Using Differential Sequence Mining to Associate Patterns of Interactions in Concept Mapping Activity with Dimensions of Collaborative Process

by

Rishabh Chaudhry

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved September 2015 by the
Graduate Supervisory Committee:

Erin A. Walker, Chair
Roberto Maldonado-Martinez
Ihan Hsiao

ARIZONA STATE UNIVERSITY

December 2015

ABSTRACT

Computer supported collaborative learning (CSCL) has made great inroads in classroom teaching marked by the use of tools and technologies to support and enhance collaborative learning. Computer mediated learning environments produce large amounts of data, capturing student interactions, which can be used to analyze students' learning behaviors (Martinez-Maldonado et al., 2013a). The analysis of the process of collaboration is an active area of research in CSCL. Contributing towards this area, Meier et al. (2007) defined nine dimensions and gave a rating scheme to assess the quality of collaboration. This thesis aims to extract and examine frequent patterns of students' interactions that characterize strong and weak groups across the above dimensions. To achieve this, an exploratory data mining technique, differential sequence mining, was employed using data from a collaborative concept mapping activity where collaboration amongst students was facilitated by an interactive tabletop. The results associate frequent patterns of collaborative concept mapping process with some of the dimensions assessing the quality of collaboration. The analysis of associating these patterns with the dimensions of collaboration is theoretically grounded, considering aspects of collaborative learning, concept mapping, communication, group cognition and information processing. The results are preliminary but still demonstrate the potential of associating frequent patterns of interactions with strong and weak groups across specific dimensions of collaboration, which is relevant for students, teachers, and researchers to monitor the process of collaborative learning. The frequent patterns for strong groups reflected conformance to the process of conversation for dimensions related to "communication" aspect of collaboration. In terms of the concept mapping sub-processes

the frequent patterns for strong groups reflect the presentation phase of conversation with processes like talking, sharing individual maps while constructing the groups concept map followed by short utterances which represents the acceptance phase. For “joint information processing” aspect of collaboration, the frequent patterns for strong groups were marked by learners’ contributing more upon each other’s work. In terms of the concept mapping sub-processes the frequent patterns were marked by learners adding links to each other’s concepts or working with each other’s concepts, while revising the group concept map.

ACKNOWLEDGMENTS

Well, I didn't do it all by myself and I was fortunate enough to be supported and guided mostly in a constructive way by some generous and obviously smarter people than me. I write this section with utmost sincerity and regards towards all these people who were kind enough to keep my boat afloat.

First up, I would like to thank Dr. Erin A. Walker for giving me this opportunity to work at the HCI lab and for being an amazing advisor and mentor for the past two years. I was involved in a couple of other projects at the lab apart from my thesis and got to learn a lot from her all along the way. She has this awesome ability to let you drive your own work and learning with continuous support and motivation from her end. The kind of effort she put in to drive me, guide me in the proper direction and provide constant feedback was astounding and made me want to continually improve in my work. Words are too meagre and I am not decent enough a writer to express my gratitude towards her.

I would also like to thank Dr. Roberto Martinez-Maldonado for his invaluable guidance and for sharing his work and dataset, upon which, I later built my research work. I bugged him a lot with queries and he was always there to clarify my doubts and show me a way forward. I would also like to thank Dr. Ihan Hsiao for her constructive feedback and suggestions to improve my thesis.

Some of my friends and colleagues also deserve to get their names imprinted in this section. I would like to extend sincere thanks to Sree Aurovindh for guiding me along with my work; Victor Girrotto for our highly philosophical and intellectual discussions and for his feedback on my initial draft; Shang Wang for being an awesome lab-mate and for driving me around sometimes in his Mustang.

Coming to the clichés, it couldn't have been possible without my family and my friends Akanksha, Pummy, Taruna, Sahil, Maitreyee, Denny, Tejas, and Tanmay because they always doubted me, took my share of the stress and cheered me up whenever I was down.

TABLE OF CONTENTS

Page

LIST OF TABLES

1. Introduction

1.1 Motivation and Thesis Goals

The long term goal of this research is to inform the design of tools which would support teachers and researchers alike in analysing how students collaborate in classroom learning and provide appropriate interventions to improve collaboration and enhance learning.

The knowledge, facilitation and improvement of collaboration skills is a key factor for enhanced learning in computer supported learning environments. The probability of some specific cognitive mechanisms like knowledge elicitation, externalization, reduced cognitive load and internalization, which lead to learning, is more when students interact with each other while collaborating (Dillenbourg, 1999). This makes collaboration skills important for higher performance in learning, critical thinking, higher retention and interest in the students (Johnson & Johnson, 1986). The use of computer-mediated tools to support collaboration is increasingly becoming an important aspect of classroom learning. These tools provide encouragement to students to pursue higher individual learning and achieve better collaboration (Dillenbourg et al., 2011). But collaboration does not necessarily happen just because CSCL environment is conducive to it. Effective collaboration in CSCL needs development of certain skills such as providing explanations and seeking information, joint coordinated effort, negotiation among others which makes teachers a pivotal cog in enhancing students' performance by monitoring the collaborative process, understanding student strategies and providing appropriate feedback or interventions.

The role of teachers in classroom orchestration has always been of the utmost importance, but with the advent of CSCL it has evolved even further in the past 3 decades (Dillenbourg & Fischer, 2009). As much as CSCL is about supporting collaboration, group cognition and the idea that students co-construct meaning, it doesn't endorse "teacherless learning"(Dillenbourg & Fischer, 2009). This necessitates a shift in focus for teachers from providing support on content to understanding how students think and learn, and channelling their suggestions based on that understanding. They need to have better awareness about the flow of knowledge and collaboration dynamics during small group learning activities (Dillenbourg et al., 2011). Their role is crucial in advancing the group collaboration skills of the students and improving their group dynamics (Dillenbourg et al., 2011) and hence there is a rising need for them to be equipped with tools that facilitate this knowledge growth.

To address this, the large amount of data that can be captured from the interactions of students with the tools, technologies and their peers can be exploited to support the teachers and researchers in monitoring and gaining insights into the collaborative and learning process. In order to leverage this data and support teachers in facilitating classroom activities, it is important to have a deep understanding of what constitutes a good collaborative process. Meier et al. (2007) defined nine process dimensions across five aspects of collaborative processes which are rated quantitatively to measure the degree of collaboration. The dimensions were derived from data from a study combined with theoretical considerations on literature from collaborative learning, CSCL and group cognitive processes. The data was taken from a study involving psychology and medicine students collaborating over a desktop based videoconferencing system and involved a qualitative analysis of the video recordings which resulted in the rating scheme. This rating scheme was then applied to a new study for evaluation. The ratings were given by humans. This multidimensional model of collaboration given by Meier (2007) covered aspects of communication, joint information processing, coordination, interpersonal relationship and motivation.

