QU, MIN

EXPLORING STUDENTS’ NAVIGATION PROFILE IN A COMPUTER-SUPPORTED COLLABORATIVE LEARNING CONTEXT

Master’s Thesis in Education

FACULTY OF EDUCATION
Master's Degree Programme in Learning, Education and Technology

2018
**Faculty of Education**

**Thesis abstract**

<table>
<thead>
<tr>
<th>Department of Educational Sciences and Teacher Education</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>Master's Degree Programme in Learning, Education and Technology</td>
<td>Qu Min</td>
</tr>
</tbody>
</table>

**Title**

Exploring Students’ Navigation Profile in a Computer-Supported Collaborative Learning Context

<table>
<thead>
<tr>
<th>Major subject</th>
<th>Type of thesis</th>
<th>Year</th>
<th>Number of pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning, Education and Technology</td>
<td>Master’s Thesis</td>
<td>2018</td>
<td>47</td>
</tr>
</tbody>
</table>

**Abstract**

As technology advances rapidly, computer-supported collaborative learning (CSCL) approaches are more and more implemented in educational contexts. However, the assessment of learning processes in CSCL is still a challenge for both teachers and students. Trace data (e.g., from log files), which is objective and can be collected in an unobtrusive way, provides great opportunity to investigate and assess learning process.

The present study explored high school students’ (N = 12) navigation behavior on a web-based learning environment during an advanced physics course. The course was implemented in a collaborative learning context and was loosely scripted. Students were instructed to work collaboratively on certain tasks during each lesson session. The study investigated students’ navigation profiles at three different levels, (i.e., class level, group level, individual level), and the relationship between students’ navigation profiles and their final learning outcomes.

The study used a quantitative research methodology. The log data of students’ navigation behavior was automatically recorded by the Open edX learning environment during the whole course. Log file was preprocessed (filtered and features extracted) before conducting descriptive and correlation analyses.

The findings of this study suggest that overall students were following the collaborative script during the whole course and some navigation behavior (i.e., navigated to course plan chapter) manifests students’ presence of planning and monitoring behavior. It was also found that each collaborative group consisted of different combination of individual navigation profiles and there was a significant correlation between students’ total navigation frequency and their final exam grade. The implication for students’ navigation profile as an assessment tool in CSCL are discussed.

The small sample size imposes a limitation of the generalizability of the results. In future research, it is suggested to investigate students' navigation behavior from multiple dimensions (e.g., sequential pattern, linearity of navigation) rather than a single factor (navigation frequency). Some other research possibilities are also proposed.

**Keywords**
collaborative learning, computer supported collaborative learning, navigation, scripting, web-based learning environment
# Contents

1 Introduction........................................................................................................................................... 1

2 Theoretical Framework...................................................................................................................... 4

  2.1 Collaborative learning ............................................................................................................... 4

  2.2 Assessment of Computer-Supported Collaborative Learning (CSCL)....................................... 5

  2.3 Scripting CSCL.......................................................................................................................... 8

  2.4 Navigation behavior in web-based learning environments ......................................................... 9

3 Aim and Research Questions........................................................................................................... 13

4 Methodology........................................................................................................................................ 14

  4.1 Participants and Context............................................................................................................. 14

  4.2 Implementation of an advanced physics course ......................................................................... 14

  4.3 Open edX learning environment................................................................................................. 15

  4.4 Data Collection.......................................................................................................................... 16

  4.5 Data Analysis............................................................................................................................ 16

    4.5.1 Log file preprocessing (including feature extraction) ......................................................... 16

    4.5.2 Descriptive and correlation analysis...................................................................................... 18

5 Results.................................................................................................................................................. 19

  5.1 Students’ navigation profiles at class level.................................................................................. 19

  5.2 Students’ navigation profiles at group level ............................................................................. 21

  5.3 Students’ navigation profiles at individual level......................................................................... 24
5.4 Relationship between students’ navigation profiles and their learning outcomes

6 Discussion ................................................................. 29

6.1 Limitations .................................................................. 33

6.2 Future work ................................................................ 33

7 Conclusion .................................................................... 35

References ...................................................................... 37
1 Introduction

Nowadays, many types of organizations like schools, companies, and non-governmental organizations are starting to make a shift from teacher-centered individual learning to student-centered collaborative learning (Wright, 2011). Theoretically, there are benefits of collaborative learning from both educational and psychological points of view. Small group settings provide learners with the opportunity to elaborate and share their understanding with peers and to be exposed to different constructs within a conceptual framework, which is crucial for learners’ knowledge retaining and comprehension (Slavin, 1990). From the idea of social constructivism, learners can only internalize knowledge and establish a conceptual framework through a social discourse (Munz, 2017). Also, the motivational benefits of learning together are well researched and confirmed (Blumenfeld, Kempler, & Krajcik, 2002; Järvelä & Järvenoja, 2011).

With the rapid development of technology, computers and software can bring learners together and provide more opportunities for knowledge co-construction and social interaction (Dourish, 2006). Computer-supported collaborative learning (CSCL) is a form of collaborative learning in which students use computers or other personal digital devices as a support to learn together (Dourish, 2006). Unlike traditional e-learning approaches, CSCL emphasizes the collaboration dimension, which requires both students’ cognitive presence and social interaction. Therefore, CSCL can either take the form of distant learning which involve students from different physical places or be implemented in a face-to-face setting (e.g., a classroom) where students learn together by using computers and tablets (Dourish, 2006).
Although CSCL is a promising approach in educational contexts, the assessment of CSCL is a challenging task for both teachers and students, and most of the assessment employed is summative, which is “isolated from the learning process” and “takes place only at the end of a course” (Strijbos, 2011). In contrast, formative assessment is integrated into the learning process, which leads to a comprehensive learning profile instead of a single score based on an exam or a final product (Harlen & James, 1996; Shute, 2006). It is believed that learners would be stimulated to adapt their learning accordingly if their individual and group learning profiles are evaluated and presented to them in real time during the learning process (Gress, Fior, Hadwin, & Winne, 2010).

