. The coefficients of the first discriminant function (LD1) indicate that LD1 affected all five dependent variables in the same way, with the focus on integration and exploration messages. However, the second discriminant function (LD2) affected integration and exploration messages in the opposite direction than triggering event, resolution and non-cognitive messages. The coefficients for integration, triggering event, and non-cognitive messages are much bigger than those for the exploration and resolution messages, indicating their much stronger significance for the LD2 scores.

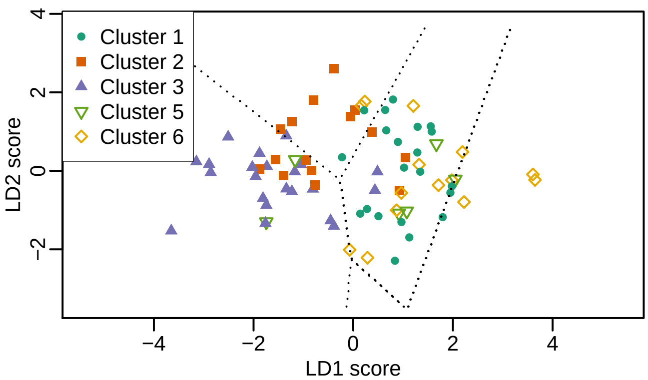
The students' scores in the first two discriminant functions in Fig. 5a were less separated compared to the DFA analysis of the technology use reported in Fig. 4a. The scores in the first discriminant function (Fig. 5c) show that the students from clusters one (task-focused users) and six (socially-focused intensive users) had similar scores; likewise, the students from clusters two (content-focused no users) and three (no users) with the students from cluster three having somewhat lower scores. Interestingly, students from cluster five (content-focused intensive users) had very disperse scores for the first discriminant function. The second discriminant function (Fig. 5d) reveals mixed scores, with the students from cluster five (content-focused intensive users) and two (content-focused no users) in general having somewhat lower and higher scores, respectively.

## 6. Discussion

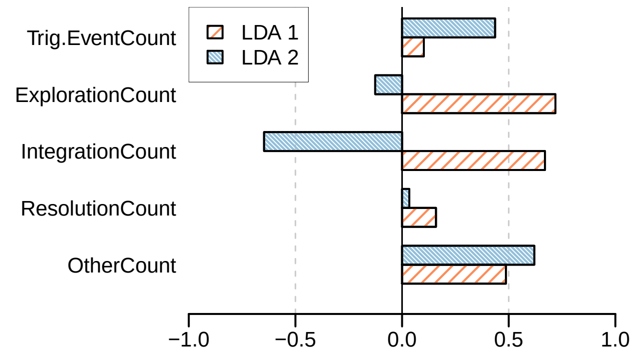
### *6.1. Research question 1: technology-use profiles withing communities of inquiry*

Based on the results of clustering, we can confirm the existence of the different technology-use profiles within the communities of inquiry. Generally speaking, our findings are aligned with the existing research of students' technology use, with some interesting differences which are discussed in the reminder of this section. What is particularly interesting are the magnitudes of the obtained effect sizes. Both MANOVA and the subsequent ANOVA effect sizes are all very large, suggesting important differences between students in terms of their technology use. This is especially evident for discussion-related activities, for which the obtained effect sizes are particularly large. This suggest stronger differences between students' technology-use profiles with respect to the use of asynchronous online discussions than to the use of the course content.

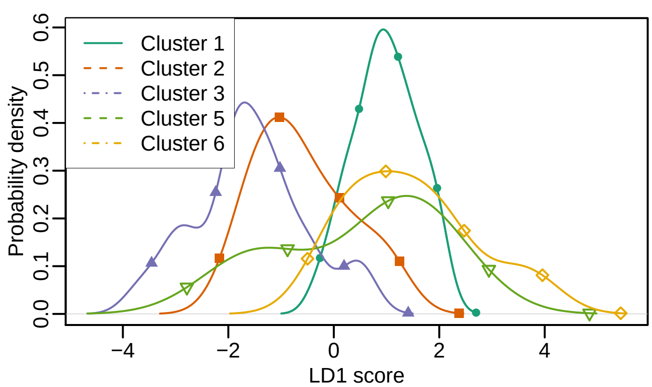
The DFA results (Fig. 4) show that the first linear discriminant (LD1) – which accounts for almost 70% of the variability in students' cluster assignments – could be best described as *the amount of the overall engagement*, with the focus on the quality of discussion-related activity. This is aligned with the



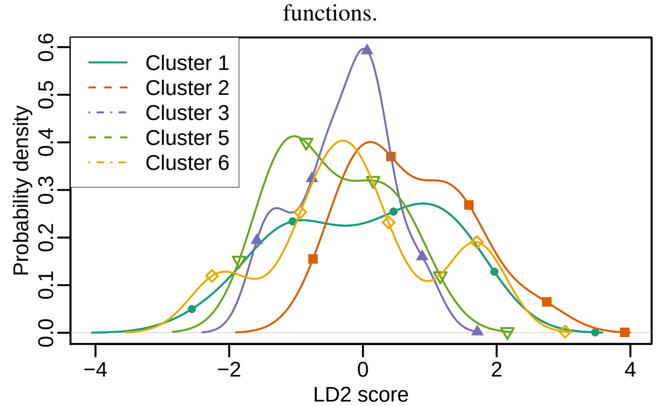

**(a)** Scores for the first two discriminant functions. LDA decision boundaries are shown with dotted lines.




**(b)** Loadings of clustering variables on the first two discriminant functions.




**(c)** Distribution of scores for the first discriminant function (LD1)




**(d)** Distribution of scores for the second discriminant function (LD2)

**Fig. 5** Results of the discriminant function analysis for the multivariate differences between clusters in the levels of cognitive presence.
1: Task-focused users, 2: Content-focused no users, 3: No users, 4: Highly intensive users, 5: Content-focused intensive users, 6: Socially-focused intensive users

previous work of Lust et al. (2013b, 2011, 2013a) and Valle and Duffy (2009) who found large differences between students in terms of the effort invested in the course. The differences in engagement were also expected, given the different conditions on which students regulate their own learning activities (Butler and Winne, 1995; Winne, 2006; Winne and Hadwin, 1998). The students from cluster six (socially-focused intensive users) have the highest LD1 scores, while the students from cluster three (no users) have the lowest scores. The students from cluster one (task-focused users) have medium scores, primarily due to their high message posting activity, while the students from cluster two (content-focused no users) also have medium scores, mostly due to their intermediate content-related activities. Likewise, the high cost of the last merging step on the clustering dendogram (Fig. 1) also points out on the substantial differences between the last two clusters. The observed differences are also consistent across all the variables suggesting that *the overall engagement with the system is the most important characteristic that defines the students' technology-use profile*.