The students' interactions with their peers and the computer mediated learning environment while engaged in a joint problem solving task constitute one of the most important aspect of collaborative learning. The group cognition (Stahl, 2006) and distributed cognition (Hutchins, 1996) theories which are reflective of collaborative learning, also give utmost importance to these interactions and talk about them as an essence of collaborative cognitive processes. Thus, the analysis of students' interactions is the most essential component for evaluation of the collaborative process. The rating scheme and the dimensions Meier (2007) defined to assess the quality of collaboration can be used in different CSCL settings for assessing collaboration using students' interactions. The rating scheme involves human evaluations which are arduous and time consuming. But deeper aspects of collaboration require human intervention to be evaluated as it involves subtleties which cannot be conveyed by only automated metrics of logged data or events. One approach going forward is to incorporate these human rated dimensions as an external reference model and then use automated interaction analysis techniques to characterize these dimensions of collaboration in a way that that reflects the already proven model of human analysis (Kahrimanis et al., 2011a). Identification of patterns of students' interactions is one of the most essential components of interaction analysis and these patterns are reflective of the collaborative learning. (Stahl, 2013). The analysis of patterns of interaction in small groups is important for understanding collaborative learning and we can use the "good" and "bad" collaboration models based on human evaluations to automatically extract frequent patterns of interactions.

As established, there is this need and opportunity for automatically associating patterns of students' interactions to higher-order collaborative process dimensions. There has been research which uses the students' interactions to analyze the collaborative process, taking into account high vs. low collaboration in groups based on all the aspects of collaboration taken as a cumulative unit. Analyzing the collaborative process across all the different aspects of collaboration would give deep insights about the strategies and interactions that characterize these different aspects of collaboration.

This thesis aims to examine the frequent patterns of students' interactions to analyse the collaborative learning process across the different dimensions of collaboration defined by Meier et al., (2007). I used an exploratory data mining technique that is known as sequence pattern mining and it was applied on the sequential data of students' interactions with the multi-touch tabletop while they were engaged in the collaborative concept mapping activity. I only focused on the first four dimensions under the “communication” and “joint information processing” aspects of collaboration, defined by Meier et al., (2007) and the results obtained were frequent patterns of students’ interactions across these four dimensions of collaboration(Sustaining Mutual Understanding, Dialogue Management, Reaching Consensus, Joint Information Processing).

The specific context of the study, which this thesis is based upon, is a face-to-face collaborative learning activity using concept maps in an enriched interactive tabletop in a small group. Martinez-Maldonado et al. (2013a) conducted this study and exploited the affordances of these tabletops to automatically capture digital traces of students’ interactions. The data was logged in the database in terms of all the physical actions, instances of verbal communication accompanied with the participants’ identifiers. Lemonier & Martinez-Maldonado et al. (2014), defined a coding scheme to encode this low level logged data into high level collaborative concept mapping processes. This resulted in sequence of concept mapping sub-processes for each group participating in the study.

There is an opportunity to apply exploratory data mining techniques to compare collaboration across the different dimensions in terms of the sub-processes and take the research beyond what has been done by the other noteworthy researchers. The idea is to take the analysis of the students’ interactions and learning traces that characterize high and low collaborative groups, or high and low achieving students ,one step further and determine which patterns of sequences of high level sub-processes of the group concept mapping activity characterize the specific dimensions of collaborative processes that differentiate the groups of students. This inspires and motivates the goal for my thesis which can be formulated as below:

1. Use sequence pattern mining to examine how frequent sequences of collaborative interactions differ between high-performing groups and low-performing groups across different process dimensions of measuring the collaboration quality.
2. Use sequence pattern mining to identify frequent sequences of collaborative interactions that characterize the important process dimensions of collaboration considering theoretical underpinnings for each dimension.

1.2 Contribution

The contribution of this work is the analysis of the collaborative process at a more granular level by associating the frequent patterns of interactions with different dimensions across which the quality of collaboration is assessed. The results can assist teachers and researchers in analysing the collaborative processes that ensue in high and low collaborative groups, enabling them to provide interventions or feedback in classrooms in real time. This thesis is only contributing in terms of presenting the information to the researchers or teachers for their perusal. The associations between the frequent patterns and dimensions of collaboration are presented considering the theoretical underpinnings for each dimension, which makes the contribution more exciting as the results are not just explained around the dimensions but analysed to examine if the patterns obtained are meaningful and meet the expectations considering the literature as well.

1.3 Organization of the Thesis

This thesis document is organized as follows: Chapter 2 talks about the relevant background research related to Computer Supported Collaborative Learning (CSCL), assessment of collaboration, data exploration and mining and collaborative concept mining which for the basis of the thesis. Chapter 3 describes the research methodology including the study and dataset upon which this work is based along with the specific data mining technique (DSM) that has been used. It also formally introduces the research questions being targeted in this thesis. Chapter 4 presents the results obtained on executing DSM on the sequential dataset prepared for this work with in-depth discussion of the results. This is followed by the conclusion and further work section.

2. Background and Related Work

2.1 Collaboration and Learning Theories

A number of theories contribute towards providing a framework for understanding aspects of CSCL and the way research has evolved in the past few decades. The constructivist theory of Group Cognition (Stahl, 2006) occupies an important place as a theoretical perspective in CSCL research. CSCL is all about supporting collaborative learning, co-construction of knowledge and group cognition (Stahl, 2013). For this thesis, I focused on the group cognition theory to define the theoretical framework. Martinez (2014b), also designed his approach based on this theory as he considered “the tabletop hardware as just a partial mediator of knowledge, and the artifacts that are built through this medium as true containers of student’s shared understanding”. Before moving forward, it’s also important to discuss collaboration and cooperation, which are key elements of CSCL.

A number of researchers have defined collaboration in different ways, which reflects different perspectives. Collaborative learning can be broadly defined as a “situation in which two people learn or attempt to learn together” or simply come together for joint problem solving (Dillenbourg, 1999). Roschelle & Teasley (1995) define collaboration as “a coordinated, synchronous activity that is the result of a continued attempt to construct and maintain a shared conception of a problem". Another term that is used synonymously with collaboration is co-operation. However researchers differentiate between these two forms of peer learning. In collaborative learning the work is done by the group as a whole and even the division of tasks/labor is horizontal which means it is across different aspects of a task. In contrast, in cooperative learning tasks are divided into independent sub-tasks so that individuals can work and construct their knowledge independently and then combine their results for the common task. The end product of collaboration is more synergistic and the group effort cannot be traced in terms of individual contributions. Researchers agree that it's difficult to arrive at a single definition of collaboration and that collaborative learning is not a single mechanism or method (Dillenbourg, 1999). Dillenbourg (1999) further explains how collaborative learning can be described in terms of a situation, which might facilitate certain interactions to occur that may trigger learning mechanisms. However, it cannot be ascertained whether or not these interactions would occur. In general, collaborative learning is associated with students giving explanations and receiving feedback, which helps students elaborate their knowledge and repair their mental models of understanding (van boxtel et al., 2000). This elaboration process is stimulated by social interactions and learning is enhanced when students have to indulge in the process of giving explanations, negotiating and reaching a shared understanding.