Trace data (e.g. log files) creates detailed time-stamped records of learners’ behavior on a learning platform for real-time analysis. Unlike self-report methods which would interrupt learner’s learning process or think-aloud protocol which would add extra cognitive workload, trace data is objective data which does not depend on the learners’ subjective view and can be collected in an unobtrusive manner during the real time their learning activities take place (Winne, Gupta, & Nesbit, 1994). The use of trace data to investigate and assess learning processes has received more and more attention. However, the majority of the studies either focus on individual level (Bagarinao, 2015; Graf, Liu, & Kinshuk, 2010; Hadwin, Nesbit, Jamieson-Noel, Code, & Winne, 2007; Jeske, Backhaus, & Stamov Roßnagel, 2014; Somyürek, 2008), or develop assessment tools to assess CSCL based on trace data of text-based messages (Garrison, Anderson, & Archer, 1999; Henri, 1992; Pozzi, Manca, Persico, & Sarti, 2007). There are few studies utilizing trace data of learners’ navigation events to assess their learning processes during a collaborative course.
Therefore, the aims of this paper are twofold; first, to explore how students navigate on a web-based learning environment during an advanced physics course; second, to investigate the relationship between students’ navigation profiles and their final learning outcomes. The possibilities of applying students’ navigation profiles as a formative assessment tool in CSCL will be discussed.
2 Theoretical Framework

2.1 Collaborative learning

Collaborative learning is defined as two or more people working together to reach a shared learning goal through joint activity and knowledge co-construction (Dillenbourg, 1999; Roschelle & Teasley, 1995). To equip the next generation with better collaborative skills, various possibilities for applying collaborative learning methods in educational settings have been explored by researchers, educators, and educational technologists across the world in recently years (Dascalu et al., 2015; Järvelä et al., 2014; Laver, George, Thomas, Deutsch, & Crotty, 2012; Zheng, Niiya, & Warschauer, 2015). However, how collaborative learning approaches can be best implemented is not fully understood and still need to be researched (Laurillard, 2012).

Some researchers have identified different benefits of collaborative learning on social emotional aspects (Jones & Issroff, 2005; Rau & Heyl, 1990). Shared responsibility and humor have been found to be factors that reduce students’ anxiety during a problem solving process (Bol et al., 2012). Emotional support from peer members provide continuous motivation for learners to participate in learning tasks (So & Brush, 2008). Comparing with individual learning environments, there are cognitive benefits found in collaborative learning environments as well. As a task becomes more complex, learners in a group perform more efficiently than individual learners, as information can be distributed across all the group members which decreases each group member’s cognitive load (Kirschner, Paas, & Kirschner, 2009; Paas, van Gog, & Sweller, 2010; Van Merriënboer & Sweller, 2005). It has been shown also that students in collaborative groups significantly outperform students who study individually on a critical thinking test (Bol et al., 2012).
However, there are situations where collaboration may fail and group learning is impeded (Dillenbourg, 1999; Nokes-Malach, Richey, & Gadgil, 2015). In social respect, some learner in a group may expect others to exert higher efforts to finish the task, which is defined as social loafing behavior (Karau & Williams, 1993). The fear of negative feedback from other group members may hinder individuals from actively sharing their ideas and solutions (Mullen, 1987). From a cognitive perspective, students may lose their own trace of thoughts when interrupted by other group members, or forget ideas due to the distraction of listening to others and waiting for their turn (Basden, Basden, Bryner, & Thomas, 1997; Diehl & Stroebe, 1987; Finlay, Hitch, & Meudell, 2000).

2.2 Assessment of Computer-Supported Collaborative Learning (CSCL)

Standardized tests for individuals can often lead to rote learning, while group problem solving can facilitate deeper modes of learning (Entwistle, 2000). Many schools, universities and multinational organizations have been making a shift from traditional educator-driven methods of education to more collaborative and learner-driven forms of learning (Frijters, Haisken-DeNew, & Shields, 2004; London & Hall, 2011). In this transition, one of the main challenge is to develop appropriate assessment tools to uncover students’ learning processes and find different indicators of successful computer-supported collaborative learning (Baker & Mayer, 1999).

Researchers have proposed several methods to conduct analysis of CSCL processes, which utilize both quantitative and qualitative data (Daradoumis, Martínez-Monés, & Xhafa, 2004; Henri, 1992; Lipponen, Rahikainen, Lallimo, & Hakkarainen, 2003). The analysis of students’
collaborative behavior in a quantitative way is easy to implement and automate. However, it may only provide surface-level information about students’ learning dynamics. On the other hand, qualitative analysis which is often much more time-consuming, allow researchers to investigate specific cases and sessions in detail, although the results are only valid within a certain context (Pozzi et al., 2007).

A framework for assessing CSCL proposed by (Pozzi et al., 2007) assesses collaborative learning process from five perspectives, which includes the participative, interactive, social, cognitive and meta-cognitive, and teaching dimensions. The framework is mainly based on Henri's (1992) model for analyzing web-based online discussions and Garrison, Anderson, Archer's (1999) framework of the ‘Communities of Inquiry’”. To apply the theoretical framework into practice, appropriate indicators for evaluating each dimension during the learning process are proposed.