The second discriminant function (LD2) – which accounted for 22% of the variability in the student cluster assignment – could be best described as *the focus on message posting activity* or *the preference towards discussion participation*. This is aligned with the findings of Wise et al. (2013) who identified three groups of students with different preferences towards active vs. passive participation in online discussions. The study of Dennen (2008) found that students' engagement in asynchronous online discussions was related to their perceived usefulness of learning through discussions, which is aligned with the notion of students' self-regulation of learning activities (Winne and Hadwin, 1998). Likewise, Bliuc et al. (2010) showed that students' conception of learning through discussions (i.e., cohesive vs. fragmented) was related to their approaches to learning and ultimately academic performance.

In the study reported in this paper, the students from cluster one (task-focused users) have the highest scores of LD2 due to their unusually high posting activity and the overall low engagement with the LMS. In contrast, the students from cluster two (content-focused no users) – given their average engagement with the LMS – have unusually low message posting activity. This is also evident from the cluster differences (Fig. 2) which indicate that different preferences guide students' choices when exercising their agency (Winne, 2006), in this case, towards online discussions. The students from cluster two (content-focused no users), three (no users), four (highly intensive users), and five (content-focused intensive users) clearly exhibit strong inclination towards content-related activities (assignments and resources) and passive reading of the discussion messages, while the students in cluster one (task-focused users) and six (socially-focused intensive users) exhibit higher inclination towards participation in asynchronous online discussions.

In order to summarize our findings and to make them more usable, Fig. 6 integrates all the previously described results into a coherent picture of cluster differences using two dimensions: i) the level of content-related activity (including discussion reading), and ii) the level of active participation in online discussions. We can see that students adopted different study tactics which might be related to the differences in their internal and external conditions, particularly metacognition and motivation as defined by the previous research (Lust et al., 2013b).

Previous research suggest the existence of three or four technology-use profiles, while our results indicate the existence of six different profiles. Indeed, Lust et al. (2011) points out on the possible existence of additional clusters that are harder to detect empirically due to the diminishing differences between them. On the other hand, it is possible that larger number of clusters identified in this study is just a reflection of anomalies and outliers in the data, particularly due to existence of some smaller clusters such as clusters four (highly intensive users) and five (content-focused intensive users). Concerning the relative sizes of the clusters, previous studies by Lust et al. (2011, 2013a,b) found that no-users are the largest group, and that intensive users are the smallest. This is consistent with the sizes of the clusters in our study, as cluster three (no users) and four (highly intensive users) reasonably well correspond to no-users and intensive users, respectively. Regarding the previously discovered selective users cluster (Lust et al., 2011, 2013a,b), our study found several user profiles that might be called selective. One possible reason might be the fact that the course in our study was a graduate level course with many of the students having previous work experience and completed bachelors degrees. It is shown that more experienced students who had previous education and work experience were able to make better use of the discussion boards (Lust et al., 2011). Likewise, given that the course in our study was from a fully online program, the students were already familiar with this type of learning environments, and thus better able to manage their own learning processes.

An interesting finding by Lust et al. (2011) is that the intensive cluster was the most diverse in terms of their technology use. Our study confirms these findings. The clustering dendogram shows (Fig. 1) that clusters one (task-focused users), two (content-focused no users), three (no users), and five (content-focused intensive users) became fully connected while the three students from the cluster four (highly intensive users) remained in isolated single-element clusters. As the students in cluster four exhibited more diversified behavior, it resulted in the larger differences in their scores on the used clustering variables, which subsequently resulted in their later merging into a single cluster.

### 6.2. *Research question 2: effects of technology use on cognitive presence*

The results of the MANOVA analysis indicate that there is a significant difference between the students in different clusters in terms of their cognitive presence. We find a large effect size, as the cluster assignment accounted for 19% of the variability in the canonically derived dependent variable, which suggests an important connection between the technology use and students' cognitive presence. Furthermore, the results of the subsequent ANOVA analyses suggest that differences were strongest for the integration phase, followed by the differences for the exploration phase and with the smallest differences for the number of non-cognitive messages. The differences in the number of triggering event messages are marginally significant, which can be explained by the grading policy for discussions and their external motivation induced by the course design – which is very common in higher education (Garrison et al., 2001; Rovai, 2007; Penny and Murphy, 2009). Likewise, the lack of significance for the resolution phase is not surprising as – due to the time constraints – in most higher education settings the resolution phase is not reached (Garrison et al., 2010a, 2001).

The DFA results (Fig. 5) reveal that the first discriminant function (LD1) – which accounts for 90% of the variability between students' cluster assignments – can be summarized as *the development of cognitive presence*. LDA1 makes a strong distinction between students from cluster one (task-focused users), five (content-focused intensive users), and six (socially-focused intensive users) on one side, and students from cluster two (content-focused no users) and three (no users) on the other side. The distribution of scores within clusters is reasonably consistent, with only students from cluster five (content-focused intensive users) scoring more diversified scores. The observed differences were expected based on the previous research (Lust et al., 2011, 2013b,a; Clarebout et al., 2013; Valle and Duffy, 2009; Wise et al., 2013), and it was not surprising that the students from less engaged clusters do not fully develop their cognitive presence. It is aligned with existing evidence of students' poor self-regulation of learning (Dunlosky and Lipko, 2007). Likewise, there is evidence suggesting that a majority of the students do not develop their cognitive presence past the exploration phase (Garrison et al., 1999).