In light of this discussion about collaborative learning, it's imperative to discuss how different theoretical perspectives contribute towards grounding this concept. All the theories in CSCL research are based on the same underlying paradigm that considers individual cognition in a social context rather than in isolation. CSCL research is highly influenced by social interactions, individual learning and their interplay supported by tools and technologies (Hartman, 2015). As part of this discussion, it also becomes relevant to describe the role of social interactions within all these theories. Further, referring to Hartman’s et al., (2015) work, linking these roles to different forms of constructivism helps categorize the nature of learning within these theories of CSCL. A large body of research on collaborative learning and CSCL is rooted in socio-cognitive and socio-cultural theories by Piaget and Vygotsky respectively, based on which we will move to our discussion about the group cognition theory.

2.1.1 The Piagetian Perspective, Vygotskian Perspective and Constructivism

Piaget posits that the individual cognitive development happens when a child/learner experiences a cognitive conflict owing to interactions with new information or different cognitive frameworks and eventually cognitive equilibrium fosters development. According to social constructivists, this cognitive conflict is facilitated by social interactions. Moshman (1982) classifies this as endogenous constructivism, as the social interactions are enabling the individual cognitive developmental process that is more reflective of cooperative learning. While this socio-cognitive approach focuses on conflict being the mediating factor enabled by social interactions, Vygotsky’s approach showcases individual cognitive change as a causal effect of social interactions. The Vygotskian perspective also known as the socio- cultural perspective says that the cognitive development happens in the zone of proximal development which is the distance between the accomplishments of a student when he/she works alone with respect to what he/she can achieve working with an advanced expert peer. According to Vygotsky’s principle, individuals develop cognitive processes within the social context - aided by expert peers, or cognitive aids and then later use these abilities individually. This is quite consistent with the current state of CSCL that considers the role of social collaborative interactions as an integral part of individual learning processes with technology playing a mediating role.

2.1.2 Group Cognition and Dialectical Constructivism

The earlier sections talk about the different theories that contribute and lead to the perspective that characterizes the concept of collaboration. Collaboration can be associated with dialectical constructivism where social interactions are viewed as a constitutive factor and an integral part of the individual cognition.

Researchers have laid emphasis on social interactions being integral component of group and individual leaning processes. (Vygotsky, 1978; Webb, 1996). Dialectical constructivism emphasizes the constitutive role of social interaction in individual cognitive processes and knowledge construction. Along the same lines, group cognition theory (Stahl, 2006) also talks about individual cognition and social environment exhibiting a dialectical relation. Martinez et al. (2012 d) described the group cognition model based on Stahl's theory. It constitutes of two main cycles: the personal understanding (1) that occurs inside individual's mind and the social knowledge building cycle (2) which includes all the sub- processes that may be present when building social understanding.

Group cognitive processes cannot be gauged from and described by the individual cognitive processes or social interactions alone. Thus these processes must be conceptualized as an amalgamation of both, the individual cognitive process and the social interactions. Collaborative learning is most often associated with these approaches, which emphasize the synergistic, reciprocal and dialectical character of social interactions and individual cognitive processes.

Furthermore, in face-to-face interactions, within a short period of time a large number of cognitive artifacts are generated by the group cognition processes. Each member of the group has to express their thoughts and views to persuade other members about their understanding. Learners, by externalizing their personal understandings, integrate their knowledge with the group understanding as a whole to then appropriate some of it into their own personal understanding. By externalizing their thoughts, potentially, they leave digital traces of the collaborative knowledge building process. As discussed, these digital traces have a huge potential and can be exploited to analyze and support the collaborative knowledge building activity and to support the researchers and teachers in understanding the collaborative and cognitive processes.

To summarize on a higher level, group cognition (Stahl, 2006) focuses on the small groups of more than 2 members, which create shared understanding or meaning across the interactions of the students and this shared meaning is also contained in the cognitive artifacts (abstract or physical). These artifacts are also an integral part of the collaborative activity. Martinez et al. (2013a), whose work I am building upon, used this theory to inform the design of tools, technologies and learning activity and also to define the collaborative learning scenario he aimed to analyze and support.

As already mentioned, the group cognition theory motivates and is reflected in different aspects of the theoretical framework, which would be discussed in the coming sections. The fact that social interactions are one of the most important constituents of collaborative learning also becomes evident and supports the motivation of this thesis to examine the collaborative process in terms of students' interactions and individual cognitive processes. I have leveraged these principles to inform the process of analyzing the collaborative process using the patterns of interactions of students engaged in a collaborative concept mapping activity.

2.2 Assessment of the Collaborative Process

Numerous researchers have already established the importance of analyzing the collaborative process and the interactions that are useful for learning and solving problems in CSCL has (Meier et al., 2007). From a research perspective, the analysis of collaborative process involves identification of key aspects of collaboration that are essential for collaborative learning to happen and a method to assess these aspects.

Based on the group cognition theory (Stahl, 2006) discussed in section 2.1, which stated when groups of students collaborate, the learning process and shared understanding is contained within the students dialogues and the cognitive artifacts, the analysis of these interactions and artifacts results in more awareness about the collaborative process which is of utmost importance. Researchers have defined indicators of collaborative learning that focus on the collaborative process and interactions rather than only the conditions and outcomes. For instance Dimitracopoulou et al. (2004) defined certain indicators of group interactions, which included collaboration intensity, level of participation, transactivity, division of labor to list a few. These indicators can be used to aid the assessment or evaluation of collaboration and even promote awareness about the collaborative process. Martinez Maldonado (2014) used basic statistics on the low level data of students' interactions, group artifacts and individual artifacts obtained as part of a collaborative concept mapping activity to differentiate between groups based on these indicators.

In terms of defining generic assessment methods for CSCL, Meier et al. (2007) defined a rating scheme for the assessment of the quality of collaboration applicable in a multitude of collaborative settings. Meier identified the dimensions of collaboration that determine the quality of collaborative learning encompassing a broad set of collaboration settings and gave a rating scheme to quantifiably measure the quality of collaboration**.** The process dimensions were defined based on an empirical derivation from the data of a collaborative problem solving activity as well as theoretical considerations. Considering the fact that quality of collaborative process has to be assessed instead of frequency measures, and how time consuming it is to transcribe audio or video recordings, a rating scheme was preferred over coding schemes. As compared to coding schemes that are employed to assess the frequency of specific behavioral indicators or types of utterances, a rating scheme allows a more direct assessment of process quality. Rating schemes in general, apply to a larger data-set and rely more on the skills and knowledge of humans whereas coding schemes makes the coders follow strictly defined rubrics for coding. Thus, rating approaches are better in terms of the true evaluation sense (Kahrimanis et al., 2011a).