In the participative dimension, participation refers to the extent to which the students engage with the task and how well they maintain interaction with others (Care, Scoular, & Griffin, 2016). The participation level in collaborative learning task is regarded as a good predictor of students’ learning outcomes (Cohen, 1994) and provides a general but important description of their engagement in the learning process (Pozzi et al., 2007). Pozzi et al. (2007) take both learners’ visible activities and their absence or continuity into consideration. There are three categories of indicators to assess the participative dimension in the framework, namely, 1) indicators of active participation (the number of messages sent, the number of sessions attended, the number of documents uploaded, etc.), 2) indicators of passive participation (the number of documents
downloaded), and 3) indicators of continuity (the distribution of participation along time) (Pozzi et al., 2007).

In CSCL learning environments, interaction refers to the process in which students engage in building a relationship among them and construct shared knowledge (Pozzi et al., 2007), which should be evaluated by how much they scaffold each other’s cognitive processes rather than how frequently they interact with each other (Dillenbourg, 1999). Therefore, the most essential indicators of the interactive dimension come from content analysis of messages and documents exchange dynamics among students during the learning process, e.g., the number of documents downloaded before posting, the number of answers to other students’ messages, and qualitative analysis of how a student takes other group members’ idea into consideration (Pozzi et al., 2007).

Social presence is identified as whatever ‘is not related to formal content or subject matter’ (Henri, 1992). Garrison et al. (1999) go beyond Henri’s definition to define social presence as ‘the ability of participants in a community of inquiry to project themselves socially and emotionally, as “real” people with their full personality, through the medium of communication being used’. The cues of social presence are often demonstrated through thematic units (Anderson, Rourke, Garrison, & Archer, 2001) which are characterized by affection (e.g., expression of emotions, expression of intimacy) and cohesiveness (e.g., vocatives, salutations) (Pozzi et al., 2007).

Cognitive presence refers to the degree to which students can develop and affirm shared understanding through supported reflection and communication in a group (Garrison, Anderson, & Archer, 2001). The indicators are grouped into four phases following students’ cognitive
process, namely, revelation (e.g., recognizing a problem), exploration (e.g., expressing agreement/disagreement), integration (e.g., connecting ideas), resolution (e.g., testing solutions) (Pozzi et al., 2007). Meta-cognitive dimension is considered as a crucial element of the cognitive process which is aiming to monitor, evaluate and assess students’ own learning (Garrison, 2003). However, it is extremely difficult to find indicators to reveal learners’ strategic thinking due to the nature of meta-cognitive skills and therefore further research is needed to capture related traces (Pozzi et al., 2007).

Teaching presence refers to ‘the design, facilitation, and direction of cognitive and social processes for the purpose of realizing personally meaningful and educationally worthwhile learning outcomes’ (Anderson et al., 2001). In general, the indicators for teaching presence can be categorized into direct instruction, facilitating discourse, and organizational matters (Pozzi et al., 2007).

2.3 Scripting CSCL

Collaboration scripts are considered as the essential instructional design components in CSCL, which lead students to perform group work regarding task assignment, workflow, final products, etc. This arrangement can be made through teachers’ instructions or embedded in the learning platform (Dillenbourg & Jermann, 2007). According to Kollar, Fischer, and Hesse (2006), there are at least five elements involved in scripts for collaboration, namely, aims of learning, type of learning activities, sequence of learning activities, role assignment, and type of representation. Different scripts can be categorized into macro scripts and micro scripts (Dillenbourg & Hong, 2008). Macro scripts deal with the formation of the learning group and the sequence of the
learning activities whereas micro scripts specify roles and activities within each group to facilitate collaborative learning (Weinberger, 2011).

Several studies have shown that software-embedded CSCL scripts have the potential to improve group functioning and individual learning outcome (Rummel, Spada, & Hauser, 2009; Stegmann, Weinberger, & Fischer, 2007; Weinberger, Ertl, Fischer, & Mandl, 2005). However, scripts may also cause unwanted side effects. Epistemic scripts could reduce learners’ ability to build on each other’s reasoning (i.e. transactivity) during group discussion and finally harm individual learning gains (Weinberger et al., 2005). Over-scripting can demotivate learners and impede their self-regulated thinking (Kearney, 2004; Rummel et al., 2009).

In order to implement scripts as an effective instructional approach for CSCL, scripts should build on various underlying principles, namely, regulate learning activities, provide complementary procedural knowledge, provide process-oriented instruction, alleviate coordination, and foster awareness (Weinberger, 2011).

2.4 Navigation behavior in web-based learning environments

In order to better assess students’ learning processes in web-based learning environments, it is important and valuable to scientifically identify relevant information about learners’ activities in such environments, which is also called traces (Bousbia, Rebaï, Labat, & Balla, 2010). Analyzing those traces, namely a collection of learners’ actions, generates high-level knowledge about those actions, which researchers refer to as learning indicators (Bousbia et al., 2010). For example, researchers examine trace data to identify self-regulated learning activities and reveal students’ different self-regulation strategies (Hadwin et al., 2007).
Navigational behavior in web-based learning environments can be defined as students’ moving actions from page to page driven by various factors (motivation, emotion, cognition) to deliberately reach their learning goals (Bagarinao, 2015). One of the major advantage of web-based learning environments is that it allows learners to access course content in a flexible and non-linear manner (Somyürek, 2008). It is sensible that different students have multiple navigational patterns during their learning process (Carbó, Mor, & Minguillón, 2005).