The ANOVA results (Table 9) indicate that the students in clusters two (content-focused no users) and three (no users) have significantly fewer exploration and non-cognitive messages than students from clusters one (task-focused users) and six (socially-focused intensive users), and that students in cluster three (no users) have significantly fewer integration messages than students from clusters one (task-focused users), two (content-focused no users), and six (socially-focused intensive users). Based on this, it is evident that the lack of cognitive presence development is more emphasized for the students in cluster three (no users), which is aligned with their overall low level of the use of the LMS.

The second discriminant function (LD2) (Fig. 5b and Ta-

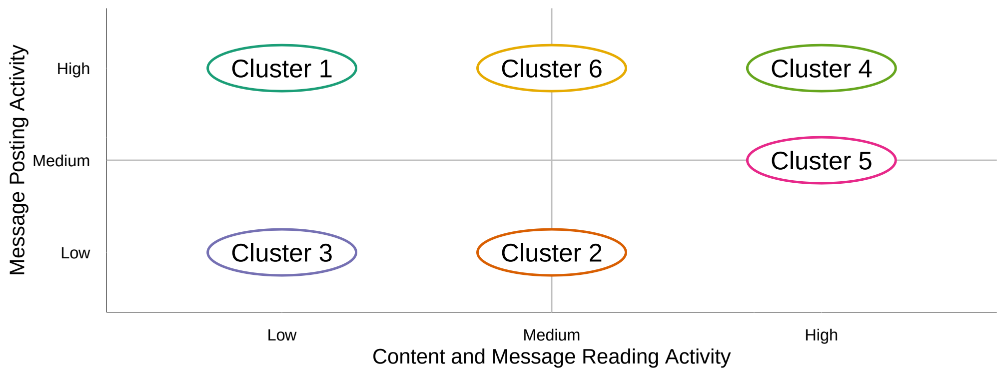



**Fig. 6** Cluster Matrix: Activity Focus and Activity Level.
1: Task-focused users, 2: Content-focused no users, 3: No users, 4: Highly intensive users, 5: Content-focused intensive users, 6: Socially-focused intensive users

ble 10) – which explain 9% of the variability of students' cluster assignment – can be best summarized as the relative *lack of integration*, given the overall level of development of cognitive presence. In general, LD2 scores of students from all the clusters are far less separated. This find is reasonable given the smaller amount of variance explained by LD2. Still, it can be seen that students from cluster five (content-focused intensive users) had slightly lower LD2 scores than other students, and that cluster one (task-focused users) and two (content-focused no users) students had slightly higher LD2 scores. Given their high engagement and clear preference towards content-related activities, one likely explanation could be their lower perceived usefulness of learning through discussions (Dennen, 2008) and more fragmented conception of learning through discussions (Bliuc et al., 2010).

What is very interesting is that *we do not observe a clear connection between the overall engagement and the development of cognitive presence*. Although students from cluster two (content-focused no users) and three (no users) have a lower level of the use of the LMS and had lower levels of cognitive presence, this is not the case for the students in cluster one (task-focused users). Similar results are already found in the literature. Study by Valle and Duffy (2009), concluded that adoption of different learning profiles have very little impact on the final course grade, with even the students who struggled with online environment finishing the course with similar final grades. Likewise, the study by Wise et al. (2013) found no difference in terms of the final grades between students with different profile of participation in online discussions. Given that the course under investigation in our study is graduate level course, it is likely that students in our study posses higher meta-cognitive awareness and motivation (Perkins, 1985; Winne, 2006; Winne and Hadwin, 1998), as well as broader domain knowledge and skills (Lust et al., 2011).

### 6.3. Cluster interpretations

This section provides a more detailed interpretations of the identified clusters by building on the existing research on the CoI model and learning technology use and, and important constructs that are identified in the previous studies (e.g., motivation, self-efficacy, self-regulated learning, and goal orientation). The implications that can inform future research and practice are also discussed.

#### 6.3.1. Cluster one (Task-focused users)

The students from cluster one show a below average use of the LMS, except for the number of posted messages and time spent on writing new messages. Even though they have above average message posting activity, their reading activity is below average. They resemble very closely task focused, "get it done" students from the Valle and Duffy (2009) study and they tend to spend only necessary time for completion of the course. As Valle and Duffy explain, they seem to be in a hurry to complete the course, but still show positive study strategies comparable to mastery-oriented and self-driven students. Given that a large number of the students in our study are also working full time, this might be one of the reasons for the observed behavior of this cluster of students.

Regardless of the LMS use, the levels of cognitive presence are very high for the students in this cluster. This further supports findings by Valle and Duffy (2009) of the use of positive study strategies. The students from cluster one have significantly more triggering events and non-cognitive messages than the students from clusters two (content-focused no users) and three (no users), and significantly more integration messages than the students from cluster three (no users). They were very similar to the students from cluster six (socially-focused intensive users), while being significantly less active overall. Similar results are reported by Lust et al. (2011) who did not find significant differences between intensive and incoherent users in terms of their academic performance. Given that quality rather than quantity of the LMS use is indicative of students' metacognitive skills (Clarebout et al., 2013) (also consistent with the learning strategy use as suggested by Kuhn (1995)), it seems likely that the students in cluster one possess higher metacognitive skills that enable them to control and monitor their learning activities effectively. As pointed out by Lust et al., the differences in tool-use patterns are not necessarily a problem, especially in learning environments which focus on active, self-controlled learning. However, if its affects students performance, it might suggest that those students are not profiting from the whole range of tools that are available to them (Lust et al., 2011).