Martinez et al., 2014 used this rating scheme to assess the quality of collaboration for his work, which will be discussed in the next chapter. This rating scheme comprised of nine dimensions: sustaining mutual understanding, dialogue management, information pooling, reaching consensus, task division, time management, technical coordination, reciprocal interaction, and individual task orientation over four aspects of collaboration namely, communication, joint information processing, coordination and motivation. Transcribed dialogue content from a study involving a psychology and medical student collaborating on a medical case using a desktop-teleconferencing was taken as the data source to extract these dimensions. The content of these dialogues was analyzed against roles of communication, interpersonal relationship, coordination, joint information processing and motivation. As a validation and evaluation method, the rating scheme was then applied to rate data for a new study involving 40 collaborating dyad. Based on the findings on inter-rater reliability, consistency, and validity from this evaluation, the researchers argued that this rating scheme is suitable to be applied in diverse CSCL settings.

For understanding the general processes that entail these dimensions and relate them to the specific concept mapping processes, it is imperative to discuss the theoretical underpinnings behind the first four covered under the two aspects, communication and joint information processing, which I am focusing on in my thesis. This section would be helpful in shaping up the hypothesis for this thesis.

2.2.1. Communication Aspect of Collaboration

Meier identifies the need for establishing “common ground” and “turn taking” when the learners converse. The coordination of content and the process is important for effective communication, which can both be encompassed by the “coordination” aspect of collaboration. But Meier lays focus on the basic communication processes to separate them from higher-level coordination.

As per the communication theory by Clark (1996), pupils engaged in a communication have to take care of both coordinating the communication content and the communication processes through which the content is produced. So at one (the low) level, the speakers and listeners have to coordinate between generating utterances and paying attention, understanding and giving feedback respectively. At the higher level, there has to be co-ordination between what speakers say and mean and what the listeners understand and their feedback.

Co-ordination on content is based on the assumption that the learners already share information or common ground, i.e. they have a set of mutual beliefs, mutual knowledge and assumptions. And as for co-ordination of the processes, they have to continually maintain pace with their evolving “common ground”. So all the actions are built upon sustaining common ground and how it gets aggregated over time during the activity.

The process of updating the common ground moment by moment is called grounding. The learners try to establish a set of mutual beliefs or common ground. Speakers make sure what is being said has been understood and listeners pay attention to understand and give positive/negative feedback or ask for clarifications. Different situations, environment and the media do affect grounding but in general, the process of contributing to a conversation has two phases:

Presentation Phase: "A presents an utterance u for B to consider. He/She does so on the assumption that, if B give evidence e or stronger, he can believe that she can understand what he means by u." (Clark, 1996)

Acceptance Phase: "B accepts utterance u by giving evidence e that she believes she understands what A means by u. She does so on the assumption that once A registers the evidence, he will also believe that she understands." (Clark, 1996)

These phases define the basic flow of communication in a co-located setting and the grounding mechanisms are also embedded within these, and could be witnessed in multiple hierarchies, which would be beyond the scope of this discussion.

2.2.2 Joint Information Processing

Meier defined two dimensions for joint information processing: information pooling" (eliciting information and giving appropriate explanations) and "reaching consensus" (reviewing and assessing information in order to make a joint decision). It requires participants to pool their knowledge and process it as a group and reach a joint decision, which requires a transactive memory system.

The concept of transactivity arises from the famous theoretical model of transactive memory systems given by Wegner et al. (1987) who states that transactive memory system aids the pooling of information. Wegner et al. (1987) talks about how individual minds can come together to form larger, organized and social shared memory systems or group mind. The development of an effective transactive memory system amongst a group of learners involved continuous communication and sharing of each other's' knowledge bases. Effectively, it's analogous to an external memory drive for a computer system where each member is aided by the others' knowledge base.

Transactivity takes into account the social modes of co-construction of knowledge among learners and illustrates the effectiveness and ways in which learners refer to each other's' contributions. Transactivity, in general is reflected in collaborative settings when learners build upon or refer to each other's contributions and has been found to be positively related to knowledge acquisition in collaborative scenarios (Teasley, 1997).

Meier talks about eliciting information, giving explanations and building consensus which are basically different social modes of co-construction of knowledge. Elicitation, externalization, quick consensus and integration/conflict-oriented consensus building are actually defined to be the different degrees of transactivity. Externalization is the process of mutually exchanging knowledge previously exclusive to each partner whereas elicitation is the process of explicitly seeking out information from your partner through questioning. A quick consensus is achieved when both partners accept the other partner's knowledge at its face value and the conversation proceeds to the next chapter of learning whereas integration occurs when a joint decision is reached by disagreement and agreements and is the highest mode of co construction.

2.3 Data Exploration, Analytics and Mining Collaborative Learning Data

There has been a large body of research in CSCL, which has been instrumental in developing tools and technologies that support collaboration in networked environments (Jermann et al., 2000). There also has been an evident shift towards supporting collaborative activities in face-to-face classroom environments where teachers orchestrate the collaborative activities with the help of computer-supported tools.

The data that can be captured when students interact with these systems and with their peers can be put to meaningful use for analysis and supporting the learning process. Roberto et al. (2012) classifies 3 primary uses of this students' data: self-regulation done by students, scaffolding, teaching and assessment performed by teachers and deeper analysis, design oriented modeling and interventions by researchers. The data could be in the form of videos, low-level logged events and students' actions and even their gestures and eye movements that can all be captured automatically and analyzed manually or automatically through the use of tools and technologies. From our discussion in section 2.2, as much as human intervention is required for assessing the deeper level aspects of collaboration, the human evaluation is bogged down by the fact that it is time consuming and cannot be applied to support and analyze the collaboration process in real time. Moving towards the automated analysis of interactions, one approach could be to use the human evaluations as an external reference to define the good and bad collaboration and then estimate metrics or find patterns of interactions that can lead to these good or bad models of collaboration defined by human analysis. For automatically rating the collaboration quality, Kahrimanis et al. (2010), developed an approach wherein he defined automated metrics of interaction from the automatically captured logs of students' data and a model was trained using the collaboration quality ratings given by human raters. The data was categorized in to events like messages, actions, etc. Some of the automated metrics were number of these events, symmetry in these events and so on.

Before even going into specific techniques for the automated analysis of the collaboration process, one needs to know about the different approaches and needs behind the design of collaborative systems for learning. Soller et al. (2005) distinguishes between two different approaches behind designing supportive collaborative learning system. One focuses on capturing students' interaction data and present visualizations that create awareness about the process in general. With this approach the students and teachers have to interpret and decide on future course to improve the collaborative process. The second approach is about creating models on students' interactions and come up with appropriate guidance to the teachers and students to improve the collaborative learning. With this approach the analysis of students' interaction data is usually automatic and done as part of a black box which can potentially run some data mining algorithms and present results that inform the design of appropriate interventions.