The analysis of students’ navigation behavior is the foundation to better monitor students’ online learning process, which provides guidance and evidence to develop better instructional design, organize course content better, and scaffold students along different learning phases (Bagarinao, 2015; Carbó et al., 2005; Neuhauser, 2002).

System log data provides a great opportunity to depict students’ navigation behavior in web-based learning environments due to its complete records of students’ interaction with platform. Behavioral evidence of students’ self-regulation strategies can be found by capturing their navigation pattern, e.g., time in learning module, forward jumps, and backward jumps (Jeske et al., 2014). This type of data is also utilized by researchers to study whether students’ navigation profiles link to their academic achievement. Bagarinao (2015) identifies the most frequently consulted page(s) and correlates navigation frequency with students’ learning outcomes.

Some researchers treat navigation pattern as an output variable and examine how a student’s characteristics affect navigation behavior (Alomyan, 2004). Somyürek (2008) uses the values of students’ stratum, compactness and revisit percentage to differentiate their navigation pattern
and examines how individual difference, namely, gender, prior knowledge affect students’ navigation pattern. Bousbia et al. (2010) draw from studies about web browsing semantics and categorize learners’ navigational behavior into four navigation classes (over-viewing, studying, deepening and flitting) by calculating the navigation type indicator and automatically identify learners’ behavior during each training session, which benefit both course designers (e.g. adapt course content) and learners (e.g. cultivate learners’ metacognition).

Data mining techniques such as clustering and sequential pattern mining are applied based on students’ characteristics of navigation behavior to get an overview of different clusters of students and their common learning pattern and better assist instructors in their pedagogical design (Hung & Zhang, 2008; Lin, Hou, & Tsai, 2016; Poon, Kong, Yau, Wong, & Ling, 2017). Features of students’ navigation behavior in web-learning environments are also used to build predictive model to predict student performance by researchers and the frequency of visiting course materials is identified as the most dominant variable in their performance predictive model (Hung & Zhang, 2008).

In this study, we use navigation log file to examine how students navigated on a web-based learning environment during a face-to-face collaborative learning course. Open edX learning environment was chosen to implement the course, as it is open source allowing for customization of the environment. The study is integrated into the ongoing SLAM project funded by the Academy of Finland and led by the Learning, Education, and Technology (LET) research unit of the University of Oulu. The SLAM project harnesses advanced technologies to enhance strategic regulation of individual and collaborative learning. With advanced
technologies, the project utilizes new ways to use intelligent tools to analyze and visualize what is learnt and how it is learnt.
3 Aim and Research Questions

The aim of the study is to explore students’ navigation behavior on a web-based learning environment during an advanced physics course and to investigate if there is a connection between individual students’ navigation profile and their final learning outcome.

In accordance with the aim of the study, the research questions are as follows:

1. How do students navigate on the Open edX learning environment during an advanced physics course at the class level?
2. How do students navigate on the Open edX learning environment during an advanced physics course at group level?
3. What types of individual navigation profiles can be identified in a collaborative group?
4. What is the relationship between student’s navigation profile and their final learning outcome?
4 Methodology

4.1 Participants and Context

The participants of the study were 16-18 year old high school students (N=12) taking an elective, advanced physics course. There were three females and nine males among the participants. The students were assigned to four heterogeneous groups of three students, based on their previous grades in physics, and their score in the Motivated Strategies for Learning Questionnaire, which they answered before the study. The study benefited from the ecologically valid learning setting of the school.

4.2 Implementation of an advanced physics course

Students enrolled in this advanced physics course at their will, since it is a regular, elective course offered by the school. All the lessons were carried out face-to-face. There were 20 sessions in total including 19 lessons and one exam session. During each session, students’ learning activities were loosely scripted focusing on one certain topic within a chapter: they were instructed to work collaboratively to answer questions of group assignment on the Open edX learning environment. A general collaborative script of learning activities are as follows: 1) group members read questions of group assignment on certain topic, 2) discuss and answer two preparing questions before starting the actual group work, 3) write up answers to actual questions of group assignment together, 4) evaluate how their group work went on. However, there is no role distribution and rotation involved, since the script for this course is loose.
4.3 Open edX learning environment

The advanced physics course was managed via Open edX learning environment. The online course had several different chapters as we can see on the left side of Figure 1. Besides the ‘Course plan’ chapter, ‘Presentation of the learning environment’ chapter, and ‘Introduction’ chapter, chapter one to six mainly consist of learning content and task activities in this course. Under each chapter, there were several sections and there were several pages within each section, which students could navigate to and consult. For example, in Figure 1, there were three sections under chapter three and there were three pages within the second section.

![Course layout on Open edX learning environment](image)

Figure 1. Course layout on Open edX learning environment
4.4 Data Collection

Students’ navigation behavior on the Open edX learning environment during the whole course were all automatically recorded by the system in a log file. The exam grades of student were also collected by the Open edX learning system (students answered exam questions online in this platform). All the data were collected in an unobtrusive way.

4.5 Data Analysis

The data analysis process was mainly divided into two phases: log file preprocessing and descriptive analysis. The following two sections will explain them in detail.

4.5.1 Log file preprocessing and feature extraction

Log files recorded all student events including course navigation events on the Open edX learning environment during the advanced physics course. There were 14,590 logs in total; however, many of them were non-navigational event logs or other logs irrelevant to the purpose of this study. Therefore, a filtering process was performed using Microsoft Excel to get a reduced, clean log file, which only contained students’ navigational events. The cleaned data table for this research has eight variables/field and 1623 events and Table 1 is a description of those variables.