Looking at the cluster sizes (Fig. 1) we can see that cluster one is relatively large (N=21) – suggesting that a considerable number of students were enough meta cognitively skillful in the use of asynchronous online discussions. This is aligned with the results of Valle and Duffy who also looked at effectiveness of students in online learning settings. A possible explanation is that their previous experience with this particular learning modality – given that the students were enrolled in a fully online condition – enables them to be more effective with their learning. This also explains why studies that looked at metacognitive skills of students in blended learning courses (Lust et al., 2013a) showed that a majority of students did not adequately regulate their tool use. Still, this interpretation warrants more empirical and theoretical attention in future studies.

The fact that the students from cluster one spend very little time reading other students' writings is very interesting. This might suggest that there was much of "scanning activity" by the students. That type of activity would unlikely results in a deep comprehension (Hewitt et al., 2007). As indicated by Entwistle (2009), effective learning can be combined with both deep and surface approaches to learning. Also, students who post solely to meet the course requirements and read the bare minimum of other students' writings are less likely to perceive online discussions as a valuable learning activity (Dennen, 2008). However, those types of students benefit particularly by requesting the student participation in asynchronous online discussions in the course design (Wise et al., 2013). This might an important reason for their high levels of cognitive presence, which shown empirically to be shaped by teaching presence (Garrison et al., 2010b).

#### 6.3.2. *Cluster two (Content-focused no users)*

The students from cluster two are characterized by the overall average LMS use. They only show above average scores for the assignment-related variables. With respect to discussion-related activities, they spend about an average amount of time reading discussions, but seemed reluctant to contribute themselves. They are most similar to the selective and incoherent users from Lust et al. (2013b) and Lust et al. (2011) studies, respectively.

The average time spent reading discussions and the below-average time spent actively participating suggests fragmented conceptions of learning through discussions (Bliuc et al., 2010) and the lack of meta-cognitive knowledge and skills required for successful participation in online discussions (Hew et al., 2010). The study by Lust et al. (2013b) indicated that metacognitive skills were mostly influencing students tool use patterns through goal orientation. Given the focus of the students from cluster two on assignments and their similarity to selective users of Lust et al. (2013b), it seems likely that the students in cluster two have high inclination towards performance goal orientation, which is in turn shown to be related to surface approaches to learning (Phan, 2008).

With respect to the level of cognitive presence, the students from cluster two show generally lower levels of cognitive presence. The ANOVA results (Table 9) show that they have significantly lower number of integration and non-cognitive messages than students from clusters one (task-focused users) and six (socially-focused intensive users). This is also confirmed by the DFA results (Fig. 5) given the low scores of students from this cluster on the first discriminant function (LD1) (Fig. 5c). Furthermore, the high scores of cluster two on the second discriminant function (LD2) (Fig. 5d) indicates a lower proportion of integration messages relative to the number of triggering event and non-cognitive messages. This offers further evidence of the lack of skills for successful participation in online discussions and their need for better instructional support in order to overcome challenges that are associated with participation in discussions and inquiry-based learning (Cho and Kim, 2013).

#### 6.3.3. *Cluster three (No users)*

With 22 students, cluster three is the largest in size. The primary characteristics of students from this cluster is their below-average overall engagement with the LMS, as their scores for all variables (except for number of logins) are below the overall mean. Furthermore, ANOVA results indicate that for many of the variables, they have significantly lower scores than students from other clusters (Table 6). In addition, they have the lowest scores for the first discriminant function (Fig. 4) which confirms their low use of the LMS. Looking at the findings from the previous studies, the students from this cluster are most similar to no-users from the Lust et al. (2013a,b, 2011) studies and procrastinating, "minimalist in effort" students from the Valle and Duffy (2009) study. The low use of the LMS is also reflected in their lower levels of cognitive presence. The students from cluster three have significantly less exploration, integration, and non-cognitive messages than students from clusters one (task-focused users) and six (socially-focused intensive users), as well as a significantly lower number of integration messages than students from cluster two (content-focused no users). The size of this cluster and the limited cognitive presence development is aligned with the previous research by Garrison et al. (2001).

There are several possible explanations for observed characteristics of students in cluster three. One likely cause might be the lack of intrinsic motivation to engage in learning as – according to Valle and Duffy (2009) – the way in which students learn online might indicate their commitment to learning. Much like students from cluster two (content-focused no users), it is possible that students in this cluster have high levels of performance goal orientation and surface approaches to learning (Phan, 2008). Finally, it is interesting to note that the low use of the LMS is especially evident with respect to discussion-related activities. This might suggest that the students from cluster three lack metacognitive skills required for successful learning in this particular context (Valle and Duffy, 2009). Drawing on the results from Bliuc et al. (2010), it is also very likely that they follow fragmented approaches to learning in online discussions, which is also aligned with the research indicating weakness in regulation of learning by a large proportion of students (Dunlosky and Lipko, 2007; Lust et al., 2013a). As suggested by Cho and Kim (2013), the students three may also require a better instructional support for interaction with others – similar to the students from cluster two (content-focused no users).

#### 6.3.4. Cluster four (Highly intensive users)

With only three students, cluster four is the smallest in our study. Although it was excluded from the statistical analyses, it is clear that students from this cluster are characterized by a very high level of motivation and the use of the system (Fig. 2). Students from this cluster log into the LMS much more often than students from other clusters, and they also frequently check the state of the discussions and assignments. Looking at Fig. 2 and Table 4, it is clear that the most of their count measures are very high, some even several standard deviations above the overall mean (e.g., the number of course logins, assignment views, and discussion views). Due to their overall high use of the LMS, they show similarities to the intensive active students from Lust et al. (2013b) study, and the mastery oriented, "self-driven" students from the Valle and Duffy (2009) study. The results of both Lust et al. (2013b) and Valle and Duffy (2009) studies suggest that mastery goal orientation is associated with this cluster of students. In particular, Lust et al. (2013b) suggest the association with mastery-avoidance goal orientation as it can explain the overly high use of the LMS system.