It has been proven that data mining and artificial intelligence techniques can be used to better understand the social interactions and cognitive processes within groups of students that lead to high or low collaboration. (Anaya & Boticario, 2011; D'Mello et al., 2011). Over the last two decades there has been significant research going on in this field and I will talk about some notable work relevant to this thesis published by noteworthy researchers.

Soller et al. (2002) did some initial research work in this area using Hidden-Markov Models to identify knowledge sharing instances of students in a collaborative environment. Anaya and Boticario (Anaya & Boticario, 2009) applied both supervised and unsupervised learning methods to gain insights into the collaboration process using statistical indicators of students' interactions in online collaborative forums as the data source. Researchers have also worked on analyzing students' learning behaviors and strategies in terms of the sequential patterns of their actions or learning traces captured by Computer Based Learning Environments (Martinez et al., 2013b; Perera et al., 2009). An exploratory data mining technique which can be used with sequential events or temporal data is sequential pattern mining (Martinez et al., 2013b; Kinnebrew et al., 2012). It finds sequential patterns that occur in a dataset with at least a minimal level of frequency called support (Agrawal & Srikant, 1995). Reflective of research in this area is Perera's et al. work (Perera et al., 2009) which modeled key aspects of teamwork for groups working with an online project management system by proposing alphabets to represent sequential events that can distinguish strong from weak groups. Kinnebrew et al. (2012) presented the differential sequence mining method (DSM) that automatically compares patterns of students' interactions for high and low performing learners and also contextualizes the sequence mining with information about student's performance in terms of productive and nonproductive behavior over the course of learning interactions. Martinez et al. (Martinez et al., 2013 b), whose dataset I will be using in my thesis, also used differential sequence mining to characterize the high and low collaborating groups based on the sequences of students' actions captured in the form of raw low-level logged data from the face-to-face group concept mapping activity on the interactive tabletops.

The use of alphabets by Perera et al. (2009), Kinnebrew et al. (2012) and Martinez et al. (2013a) to represent the sequential actions that can make use of information about authorship, interactions between learners apart from the students' actions is novel. The frequent patterns of interactions can then be associated with different behaviors in the collaborative process and used to differentiate between high and low collaborative groups.

2.4 Concept Mapping and Collaboration

This section gives a brief about concept mapping and then moves towards collaborative concept mapping and how concept maps as learning tools are extremely useful in collaborative settings. I have used the high level collaborative concept mapping processes extracted from the low level students' data to analyze the collaborative learning process. As we have already discussed about the assessment of collaborative processes in the previous sections, it is imperative to discuss different collaborative concept mapping processes in light of the different dimensions for assessing the quality of collaboration that have already been discussed and establish associations between these two taking into account the theoretical underpinnings. This section will end with a brief discussion about the studies on collaborative concept mapping which will set-up a base for understanding the work in this thesis.

2.4.1 Collaborative Concept Mapping

This thesis is completely based upon the work that combines face-to-face collaborative learning facilitated through interactive tabletops using concept maps as a learning tool, which foster meaningful learning amongst students (Martinez-Maldonado, 2014a). There is a large body of research supporting the fact that concept maps are extremely effective learning tools (Novak, 1995). A concept map is a type of graphic organizer that uses labeled nodes to denote concepts and links to denote relationships among concepts.

Concept mapping as a process or strategy facilitates students to externalize their thoughts in a visual form to improve students’ understanding. It allows them to link different ideas, consider multiple perspectives and organize them in a structured way (Martinez-Maldonado, 2014a). While collaborating on a concept mapping activity, students indulge in joint construction of knowledge by exchanging information and negotiating meaning (Cañas & Novak, 2008; van boxtel et al., 2000). Thus, concept maps can serve as vehicles of discussion and negotiation of meaning between students (Novak, 1995). They can be used for facilitating collaborative learning, offering students the opportunity to discuss ideas, present knowledge from diverse perspectives, identify misconceptions, build consensus and reach a joint decision (Chaka, 2010; Novak, 1995; Stahl, 2006).

As already mentioned, I have used the high level concept mapping processes extracted from low level logged data by Martinez et al. (2006). Martinez in one of his work proposed an approach to inform the design of visualizations to support the teachers to analyze the collaborative concept mapping process. The digital traces of students' interaction over a multi-touch interactive tabletop while engaged in collaborative concept mapping activity were exploited for demonstrating the visualizations. For this specific work, the low-level meaningful actions were abstracted in to high-level processes representing the collaborative concept mapping process. These processes were then used to reveal the sequence of interactions in terms of 8 sub-processes identified, namely: sharing individual maps, consulting own maps, generation of concepts, organization of concepts, generating links, revising and layout of the map and speaking.

2.4.2 Collaborative Concept Mapping and Aspects of Collaboration

Moving forward to the discussion about collaborative concept mapping processes with the backdrop of the dimensions defined my Meier for assessing the quality of collaboration; the focus is on explaining how the different processes meaningfully relate to the different theories discussed with respect to the different dimensions. These associations form the base of automatically assessing the quality of collaboration in terms of the patterns of the high-level concept mapping processes.

As per the rating scheme of Meier, with respect to the communication aspect of collaboration, the focus is assessing the processes involved wherein learners or interlocutors converse and are continually engaged in pursuit of reaching a “common ground”. Therefore, the two dimensions defined were “dialogue management” (assess turn taking and other aspects of the communication process) and “sustaining mutual understanding” (assesses the grounding process).

From the communication theory, the presentation phase and acceptance phase were discussed as the two phases which constitute the process of contributing to the conversation. In relation to the concept mapping processes, we can decipher that effective turn-taking in the collaborative activity is reflected by episodes of symmetric talking between participants with other interwoven processes. The discourse should also conform to the sequence of presentation phase followed by acceptance phase.

The grounding process is also reflected in terms of the presentation and acceptance phase, so it’s important to talk about collaborative concept mapping with respect to these phases as well. The presentation phase can be represented with processes like “Adding a Concept”, “Sharing Individual Map” and “Adding Links” with talking in parallel and the acceptance phase should mostly be characterized with “Link” or “Revise” in conjunction with talking which represents evidence of grounding and a signal of understanding. For effective turn-taking discourse should conform to the sequence of presentation phase followed by acceptance phase.