Table 1. A Description of Variables in the Data Table

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>EventID</td>
<td>Student event ID</td>
</tr>
<tr>
<td>Date</td>
<td>Event date</td>
</tr>
<tr>
<td>Time</td>
<td>Event time</td>
</tr>
</tbody>
</table>
The use of UserID instead of students’ actual names allows for students’ anonymization for the study, tackling the ethical issue of privacy. In order to reduce information redundancy and map session time information more conveniently, original ‘time’ field/variable which contained both time and date information was recoded into two separate variables ‘Date’ and ‘Time’ in the data table. New variables called ‘SessionID’, ‘GroupID’, ‘NavigationTo’, and ‘Chapter’ were added into the data table based on session time information, group division information and original variables ‘event_type’ and ‘event’. There are 20 sessions in total in this advanced physics course and 15 sessions were chosen for this research. Session 1, 2 were excluded because log events in these two sessions were mostly staff enrolling student. Two home sessions (session 4, 5) and the exam session (session 20) were also excluded. The workflow of preprocessing log file is briefly demonstrated in Figure 2.

<table>
<thead>
<tr>
<th>SessionID</th>
<th>Course session ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>UserID</td>
<td>Students’ user ID on the Open edX environment</td>
</tr>
<tr>
<td>GroupID</td>
<td>Which group a student belongs to (from 1 to 4)</td>
</tr>
<tr>
<td>NavigationTo</td>
<td>Which page a student navigated to/consulted</td>
</tr>
<tr>
<td>Chapter</td>
<td>Which chapter/section a student navigated to/consulted</td>
</tr>
</tbody>
</table>

Figure 2. The Workflow of Preprocessing the Log file using Excel
4.5.2 Descriptive and correlation analysis

The cleaned data table was analyzed with Excel, mainly using the pivot table function which allows to explore and summarize data from different angles. Several plots were also generated by Excel. Variables of interest (e.g., total navigation frequency) were aggregated and extracted, and the teachers provided exam grades. A correlation analysis was conducted on certain variables by using SPSS software.
5 Results

The results are generated from both descriptive statistical analysis and relational statistical analysis. The results from descriptive analysis depict students’ navigation profiles at class level, group level, and individual level. The results from correlation analysis shows a relationship between students’ navigation behavior and their final learning outcome.

5.1 Students’ navigation profiles at class level

Figure 3 is a bar chart that shows the total navigation frequency of the whole class along 15 sessions. The X axis represents each lesson during the whole course (session 1, 2, 4, 5, 20 were excluded for the reasons stated in “Log file preprocessing” section). The Y axis represents students’ navigation frequency which indicates the number of times they navigated from one page to another on Open edX learning environment.

The overall trend of students’ navigation frequency was decreasing, with a sudden drop in session 8 and a steep rise in session 16. The navigation frequency of the whole class (12 students) reached 151 or higher in 4 out of 15 sessions. Figure 3 also demonstrates the exact number of overall navigation events in each session. The navigation frequency of the first two sessions were relatively high which were 162 and 189 respectively, while the number of navigation events in other sessions were mostly below 100. Besides the first two sessions, session 9 and session 16 also had very high frequency both of which exceeded 150. Students’ navigation frequency in the session 19, which was the last session before exam was much lower than other sessions.
Figure 3. Bar chart of navigation frequency of the whole class along sessions

Figure 4 shows the exact chapter students’ navigation events took place in each session. As we can see from the plot, in general, the students’ navigation activity mainly focused on a same chapter in each session. From the beginning to the end of the course, namely from session 3 to session 19, the students left their footprints in every chapter from chapter 1 to chapter 6 and two or three session time were spent on each chapter. The last two sessions, namely 18 and 19 were somewhat different from other sessions, as the students were not only focusing on chapter 6 and navigation events in chapter 4 and chapter 5 had considerable proportion of presence in these two sessions. The navigation frequency of Course Plan chapter was also plotted as a separate line in Figure 4. It is obvious to see that students had consistent navigation to Course Plan chapter from early phase (session 6) to the end of the course (session 19). Also, interestingly, students’ navigation events to Course Plan chapter were more clustered in the middle phase of the course, while students consulted it less frequently in the early and final stage of the course.
In summary, at class level, students’ navigation frequency started very high in the first two sessions and the overall trend was decreasing with certain spikes in a few sessions. During each session, the majority of students’ navigation focused on one particular chapter, which confirmed that this course was loosely scripted in each session instead of a self-paced course. Although students navigated within a main targeted chapter in each session, interestingly, a small portion of navigation events indicate that students navigated to the previous chapters (i.e. previous learning content) as well during their learning process in most of the sessions. In addition, students were consistently navigating to the course plan chapter during the whole course.

5.2 Students’ navigation profiles at group level

Figure 5 shows how different groups navigated during the course and there is no clear pattern found in this plot. However, there are certain group features that can be identified, for example,
group 2 had very high active level of navigation in the beginning of the course and dropped dramatically along first few sessions and remain relative low active level until the end of the course. The distribution of group 3’s total navigation frequency had notable ups and downs, where at their peaks, their navigation frequency outnumbered all other groups’. It seems that group 1 and group 4 were more steady and stable during their learning process. From Figure 6, which is a boxplot of different groups’ navigation frequency, we can see group 3 was more active overall than other three groups during this course.