The students in cluster four are also characterized by the largest number of discussion views, 6.3 and 3.5 times as many as those of the students from clusters one (task-focused users) and six (socially-focused intensive users), respectively (Table 4). They also spend most of their time reading discussions, 3.9 and 1.8 times as much as that of the students from clusters one and six, respectively. However, in terms of the average time spent on reading discussions, cluster four students spend substantially less time on average than students from clusters one and six. Students from cluster four spend on average 2.4 minutes on each discussion reading activity, while students from clusters one and six spend 3.9 and 4.7 minutes, respectively (Table 4). Therefore, they were mostly similar to the "broad listeners, reflective speakers" cluster from the study of Wise et al. (2013) who showed similar inclination towards broad reading of discussions.

With respect to active participation in the discussions, the students from cluster four write a similar number of messages as the students from cluster six (socially-focused intensive users), but spend much less time on writing their responses. The less time spent on writing messages might be caused by: i) posting many non-cognitive messages, which typically require much less effort to write (Joksimović et al., 2014), ii) posting slightly fewer integration messages, which require the most time to write, and iii) longer discussion reading time which is often related to high comprehension and depth of learning (Wise et al., 2013). The number of messages in different phases of cognitive presence are similar to those of students from cluster one and six (Table 8), with a slightly smaller number of the integration messages. The students in cluster four also have the highest number of non-cognitive messages. As Wise et al. (2013) results showed, "broad listeners, reflective speakers" were the most frequent posters who attended to almost all of peers' messages with the most posts per discussion. Thus, the large number of non-cognitive messages – which are typically acknowledgments of others' contributions – can be explained by their high attendance to postings of other students. Although our data do not warrant drawing conclusive interpretations, it is likely that the intensive discussion reading – coupled with lower numbers of integration messages and large numbers of non-cognitive messages – is due to the high levels of motivation and lower levels of meta-cognitive skills for online discussions. However, this needs to be further investigated in the future studies.

This group is the most diverse in terms of their technology use, as indicated by the late merging of this cluster in the dendogram (Fig. 1). This is aligned with the findings of Lust et al. (2011) who also found the highest divergence in the most engaged group. The diversity in technology-use is also shown to be associated with higher meta-cognitive activity (Lust et al., 2013a), as well as to be an indicator of metacognitive-monitoring (Hadwin et al., 2007).

#### 6.3.5. Clusters five (Content-focused intensive users)

Cluster five is mainly characterized by the focus on the course content and passive reading of online discussions, much like the students from cluster two (content-focused no users); however, with the overall much higher use of the LMS. The students from cluster five have around average number of LMS logins and course views, coupled with high number of assignment and resource views. The students from this cluster also spend more time on assignments and static resources than the students from any other cluster. They also spend more time reading discussions than most of the students (i.e., all students except students from cluster four). However, their active participation in the discussions through message posting is just slightly above the overall mean value.

Similarly to cluster two students (content-focused no users), students from cluster five show similarity with incoherent users from the Lust et al. (2011) study and selective users from the Lust et al. (2013b) study – given by their clear inclination towards the use of static course content and passive reading of online discussions. The Lust et al. (2011) study found no differences between the incoherent users and the intensive users, and our study showed similar results (i.e., posthoc analysis found no difference between cluster five and any other cluster, as shown on Table 9b).

Based on the Lust et al. (2013b) results, the students from cluster five likely have higher levels of performance-approach orientation than other high technology users (i.e., socially-focused intensive users from cluster six). Similarly, they likely have higher levels of mastery-avoidance orientation than students from clusters two (content-focused no users) and three (no users). This could explain their focus on static course information and discussion "consumption" rather than active participation. Given their reluctance to participate, it is likely that those students lack skills that are required for successful participation in the discussions (Hew et al., 2010). Drawing on Bliuc et al. (2010) results, it is likely that they have high levels of fragmented conceptions of learning in asynchronous online discussions. Thus, the cluster five students might need better instructional support and scaffolding in order to successfully participate in asynchronous online discussions (Cho and Kim, 2013).

The median values of the numbers of messages in each phase of cognitive presence (Table 8) indicate that the students from cluster five have a similar number of messages as the students from clusters one (task-focused users) and six (socially-focused intensive users), except for the lower number of integration messages. This is also visible in the DFA results, as the LD1 scores of the students in cluster five are similar to those of the students from clusters one and six, but more disperse. The lower number of integration messages also reflects on their LD2 scores, which are the lowest for the students in this cluster. Still, as Tukey HSD analysis do not indicates any statistically significant difference from the students from other clusters, we can not draw conclusive inferences from the observed data and future research to further examine this cluster is warranted.

#### 6.3.6. *Cluster six (Socially-focused intensive users)*

Students from cluster six are characterized by an average content-related activity and above average discussion-related activity. Their scores on all discussion-related variables (except the number of discussion views) are around one standard deviation above the overall mean values – indicating a strong commitment to learning through asynchronous online discussions. By having a limited number of sessions in which they spent a significant amount of time, they show a similarity with "concentrated listeners, integrated talkers" from Wise et al. (2013) which suggest the depth of their reading activities. However, they also show a certain similarity with the selective users from Lust et al. (2013b) study, as they are clearly inclined towards learning through online discussions. Based on this, it is likely that those students have cohesive conceptions of learning through discussions (Bliuc et al., 2010), as well as mastery goal orientation. It is interesting to note that they have a higher number of forum searches, suggesting a more strategy approach to the use of online discussions.

The development of cognitive presence of the students from cluster six is characterized by the high levels of cognitive presence, as shown by the ANOVA (Table 9) and DFA (Fig. 5) analyses. They have a significantly more exploration, integration, and non-cognitive messages than the students from clusters three, as well as higher number of exploration and non-cognitive messages than students from cluster two (content-focused no users). In addition, their LD1 scores are very high, confirming their overall high development of cognitive presence. As such, those students might be good candidates for student moderators that are given a responsibility to guide discussions in the productive directions and to assist other students in their own learning (Schellens et al., 2007). Given that student-centered discussions are shown to better foster the development of cognitive presence than instructor-centered discussions (Schrire, 2006), this seems as one promising direction for further research that warrants a further empirical examination.