Meier defined two dimensions defined for joint information processing: information pooling” (eliciting information and giving appropriate explanations) and “reaching consensus” (reviewing and assessing information to come to a mutual agreement) and we discussed earlier how the concept of transactivity relates to these two dimensions. Transactivity is reflected in collaborative concept mapping when learners build upon or refer to each other’s’ contributions. More specifically transactivity is evident in the collaborative concept mapping activity when learners consult each other’s’ individual maps, interact with/co-manipulate each other’s’ objects (concepts, links) and add links to their partners’ concepts. Martinez et al. (2012) posits that in a collaborative concept mapping activity transactivity can be strictly measured by taking into account the number of links added by learners using concepts that were added by different learners in the group map. In one of his more recent works, the same author also goes one step further and takes into consideration other interactions by learners on each other’s objects.

3. Research Methodology

For this thesis, I have used data collected from a study carried out by Roberto Martinez-Maldonado, a post-doctoral candidate at the University of Sydney using his concept mapping system built around interactive tabletops. This chapter will introduce the system setup, describe the study and then talk about the dataset that has been used to analyze the collaborative learning process. I will also talk about the specific exploratory data mining technique, namely sequential pattern mining which can help discover patterns of students' interactions and gain insights into the strategies of small groups of students effectively collaborating or even facing problems.

3.1 Study Description

The study was conducted by Martinez et al. (2014b) and captured traces of students' activity on the multi-touch tabletop. It was aimed at building an interface that can automatically analyze and generate visualizations of the face-to-face collaboration and concept mapping process to support the teachers and researchers (Lemonier et al., 2014a; Martinez et al., 2014b). 60 university students from science background, grouped into triads, participated in the study and were asked to build a group concept map on the tabletop around a focus question. The focus question was: *What types of food should we eat to have a balanced diet?* Students were asked to read a one-page article based on the Australian Dietary Guidelines 2011. The study also had pre and post-individual concept map construction phases. Students built an individual map before and after the group concept mapping activity using the CmapTools and these individual sessions took around 30 minutes. For this thesis, I will only focus on the group concept mapping part of the study.

The group activity was partitioned into two phases, namely:

Brainstorming: Students were asked to create general concepts for the group joint map. No propositions were created in this phase. It was advised that 5-10 minutes be spent in this phase.

Linking: Students could add new concepts and add propositions. It was advised to spend 20-25 minutes in this phase. These phases were clearly distinguished in the study with different learning goals and range of student activities.

Martinez et al. (2012b), talks about the synergy between collaborative learning, concept mapping and the use of interactive tabletops, which offers to support and analyze the process of collaborative knowledge building. To explore this potential he has already built a system, Cmate, which can be used for face-to-face collaborative group concept mapping activity on interactive multi-touch tabletops.

Cmate (Martinez-Maldonado et al., 2010) is a tabletop application that provides a platform to the learners to collaboratively construct knowledge through concept maps. The application has a personal menu for each user that can be used to add any new or existing concepts by the students. Students can also view a private screenshot of the map, which helps them in recalling and sharing their thoughts with their group members. The students also have access to their individual maps built earlier, while engaged in the group concept mapping activity. They can add links between concepts and even delete concepts and links. In addition to using a new link for propositions, the students can also choose from the top six linking words based on their usage history that appears on their personal menu. This data from the tabletop, of students' actions was logged in the database. For capturing the nuances of face-to-face interactions and multiple touches on the tabletop, Martinez built a system called Collaid.

Collaid (Martinez-Maldonado et al., 2011), which is a sensing system, was also integrated with the tabletop, capturing the verbal and face-to-face interactions between the students. The tabletop was a regular 46-inch LCD interactive touch screen with the ability to detect multiple touches at a time. Collaid extended this tabletop using overhead sensors to associate the touches to the students performing that action. A microphone was also attached to the tabletop to record the audio and distinguished sounds coming from different locations. The microphones attached with the set-up also logged the instances in the database when the students were engaged in conversation.

The available dataset consists of the following:

1. **Low Level Student Logs:** These consist of all the activities, actions performed by all the students in the groups while partaking the concept mapping activity on the multi-touch interactive tabletop.

2. **Ratings for High and Low Collaborative Groups**: This data consists of the ratings for all the groups across the mentioned nine dimensions of collaboration.

3. **Human Coding of the Sub-Processes**: This consists of time-sequenced coding of the 8 identified sub-processes of the collaborative concept mapping process for each group session.

Martinez et al. (2014a) identified and analyzed the captured data, which may exist in multiple levels. The lowest level of abstraction is the raw data from the logs, which contains a record of all the student actions. The low level logged data comprised of evidence of speech and all the touch actions, which may or may not change the group collaborative product. The meaningful actions extracted out of this raw data, which affect the collaborative artifact, were adding, deleting, editing and moving concepts and links and accessing individual maps. At the next level, the data is categorized into sub processes of concept mapping which is nothing but a sequence of actions of the students, which have a semantic meaning. The aim is to be able to automatically generate visualizations of the collaborative and concept mapping processes with some level of semantic understanding (Lemonier, Martinez-Maldonado, 2014).

Based on the qualitative observations by a human and the sub-processes involved in the concept-mapping activity, a coding scheme was defined and each of twenty sessions were encoded to reveal the succession of sub-processes. The focus was on understanding the meaning of the low level logged data; the content of verbal interactions was not taken into account. The 8 events that were defined are (Lemonier, Martinez-Maldonado, 2014):

1. **Sharing individual map**: Learners opened their individual concept maps and shared them with other students in the group.

2. **Consulting own map**: Learners consulted their individual maps. This is supposed to occur when students open and read from their maps even if they talk in parallel. When the map was opened but the student did not pay attention to it for a long time, the map was considered as closed.

3. **Generation of concepts**: Students add concepts to the map. This usually happens in the first few minutes of the activity and then they still add concepts during the whole activity but the frequency is low.

4. **Organization of concepts**: Students focus on just spreading concepts on the whole surface in order to delete redundancies, create groups or subgroups of concepts, emphasize hierarchies, and add missing concepts.

5. Generation of links: Students focus on creating links choosing for each the direction of the arrow and the linking word. Numerous cognitive conflicts may emerge at that stage. This sub-process may reflect different levels of understandings and misunderstandings as well.

6. **Revising and layout**: Students focus on revision of the map, making small changes to the map (some generation of new links or editing linking words). This sub-process usually emerged as an iterative process with "Generation of links". When students rearrange "aesthetically" links, they see often in the same time new problems in the meanings and have to reorganize their ideas.

7. **Adding the main concept(s)**: Students add the main concept pertaining to the focus question. This is usually added at the top or center of the whole map.

8. **Speaking**: This is just noting down the time when the students speak. The quality of the interaction and the speaker itself is not taken into consideration.

Table 3.1.1 summarizes the 8 sub-processes identified.