![Figure 5. Comparison of group navigation frequency along sessions](image-url)
The different pages in Open edX learning environment of this advanced physics course were categorized as Chapter 1, Chapter 2, Chapter 3, Chapter 4, Chapter 5, Chapter 6, Course Plan, Chapter and Other pages. Figure 7 compares groups’ navigation frequency across different page categories. Although the exact navigation frequency of different categories varies from group to group, interestingly, all the groups have similar pattern across different categories in term of navigation frequency.
In summary, there was no clear pattern identified in terms of group navigation frequency along sessions. However, there was a clear pattern found in terms of group navigation frequency across different page categories. In general, group 3 was more active than other groups, despite the largest variation among all the groups.

5.3 Students’ navigation profiles at individual level

Students’ individual navigation profiles were also identified in term of their navigation frequency. We investigated each student’s profiles of navigation frequency in each chapter. Firstly, the average navigation frequency (mean value) in each chapter in terms of all the students was calculated. An average line with a range from 0.25 standard deviation (SD) above the mean value to 0.25 SD below the mean value has been proposed, as shown in Figure 8. Then each student’s z-score of navigation frequency in each chapter was calculated. As Figure 8
shows, if a student’s z-score was above the average line, the corresponding square of that chapter was colored as green; if it was below the average line, the corresponding square of that chapter was colored as red; if it was within the range of average line, the corresponding square of that chapter was colored as grey. Thus, Figure 8 demonstrates students’ profiles of navigation frequency in terms of each chapter within their collaborative groups.

Figure 8. Students’ Profiles of Navigation Frequency in Collaborative Groups

Based on the results above, the following classification has been proposed to determine students’ overall active level of navigation during this course:

- **High** = above the average line in 4 or more chapters
- **Intermediate** = no more than 3 chapters above or below the average line
- **Low** = below the average line in 4 or more chapters

The aim of this classification is to not only take students’ total navigation frequency, but also their continuity into consideration. Table 2 shows different types of students’ navigation profile in their collaborative group and it is obvious that different types of groups emerges in terms of
group members’ individual navigation profile. For example, in group 1, each student’s active level varied from high to low, in contrast in group 2, student 15 was very active while the other two group members’ active level was low.

Table 2. Classification of Individual Navigation Profiles in Collaborative Groups

<table>
<thead>
<tr>
<th>Group Number</th>
<th>Student ID</th>
<th>Active Level of Navigation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>Student 12</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Student 14</td>
<td>Intermediate</td>
</tr>
<tr>
<td></td>
<td>Student 18</td>
<td>Low</td>
</tr>
<tr>
<td>Group 2</td>
<td>Student 15</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Student 16</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Student 17</td>
<td>Low</td>
</tr>
<tr>
<td>Group 3</td>
<td>Student 20</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Student 26</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Student 27</td>
<td>Low</td>
</tr>
<tr>
<td>Group 4</td>
<td>Student 21</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>Student 23</td>
<td>Intermediate</td>
</tr>
<tr>
<td></td>
<td>Student 24</td>
<td>Intermediate</td>
</tr>
</tbody>
</table>

In summary, different combination of individual navigation profiles formed each collaborative group and there is no such group found in which all the students’ active level of navigation was high or low.

5.4 Relationship between students’ navigation profiles and their learning outcomes

Table 3 shows descriptive statistics of three variables that we aimed to find correlations between them. The exam grade is sum of students’ group task and individual task based on 4-10 scale.

Table 3. Descriptive statistics of variables of interest
<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Navigation Frequency</td>
<td>135.25</td>
<td>46.20</td>
<td>12</td>
</tr>
<tr>
<td>Navigation Frequency of Course Plan Chapter</td>
<td>11.50</td>
<td>8.68</td>
<td>12</td>
</tr>
<tr>
<td>Exam Grade (include group task and individual task)</td>
<td>8.56</td>
<td>0.81</td>
<td>12</td>
</tr>
</tbody>
</table>

Based on the results shown in Table 4, there is significant positive correlation between students’ total navigation frequency and their navigation frequency of course plan chapter ($r = 0.822, p < .01$). There is also correlation found between variables about students’ navigation behavior during the course and their final learning outcome, namely students’ total navigation frequency and exam grade ($r = .801, p < .01$), as well as navigation frequency of course plan chapter and exam grade ($r = .700, p < .05$).

Table 4. Correlations of interested variables

<table>
<thead>
<tr>
<th></th>
<th>Total Navigation Frequency</th>
<th>Navigation Frequency of Course Plan Chapter</th>
<th>Exam Grade (include group task and individual task)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Navigation Frequency</td>
<td>1</td>
<td>.822**</td>
<td>.801**</td>
</tr>
<tr>
<td>Navigation Frequency of Course Plan Chapter</td>
<td>.822**</td>
<td>1</td>
<td>.700*</td>
</tr>
</tbody>
</table>

**p < .01, *p < .05

It is not surprising that there was strong positive correlation between students’ total navigation frequency and their navigation frequency of course plan chapter. Students who had higher total
navigation frequency during the course also tended to consult more times of course plan chapter. Therefore, we wanted to find out if the significant correlation between students’ navigation frequency of course plan chapter and their exam grade was due to the contribution of a third variable, which is total navigation frequency. Table 5 is the result of a partial correlation analysis in which the third variable total navigation frequency was controlled. There is only weak correlation between navigation frequencies of course plan chapter and exam grade and the correlation is not significant.