### 6.4. *Limitations*

The most important limitations of this study are related to its internal validity as – given its correlational nature – the claims about causality are not truly possible like in the case of randomized controlled trials. Likewise, the sample of 81 students – even though it consists of the six offerings of a course – is still small (although on a higher end of the related studies about communities of inquiry that used quantitative content analysis) and could be affecting the validity of our findings. The data originated from a single graduate-level course at a single university, so external validity of our findings could be potentially compromised by the specifics of the adopted pedagogical approach in the target course. Furthermore, despite the same course design and organization, the variations in cohort sizes between different course offers could also potentially influence the student learning activities by affecting the climate and overall volume of online discussions which in turn will have an impact on the clustering results. With respect to the construct validity, cognitive presence construct – which is a latent construct by the definition – is measured only through content analysis of discussion transcripts. Finally, the calculation of time duration measures for different activities are approximations which could be affected by the current limitations in tracking student activities in the LMS systems.

In order to define technology-use profiles, we adopted clustering techniques, as commonly done in the related studies that looked at learning technology-use profiles. However, clustering is an unsupervised machine learning technique and inherently subjective. Given that there are no upfront right and wrong clustering solutions, this leaves a space for the subjectivity in the interpretation of the cluster findings which may or may not affect the final outcomes of the study. In order to interpret our clustering solutions, we looked at the existing literature which provided necessary foundation for cluster interpretation. However, this brings certain challenges, as the reported findings might not be applicable in the context of the study presented in this paper. In future studies, the explicit use of standard instruments for measuring goal orientation, motivation, and other constructs will be used to provide more empirical evidence for cluster interpretations.

## 7. Conclusions

The analysis of the clustering process and the final clusters reported in this paper showed several interesting implications for both the educational practice and research on the community of inquiry. Aligned with the previous studies (Clarebout et al., 2013; Lust et al., 2012), our results indicate that the availability of different tools in a learning environment is not enough for their successful use. As indicated by Perkins (1985), students need to be sufficiently meta-cognitively capable, skillful, and motivated to use available tools. The preference towards static content or discussions and different use of the available tools suggests a need for different instructional interventions and support for different groups of students. Students that are reluctant to participate in online discussions or who have performance goal orientation might require more detailed instructions on how to productively participate in discussions (Hew et al., 2010), or access to various types of contextual aids, such as access to different static educational resources (Azevedo, 2005). As pointed out by Wise et al. (2013, p. 340): *"Students who are oriented toward mastery and see discussions as vehicles to support this goal are likely to participate in productive ways. In contrast, for students oriented toward performance goals, explicitly embedding desirable participation behaviors in the activity requirements and assessment scheme can help encourage more productive listening and speaking"*. Other students, such as students from cluster three (no users) might require more motivational support, which is aligned with the suggestions of Valle and Duffy (2009).

In terms of the cognitive presence development, our results indicate large differences among students in terms of their cognitive presence (multivariate $\eta^2 = 0.19$). However, low use of the LMS is not necessarily indicative of poor cognitive development. Our results suggest that the quality of activity is more important than quantity as highly meta-cognitively skilled students – such as students from cluster one (task-focused users) – can be equally successful as more engaged students (i.e., cluster four, five and six). This has also been suggested by Valle and Duffy (2009), Lust et al. (2011), and Clarebout et al. (2013) and our results provide further evidence for this.

Finally, our results suggest that hierarchical clustering can be successfully used for understanding the differences between students in online learning contexts. It is likely that this approach could be used in similar studies and to further validate our findings in other contexts. There are several reasons why hierarchical clustering appears appropriate for this types of studies: i) the number of clusters is in general very small which makes the dendogram analysis manageable and practical, ii) the analysis of the order of cluster merges allows for observing similarities between clusters, thus giving us a view at the factors behind the clustering process, and iii) it can be performed well for small data sets, as it does not depend on the statistical properties of the large data sets like other popular algorithms such as K-means or EM (Abbas, 2008; Hastie et al., 2013).

## References


Abbas, O. A., Jul. 2008. Comparisons between data clustering algorithms. The International Arab Journal of Information Technology 5 (3), 320–325.

Akyol, Z., Arbaugh, J. B., Cleveland-Innes, M., Garrison, D. R., Ice, P., Richardson, J. C., Swan, K., 2009. A response to the review of the community of inquiry framework. Journal of Distance Education 23 (2), 123–135.

Akyol, Z., Garrison, D. R., 2011a. Assessing metacognition in an online community of inquiry. The Internet and Higher Education 14 (3), 183–190.

Akyol, Z., Garrison, D. R., 2011b. Understanding cognitive presence in an online and blended community of inquiry: Assessing outcomes and processes for deep approaches to learning. British Journal of Educational Technology 42 (2), 233–250.

Anderson, T., Dron, J., 2010. Three generations of distance education pedagogy. The International Review of Research in Open and Distance Learning 12 (3), 80–97.

Anderson, T., Rourke, L., Garrison, D. R., Archer, W., 2001. Assessing teaching presence in a computer conferencing context. Journal of Asynchronous Learning Networks 5, 1–17.

Arbaugh, J., Cleveland-Innes, M., Diaz, S. R., Garrison, D. R., Ice, P., Richardson, J. C., Swan, K. P., 2008. Developing a community of inquiry instrument: Testing a measure of the community of inquiry framework using a multi-institutional sample. The Internet and Higher Education 11 (3–4), 133–136.

Azevedo, R., 2005. Using hypermedia as a metacognitive tool for enhancing student learning? the role of self-regulated learning. Educational Psychologist 40 (4), 199–209.

Bjork, R. A., Dunlosky, J., Kornell, N., 2013. Self-regulated learning: beliefs, techniques, and illusions. Annual review of psychology 64, 417–444.

Bliuc, A.-M., Ellis, R., Goodyear, P., Piggott, L., 2010. Learning through face-to-face and online discussions: Associations between students' conceptions, approaches and academic performance in political science. British Journal of Educational Technology 41 (3), 512–524.