Table 3.1.1 Eight Sub-processes of Collaborative Concept Mapping

Talk:	Oral interactions: closed loop communication, instructions, explanation, arguing, reach consensus
KQ:	Key question: focus on the KQ or/and put it at the top or in the center of the map
Gener Stat:	Generation of statements: add concepts (whatever the method: brainstorming, statements extracted from pre-map etc.,)
Organization and Layout:	Layout, structure statements: clustering, creating hierarchies, delete redundancies and add missed concepts.
Link:	Add links and propositions.
Revise:	Revising the map in construction (check with pre- concept map, simplify, give explanations and interpretations, rearrange links)
Map own:	Consulting your own individual map.
Map share:	Sharing your individual map with the group

Quantification of the quality of collaboration

A quantitative assessment of all the sessions on the tabletop was performed to differentiate between the quality of collaboration in groups. The analysis method by Meier et al. (2007) was used to rate collaboration as high or low based on 9 dimensions. The dimensions have already been discussed earlier in the section 2.4. These dimensions were assigned a whole number ranging from -2(very bad) to 2 (very good) to obtain an aggregated result for all dimensions and categorize the groups. A higher score denotes a higher degree of collaboration and vice versa. This gave 10 groups negative scores (-10 to 0). The other 10 groups had scores from 5 to 19. The averages were -4 (±3) (low collaboration) and 13 (±5), (more collaborative), where these differ by at least twice the standard deviation in each case. The sessions were tagged by two raters following the same rubrics as Meier et al., (2007) and a high inter-rater reliability was achieved (Cohen's k = 0.80). This qualitative rating scheme was useful to generate a quantitative measure to distinguish the groups. However, it still has the limitation of requiring human judgment. Researchers have been making efforts to automate this evaluation by having a machine learning model to provide these ratings which would enable teachers to perform real-time analysis using information from the ratings (Martinez-Maldonado et al., 2013b).

3.2 Research Questions

The contribution of this thesis is in terms of analyzing the collaborative concept mapping process and examining the specific patterns of interactions with respect to the different dimensions of assessing the quality of collaboration. All the research questions are motivated to leverage theoretical underpinnings to associate and examine patterns of interactions of the concept-mapping sub-processes with respect to the different dimensions given by Meier et al., (2007) as discussed in the background section.

Q1: **Is the differentiation between high and low groups with respect to the communication aspect of collaboration characterized by more equality of discussion and conformance to presentation and acceptance phase in the frequent patterns of interaction?**

In terms of the collaborative concept mapping processes, the equality of discussion would be reflected if all the learners participate equally in the communication process. Having contextual information about different learners participating in the "Talk" process would have helped gauge this aspect in a better fashion but that information is not available with this data. A secondary indicator could be higher level of "Talk" process with other processes occurring in parallel, which would be more likely to involve multiple learners in the discussion. As for conformance to the presentation and acceptance phase, Table 3.2.1 lists the collaborative concept mapping processes associated with both the phases. The frequent patterns of interactions should have stream of processes related to presentation phase followed by the processes related to acceptance phase.

The following table summarizes the collaborative concept mapping processes with respect to the phases of communication.

Table 3.2.1 Associations between collaborative concept mapping processes and phases of communication

Phases of Communication Process	**Collaborative Concept Mapping Processes**
Presentation Phase	Learners could present their knowledge base just by talking or sharing their individual maps. Generating propositions in parallel to talking would also be indicative of the presentation phase.
Acceptance Phase	Learners can register the evidence of understanding by short independent utterances. Acceptance phase could also be characterized by revisions in parallel to discussions, which would mean some form of clarifications or negotiation.

Q2:

Can we differentiate the high and low groups with respect to the joint information processing aspect of collaboration based on the level of transactivity observed in the frequent patterns?

It has already been established that eliciting information, externalization and reaching consensus are all reflected through transactivity in collaborative settings and this is the motivation behind the above research question.

As I discussed earlier in the background section, transactivity in collaborative concept mapping is reflected when learners build upon or refer to their peers' contributions. It can be measured by taking into considering learners' interactions with each other's objects. The addition of contextual information about owners of objects with every block of processes in the sequence would be useful in this regard and it will be further discussed while talking about pre-processing of data to create alphabets.

In terms of the specific collaborative concept mapping processes involved in the patterns of processes, the results will be analyzed for both high and low groups with respect to the two dimensions defined for the joint information processing aspect and discussed following abduction.

3.3 Data Mining Approach

Sequence pattern mining is an exploratory data mining technique, which can used to differentiate high and low achieving groups by identifying frequent patterns that characterize these groups. This technique has been used keeping in mind that it considers the sequential nature of data and not only statistical information about the data.

Kinnebrew et al. (2012) presented the differential sequence mining method (DSM), which automatically compares patterns that characterize high and low-achieving learners including contextual information of student's actions. Martinez et al. (2012) also used the DSM algorithm with his own alphabets targeting different research questions, which take contextual information regarding the verbal interactions and physical actions performed by learners. He also considers the authorship of objects while learners perform the specific actions.

This thesis uses Martinez's above work as a baseline and builds upon the idea of application of sequence pattern mining to analyze the collaborative concept mapping process with respect to the dimensions of assessment of collaboration.

3.3.1 Data Preparation and Pre-Processing

We discussed in section 3.1 how low-level logged interactions for the collaborative concept mapping activity were encoded into high-level sub-processes.

Refer Table 3.3.1, which summarizes the 8 sub-processes, identified.

For all the 20 groups involved in the study, I extracted the initial dataset in the form of sequence of these sub-processes. The sub-processes occurring in parallel at the same instant were represented as a block of concurrent sub-processes. For instance, a block of these sub-processes would be of the form of: {Talk-Revise-Link}, where Talk, Revise and Link are the sub-processes, which occur at the same instance and the relative order between these doesn't matter. A sequence of sub-processes for a group can be represented by:

{Talk-Revise-Link} -> {Map share-Link} -> ……

where the temporal nature of the processes is represented by "->". So for {A} -> {B}, where "A" and "B" are concurrent sub-processes, "->" implies that B is followed by A in the sequence of sub-processes.

There is much more contextual information available which can be useful to gain insights about different dimensions for which alphabets can be framed. Alphabets are used to encode the sequential concept-mapping sub-processes and make use of this information. In addition to having the sub-processes, they can also have information related to ownership of objects involved in the concept-mapping processes, turn taking or parallelism.

The approach taken for encoding alphabets is based upon the level of abstraction needed for specific research questions. As already seen, the research questions pertain to one of the aspects of collaboration for which different dimensions are defined. The alphabets are also encoded taking selective contextual information, which is based on the theoretical considerations presented for each of the aspect of collaboration. The framing of alphabets is inspired by the work of Martinez which was based upon the way alphabets were framed, to mine group behaviors by Perera et al. (2009) and the suffix nomenclature introduced by Kinnebrew et al. (2012). I will now discuss how each alphabet is associated with a research question in order to investigate the relationship between concept mapping processes, authorship, turn taking and interaction with others' objects.