Table 5 Partial correlation of interested variables

<table>
<thead>
<tr>
<th>Control Variables</th>
<th>Exam Grade (include group task and individual task)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Navigation Frequency</td>
<td>Navigation Frequency of Course Plan Chapter</td>
</tr>
</tbody>
</table>

p= .716

In summary, there was a significant correlation found between students’ total navigation frequency and their final exam grade. Because of the significant correlation existed between total navigation frequency and navigation frequency of course plan chapter, we found the significant correlation between navigation frequency of course plan chapter was due to a third factor, which is students’ total navigation frequency.
6 Discussion

This study explored students’ navigation behavior in Open edX learning environment during an advanced physics course and investigated the possible relationship between students’ navigation profiles and their final learning outcomes.

Regarding students’ overall navigation behavior in this class, we found that navigation frequency of the whole class in first two sessions were rather high compared to other sessions. Since Open edX learning environment is new to most of the students, they probably had some extra navigation activities in order to get familiar with this web-based learning environment. The rather low navigation activity of session 19 is mainly due to that the workload of session 19 is low, as no scripted group work was carried out, according to the course plan.

The course was loosely scripted for each session that students were instructed to collaborate on learning tasks following certain sequence. Several studies have provided evidence that collaboration scripts have positive effect on group dynamics and individual learning outcome (Stegmann, Weinberger, & Fischer, 2007; Weinberger, Ertl, Fischer, & Mandl, 2005). The findings that during each session, students mostly navigated within one particular chapter confirmed that students were following the script defined by the teachers during the whole course. Moreover, every student successfully accomplished this course. Although students had a main targeted chapter in each session, they more or less consulted pages in previous chapter during most of the sessions, which suggests that students tried to activate or link to prior knowledge when learning new content. As regulated learning skill plays an important role in collaborative learning (Järvelä & Hadwin, 2013), we also found that students navigated to
course plan chapter consistently from very beginning of the course which was a sign of students’ planning and monitoring behavior. From the distribution of navigation frequency to course plan chapter, we can infer that students tended to monitor their process in the middle phase of the course.

Students can get great support from CSCL script which help to avoid losses during the process by assigning tasks within groups (Weinberger, Stegmann, & Fischer, 2010). Concerning groups’ navigation behavior during this course, there was a clear pattern found in terms of group navigation frequency across different chapters, which is a sign that all the groups were following a same script during the whole course. Although group 3 varied more considerably among different sessions in terms of navigation frequency, it was a more active group in general. It was probably because there were more highly active participants in this group.

Based on the results of cluster analysis, individual navigation profiles were classified as high, intermediate, low in terms of active level of navigation. The results of individual navigation profiles confirmed that group 3 was a more active group at group level. There were two members in that group whose active level of navigation was high. Each group consisted of a different combination of individual navigation profiles. For example, in group 2, student 15’s active level of navigation was high but the other two group members participated at a low level; in group 4, although there was a lack of highly active participants, there was no considerable difference between group members’ contribution. These findings are in line with previous studies. The results from some previous studies show there are some dominant members existing in a collaborative learning group (e.g., Savicki, Kelley, & Ammon, 2002), whereas other studies report more equal contribution of collaborative groups (e.g., Fjermestad, 2004). Both group and
individual achievement depend on high-level engagement and equal participation (Cohen, 1994; Webb, 1995). However, there is no such group found in which all group members' active level was high during this course. Some students probably engaged in social loafing which is the tendency to invest less effort when working collaboratively than when working by oneself (Karau & Williams, 1993) and certain group may even encountered “free rider” problem during their learning process (Salomon & Globerson, 1989). The loose script implemented in this course avoided the risk of over-scripting which could demotivate students and hinder their self-regulated behavior (Kearney, 2004; Rummel et al., 2009). However, without role assignment and rotation, it also lost the opportunity to facilitate students’ equal participation during this course (Kollar et al., 2006). The trade-off regarding scripting in CSCL is worthy of further investigation.

There is significant correlation found between students’ total navigation frequency and their final exam grade. This result meets our expectation since learner’s visible activities (e.g., navigation events) and their absence or continuity are indicators of students’ participation level in collaborative learning (Pozzi et al., 2007). And the participation level provides an important profile of their engagement and is considered as a good predictor of students’ final learning outcomes (Cohen, 1994; Pozzi et al., 2007). To some extent, the result is also in line with several previous studies. Bagarinao (2015) reported a slight but nonsignificant correlation between navigation frequency and students’ learning outcomes. However, the research was conducted in the context of an online course for individual learning. Also some researchers found that the frequency of visiting course materials is the most dominant variable in their learning performance predictive model (Hung & Zhang, 2008).
Teachers cannot perfectly predict students’ gains and behavior during their learning process, no matter how carefully a course is designed and implemented, and that’s the reason why assessment is the central part in effective teaching and learning (Wiliam, 2013). There are two tools teachers often use to evaluate students’ learning performance which are formative and summative assessment (Dixson & Worrell, 2016). Formative assessment is often referred to as “assessment for learning”, which should be used throughout the learning process rather than only at the end. Formative assessment focuses on different aspects of learning (e.g., cognition, motivation, social interaction), which “leads to a profile instead of a single score” and provides teachers opportunity to intervene during learning process (Stiggins, 1994; Strijbos, 2011). On the other hand, summative assessment is referred to as “assessment of learning”, which is not an integral part of a learning process and takes place only when a course is over. Summative assessment mostly focuses on the “cognitive aspects of learning” and are always graded as a single score (e.g., final exams, college entrance exams, term papers) (Dixson & Worrell, 2016; Harlen & James, 1996; Strijbos, 2011).