Bock, R. D., 1985. Multivariate Statistical Methods in Behavioral Research. Scientific Software International.

Bray, J. H., Maxwell, S. E., 1985. Multivariate Analysis of Variance. SAGE.

Buckingham Shum, S., Ferguson, R., 2012. Social learning analytics. Journal of Educational Technology & Society 15 (3), 3–26.

Butler, D. L., Winne, P. H., Sep. 1995. Feedback and self-regulated learning: A theoretical synthesis. Review of Educational Research 65 (3), 245–281.

Cho, M.-H., Kim, B. J., Apr. 2013. Students' self-regulation for interaction with others in online learning environments. The Internet and Higher Education 17, 69–75.

Clarebout, G., Elen, J., Collazo, N. A. J., Lust, G., Jiang, L., 2013. Metacognition and the use of tools. In: Azevedo, R., Aleven, V. (Eds.), International Handbook of Metacognition and Learning Technologies. No. 28 in Springer International Handbooks of Education. Springer New York, pp. 187–195.

Cohen, J., 1988. The analysis of variance. In: Statistical power analysis for the behavioral sciences. L. Erlbaum Associates, Hillsdale, N.J., pp. 273–406.

Dawson, S., Tan, J. P. L., McWilliam, E., 2011. Measuring creative potential: Using social network analysis to monitor a learners' creative capacity. Australasian Journal of Educational Technology 27 (6), 924–942.

Dennen, V. P., Jul. 2008. Pedagogical lurking: Student engagement in non-posting discussion behavior. Comput. Hum. Behav. 24 (4), 1624–1633.

Dewey, J., 1910. How We Think. D.C. Heath & Company.

Dunlosky, J., Lipko, A. R., 2007. Metacomprehension a brief history and how to improve its accuracy. Current Directions in Psychological Science 16 (4), 228–232.

Ellis, R. A., Marcus, G., Taylor, R., 2005. Learning through inquiry: student difficulties with online course-based material. Journal of Computer Assisted Learning 21 (4), 239–252.

Entwistle, N., Sep. 2009. Teaching for Understanding at University: Deep Approaches and Distinctive Ways of Thinking. Palgrave Macmillan, Basingstoke, Hampshire ; New York.

Field, A. P., Miles, J., Field, Z., 2012. Discovering statistics using R. SAGE, Los Angeles; London.

Garrison, D. R., 2007. Online community of inquiry review: Social, cognitive, and teaching presence issues. Journal of Asynchronous Learning Networks 11 (1), 61–72.

Garrison, D. R., 2011. E-Learning in the 21st Century: A Framework for Research and Practice, 2nd Edition. Routledge, New York.

Garrison, D. R., Akyol, Z., 2013. Toward the development of a metacognition construct for communities of inquiry. The Internet and Higher Education 17, 84–89.

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 (2–3), 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.

Garrison, D. R., Anderson, T., Archer, W., 2010a. The first decade of the community of inquiry framework: A retrospective. The Internet and Higher Education 13 (1–2), 5–9.

Garrison, D. R., Arbaugh, J., 2007. Researching the community of inquiry framework: Review, issues, and future directions. The Internet and Higher Education 10 (3), 157–172.

Garrison, D. R., Cleveland-Innes, M., 2005. Facilitating cognitive presence in online learning: Interaction is not enough. American Journal of Distance Education 19 (3), 133–148.

Garrison, R., Cleveland-Innes, M., Fung, T. S., 2010b. Exploring causal relationships among teaching, cognitive and social presence: Student perceptions of the community of inquiry framework. The Internet and Higher Education 13 (1–2), 31–36.

Gonyea, R. M., 2005. Self-reported data in institutional research: Review and recommendations. New Directions for Institutional Research 2005 (127), 73–89.
URL http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=18442595&site=ehost-live

Gorsky, P., Caspi, A., Blau, I., Vine, Y., Billet, A., 2011. Toward a CoI population parameter: The impact of unit (sentence vs. message) on the results of quantitative content analysis. The International Review of Research in Open and Distance Learning 13 (1), 17–37.

Hadwin, A. F., Nesbit, J. C., Jamieson-Noel, D., Code, J., Winne, P. H., Dec. 2007. Examining trace data to explore self-regulated learning. Metacognition and Learning 2 (2-3), 107–124.
URL http://link.springer.com/article/10.1007/s11409-007-9016-7

Hastie, T. J., Tibshirani, R. J., Friedman, J. H., 2013. The elements of statistical learning: data mining, inference, and prediction. Springer, New York, NY.

Hew, K. F., Cheung, W. S., Ng, C. S. L., Nov. 2010. Student contribution in asynchronous online discussion: a review of the research and empirical exploration. Instructional Science 38 (6), 571–606.

Hewitt, J., Brett, C., Peters, V., Nov. 2007. Scan rate: A new metric for the analysis of reading behaviors in asynchronous computer conferencing environments. American Journal of Distance Education 21 (4), 215–231.

Joksimović, S., Gašević, D., Kovanović, V., Adesope, O., Hatala, M., Jul. 2014. Psychological characteristics in cognitive presence of communities of inquiry: A linguistic analysis of online discussions. The Internet and Higher Education 22.
URL http://www.sciencedirect.com/science/article/pii/S1096751614000189

Kanuka, H., 2011. Interaction and the online distance classroom: Do instructional methods effect the quality of interaction? Journal of Computing in Higher Education 23 (2), 143–156.

Kuhn, D., May 1995. Microgenetic study of change: What has it told us? Psychological Science 6 (3), 133–139.
URL http://pss.sagepub.com.proxy.lib.sfu.ca/content/6/3/133

Lust, G., Elen, J., Clarebout, G., 2013a. Regulation of tool-use within a blended course: Student differences and performance effects. Computers & Education 60 (1), 385–395.

Lust, G., Elen, J., Clarebout, G., 2013b. Students' tool-use within a web enhanced course: Explanatory mechanisms of students' tool-use pattern. Computers in Human Behavior 29 (5), 2013–2021.

Lust, G., Juarez Collazo, N. A., Elen, J., Clarebout, G., 2012. Content management systems: Enriched learning opportunities for all? Computers in Human Behavior 28 (3), 795–808.

Lust, G., Vandewaetere, M., Ceulemans, E., Elen, J., Clarebout, G., 2011. Tool-use in a blended undergraduate course: In search of user profiles. Computers & Education 57 (3), 2135–2144.

Macfadyen, L. P., Dawson, S., 2010. Mining LMS data to develop an "early warning system" for educators: A proof of concept. Computers & Education 54 (2), 588–599.

McGill, T. J., Klobas, J. E., 2009. A task–technology fit view of learning management system impact. Computers & Education 52 (2), 496–508.

Miles, J., Mar. 2001. Applying Regression and Correlation: A Guide for Students and Researchers, 1st Edition. SAGE Publications Ltd, London ; Thousand Oaks, Calif.

Morris, L. V., Finnegan, C., Wu, S.-S., 2005. Tracking student behavior, persistence, and achievement in online courses. The Internet and Higher Education 8 (3), 221–231.

Murphy, E., Coleman, E., 2004. Graduate students' experiences of challenges in online asynchronous discussions. Canadian Journal of Learning and Technology / La revue canadienne de l'apprentissage et de la technologie 30 (2).

Nath, R., Pavur, R., 1985. A new statistic in the one-way multivariate analysis of variance. Computational Statistics & Data Analysis 2 (4), 297–315.

Penny, L., Murphy, E., 2009. Rubrics for designing and evaluating online asynchronous discussions. British Journal of Educational Technology 40 (5), 804–820.
URL http://onlinelibrary.wiley.com/doi/10.1111/j.1467-8535.2008.00895.x/abstract

Perkins, D. N., 1985. The fingertip effect: How information-processing technology shapes thinking. Educational Researcher 14 (7), 11–17.

Phan, H. P., 2008. Unifying different theories of learning: Theoretical framework and empirical evidence. Educational Psychology 28 (3), 325–340.

Romero, C., Ventura, S., García, E., 2008. Data mining in course management systems: Moodle case study and tutorial. Computers & Education 51 (1), 368–384.

Rourke, L., Anderson, T., 2004. Validity in quantitative content analysis. Educational Technology Research and Development 52 (1), 5–18.

Rourke, L., Anderson, T., Garrison, D. R., Archer, W., 1999. Assessing social presence in asynchronous text-based computer conferencing. The Journal of Distance Education 14 (2), 50–71.

Rovai, A. P., 2007. Facilitating online discussions effectively. The Internet and Higher Education 10 (1), 77–88.
URL http://www.sciencedirect.com/science/article/pii/S1096751606000704

Rubin, B., Fernandes, R., Avgerinou, M. D., 2013. The effects of technology on the community of inquiry and satisfaction with online courses. The Internet and Higher Education 17, 48–57.

Sawyer, R. K., 2006. The Cambridge handbook of the learning sciences. Cambridge University Press, Cambridge; New York.

Schellens, T., Keer, H. V., Wever, B. D., Valcke, M., Sep. 2007. Scripting by assigning roles: Does it improve knowledge construction in asynchronous discussion groups? International Journal of Computer-Supported Collaborative Learning 2 (2-3), 225–246.
URL http://link.springer.com/article/10.1007/s11412-007-9016-2

Schrire, S., Jan. 2006. Knowledge building in asynchronous discussion groups: Going beyond quantitative analysis. Computers & Education 46 (1), 49–70.
URL http://www.sciencedirect.com/science/article/pii/S0360131505000539

Senko, C., Hulleman, C. S., Harackiewicz, J. M., 2011. Achievement goal theory at the crossroads: Old controversies, current challenges, and new directions. Educational Psychologist 46 (1), 26–47.

Shea, P., Bidjerano, T., 2010. Learning presence: Towards a theory of self-efficacy, self-regulation, and the development of a communities of inquiry in online and blended learning environments. Computers & Education 55 (4), 1721–1731.

Shen, D., Cho, M.-H., Tsai, C.-L., Marra, R., 2013. Unpacking online learning experiences: Online learning self-efficacy and learning satisfaction. The Internet and Higher Education 19, 10–17.

Siemens, G., Gasevic, D., 2012. Guest editorial - learning and knowledge analytics. Educational Technology & Society 15 (3).

Tabachnick, B. G., Fidell, L. S., 2007. Using multivariate statistics. Pearson/Allyn & Bacon, Boston.

Trigwell, K., Prosser, M., 1991. Improving the quality of student learning: the influence of learning context and student approaches to learning on learning outcomes. Higher Education 22 (3), 251–266.

Valle, R. d., Duffy, T. M., 2009. Online learning: Learner characteristics and their approaches to managing learning. Instructional Science 37 (2), 129–149.

Winne, P. H., 1982. Minimizing the black box problem to enhance the validity of theories about instructional effects. Instructional Science 11 (1), 13–28.

Winne, P. H., 2006. How software technologies can improve research on learning and bolster school reform. Educational Psychologist 41 (1), 5–17.

Winne, P. H., Hadwin, A. F., 1998. Studying as self-regulated learning. In: Hacker, D. J., Dunlosky, J., Graesser, A. C. (Eds.), Metacognition in educational theory and practice. The educational psychology series. Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US, pp. 277–304.

Winne, P. H., Jamieson-Noel, D., 2002. Exploring students' calibration of self reports about study tactics and achievement. Contemporary Educational Psychology 27 (4), 551–572.

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 (2), 323–343.

Yen, J.-C., Lee, C.-Y., 2011. Exploring problem solving patterns and their impact on learning achievement in a blended learning environment. Computers & Education 56 (1), 138–145.

Zhou, M., Winne, P. H., 2012. Modeling academic achievement by self-reported versus traced goal orientation. Learning and Instruction 22 (6), 413–419.