Several studies have explored how students’ navigation profiles constructed from log files can help to capture self-regulated learning activity and serve as a formative assessment tool which affords opportunities to perform intervention and adapt self-regulation (Hadwin et al., 2007; Jeske et al., 2014). However, there is lack of effective formative assessment tool in CSCL environment. Traditionally, peers are regarded as a source of formative assessment in CSCL, despite some issues on reliability, validity, and implementation (Cho, Schunn, & Wilson, 2006; Prins, Sluijsmans, Kirschner, & Strijbos, 2005; Topping, 2013). Therefore, we propose to utilize trace-based navigation profiles as a formative assessment tool to assess CSCL process for several reasons. Firstly, unlike self-report and peer assessment approach, navigation profiles
based on log data can be obtained in real-time and in an unobtrusive manner, without interrupting students’ learning process. Secondly, students’ navigation profiles (e.g., navigation frequency, navigation target, sequence of navigation, linearity of navigation) reflect their learning in different dimensions of CSCL process (e.g., participative, interactive, cognitive dimension) (Pozzi et al., 2007), which provides students with opportunities to adapt their learning accordingly (Gress et al., 2010). Lastly, like other trace data, navigation events are “less susceptible than other methodologies” for avoiding students’ possible bias recall or perception (Winne, 2015).

6.1 Limitations

There were certain limitations in this study. Due to the small sample size and the non-experimental nature of this study that lacked the manipulation of an independent variable, the significant correlation found might be biased and it was impossible to run inferential statistical test to find cause and effect relationship. In addition, the students’ navigation profile were constructed solely based on navigation frequency, which failed to provide a holistic picture of students’ learning situation. Because there were few studies investigating students’ navigation behavior in a collaborative learning context, the proposed classification of students’ navigation profile lacked theoretical foundation.

6.2 Future work

It is recommended to explore students’ navigation behavior on web-based learning environment from different perspectives (e.g., sequential patterns of pages navigated, linearity of navigation) other than navigation frequency in future research by conducting more advanced trace analysis
(e.g., sequential pattern mining, association rule mining). There are already some relevant studies however not in a collaborative learning context (Bousbia et al., 2010; Hadwin et al., 2007; Lin et al., 2016; Poon et al., 2017; Zaïane, 2001). To conduct experimental design research in which control groups implement loose script and treatment groups implement tight script including role assignment and rotation will bring more insights on how script approach can be best utilized in CSCL contexts. The investigation of how students’ collaborative skills or previous group work experience connect to their navigation behavior during a collaborative course would also be interesting.
7 Conclusion

The aim of this study was to explore students’ navigation behavior on Open edX environment during a collaborative course and to investigate individual students’ navigation profile and final learning outcome. Overall, the findings of this study concerns students’ navigation behavior at class level, students’ navigation behavior at group level, classification of individual students’ navigation profile, and the relationship between individual students’ navigation frequency and their final learning outcome.

The physics course involved in this study was loosely scripted for each session. There is potential evidence found that the whole class and each group were following the same script defined by the teachers during the course. As several studies show that collaboration scripts facilitate group learning and enhance individual learning outcome (Stegmann et al., 2007; Weinberger et al., 2005), students in this study all successfully accomplished the course requirements. There is also evidence suggesting that the students might have used self-regulated learning skills (e.g., planning, monitoring) during the whole course, which is considered as the essential skill in collaborative learning (Järvelä & Hadwin, 2013).

Although Cohen (1994) and Webb (1995) state that high-level engagement and equal participation are crucial for both group and individual’s success in collaborative learning, we failed to find such a case where all the group members were very active during this course. The findings of this study that collaborative groups were formed of different combinations of individual navigation profile confirm previous studies (e.g. Fjermestad, 2004; Savicki et al., 2002).
Students’ participation level in collaborative learning is often regarded as a reliable predictor of their learning achievement (Cohen, 1994; Pozzi et al., 2007), and navigation events are indicators of students’ participation level (Pozzi et al., 2007). The positive correlation between students’ total navigation frequency and their final exam grade we found in this study is in line with those theoretical claims. However, more studies concerning students’ navigation behavior in collaborative learning context is needed, as most of the previous studies focused on individual learning (e.g., Bagarinao, 2015; Hung & Zhang, 2008).

Log-based navigation data is unobtrusive, informative, and objective by nature (Greene & Azevedo, 2010; Winne, 2015). Therefore, we also propose to employ students’ navigation profiles on web-based learning environment as a formative assessment tool in CSCL context. However, more research is needed on students’ navigation behavior from different perspectives in order to provide more complete learning profiles, which would reflect multiple dimensions of CSCL processes (Pozzi et al., 2007).
References


https://doi.org/10.1037/0022-3514.53.3.497

https://doi.org/10.1.1.167.4896

https://doi.org/10.1007/s11412-007-9033-1


https://doi.org/10.1080/00405841.2016.1148989

https://doi.org/10.1145/1124772.1124855


and discourse in elementary students’ computer-supported collaborative learning.

*Learning and Instruction, 13*(5), 487–509.


https://doi.org/10.1002/hrm.20455


https://doi.org/10.1007/s13398-014-0173-72


https://doi.org/10.1207/S15389286AJDE1602_4


https://doi.org/10.4135/9781452218649.n22


https://doi.org/10.3102/01623737017002239


https://doi.org/10.1007/s11251-004-2322-4


https://doi.org/10.1080/03075079312331382498


https://doi.org/10.1080/1475939X.2014.948041