Alphabet 1 for Q1

Starting with the first research question, as discussed earlier in the background section, there is no need for any other information apart from the sub-processes only, which have already been discussed in light of the aspects of communication. So the alphabet for this is simply represented by the sequences of the sub-processes ({Talk-Revise-Link} -> {Map share-Link} ->….) as described earlier.

Alphabet 2 for Q2

For targeting the second research question we added some contextual information to the sequence of processes for all the groups. The authorship information was added with the processes in the sequence. By adding this information we would achieve the ability to take students' interactions with each other's objects that are involved in different collaborative concept mapping processes into consideration. More specifically, more transactivity would be reflected if the frequent patterns are characterized by higher level of interactions with others' objects accompanied with higher levels of talking in the collaborative concept mapping processes (Molinari et al. 2008; Stahl, 2013). The two keywords, *OSAME* and *ODIFF* were used to convey if only the owner of the object is involved during the process or if different learners are participating respectively.

3.3.2 The Algorithm: Differential Sequence Mining

Having encoded the students' actions into sequences of concept mapping sub-processes and framed alphabets targeting specific research questions, I got 6 datasets of 20 sequences each. There were 20 groups and these sequences were either for high or low collaborating groups for the 4 dimensions I was analyzing. I applied the differential sequence-mining algorithm (DSM) given by Kinnebrew (2012) for all the 4 dimensions to get patterns of the processes that can differentiate high vs. low datasets for each dimension.

A contiguous (consecutive or non-consecutive) subset of a sequence of events that is ordered qualifies as a frequent pattern if it meets a support factor. This support factor is called the sequence support (s-support), which gives the number of sequences in which the pattern occurs irrespective of the number of times it occurs within each group sequence. Following Martinez's et al. (2013) and Kinnebrew's el al. (2012) work I set the s-support threshold to .5 to consider patterns present in at least half of the high or low groups. The DSM algorithm also captures the frequency of the patterns within each group's sequence for the dataset without overlaps and it is called the instance-support (i-support). The i-support of a pattern for a group of sequences is given by the mean of the i-support values of that pattern across all the sequences for the high or low group. For pattern matching, again following previous work, I used 1 as the maximum error gap to allow patterns with the sub-sequences having an edit distance 1 or 0, which means at most one different sub-process is allowed in the sub-sequence. A set of frequent patterns, meeting the threshold s-support value for each dataset is returned as an output of the DSM. These patterns differentiate high collaborative groups from the low collaborative ones across the four dimensions of collaboration.

4. Results and Discussion

4.1 Statistical Analysis

To start the data analysis for this thesis, a preliminary statistical analysis was conducted to extract any relations that may exist between collaborative level of groups (high/low for different dimensions), and the aggregate counts of the sub processes in the students' interaction sequences. The aim of this activity was to examine if the aggregated counts of the sub processes in the group activity patterns had any effect on the group collaboration score for the different dimensions. To do this, first I aggregated the count of the sub processes in the sequences for each of the four dimensions for every group. For each group, the percentage of counts of the sub processes in their group sequences was calculated. These were then grouped together into high and low collaboration groups, and the average for every sub process across all groups was calculated. The results are shown in the tables from 4.1.1 to 4.1.4 below. Table 4.1.1 shows that for "sustaining mutual understanding", the sub-process "Talk" accounted for 42.90% of the total sub-processes for the high groups whereas for low collaborative groups it only accounted for 34.90%. The total aggregated count of all the sub-processes for the high groups was 1928 and for the low groups it was 1200 which gives the total for all the groups to be 3128.We can look at other sub-processes for different dimensions in a similar fashion.

Overall the averages and standard deviations just give a sense how the counts of sub-processes vary across dimensions for high and low groups. We need further statistical tests to determine if meaningful differences exist between the high and low collaborative groups based on these aggregated counts.

Table 4.1.1 Average values of Sub-Processes for Sustaining Mutual Understanding

Sustaining Mutual Understanding	High		Low	
	Average	Std. Dev.	Average	Std. Dev.
Talk	42.90%	7.60%	34.90%	7.60%
KC	0.90%	0.50%	0.90%	0.50%
Map share	6.20%	5.90%	5.00%	3.70%
Map own	12.90%	11.30%	12.90%	9.00%
Link	14.90%	3.80%	20.70%	7.40%
Gener stat	7.90%	5.10%	8.20%	4.50%
Organiz stat	4.50%	1.90%	5.40%	4.30%
Revise	9.80%	4.30%	11.90%	4.40%

Table 4.1.2 Average values of Sub-Processes for Dialogue Management

Dialogue Management	High		Low	
	Average	Std. Dev.	Average	Std. Dev.
Talk	46.78%	7.90%	25.92%	9.40%
KC	1.14%	0.50%	0.50%	0.40%
Map share	7.16%	5.10%	5.00%	6.00%
Map own	16.44%	11.20%	10.08%	9.60%
Link	18.10%	5.20%	14.58%	7.80%
Gener stat	8.77%	5.00%	6.50%	4.90%
Organiz stat	5.45%	2.20%	4.00%	4.70%
Revise	11.51%	4.50%	8.75%	4.70%

4.1.3 Average values of Sub-Processes for Joint Information Pooling

Brainstorming				
Joint Information Pooling	High		Low	
	Average	Std Dev.	Average	Std dev.
Talk	31.80%	6.60%	19.20%	17.80%
KC	0.70%	0.70%	0.70%	0.00%
OSAME	20.20%	4.10%	12.20%	11.40%
Map share	6.10%	6.80%	6.50%	0.50%
Map own	5.70%	7.50%	6.60%	1.30%
Gener stat	13.80%	9.20%	11.50%	3.30%
Organiz stat	13.40%	9.70%	11.50%	2.70%

Link				
Joint Information Pooling	High		Low	
	Average	Std Dev.	Average	Std Dev.
Talk	25.90%	7.30%	22.90%	9.90%
KC	0.10%	0.10%	0.10%	0.20%
OSAME	17.60%	2.60%	19.30%	2.70%
Map share	2.70%	3.30%	0.90%	1.40%
Map own	7.70%	8.90%	5.20%	4.70%
Link	18.00%	6.40%	21.50%	8.80%
ODIFF	13.00%	2.30%	14.70%	2.30%
Gener stat	0.70%	0.80%	1.00%	0.90%
Organiz stat	0.80%	1.30%	1.80%	3.80%
Revise	13.70%	4.90%	12.60%	6.30%

4.1.4 Average values of Sub-Processes for Reaching Consensus

Brainstorming				
Reaching Consensus	High		Low	
	Average	Std Dev.	Average	Std Dev.
Talk	30.00%	4.00%	15.60%	10.70%
KC	1.00%	1.00%	0.60%	0.30%
OSAME	19.00%	4.00%	10.10%	6.50%
Map share	7.00%	7.00%	5.70%	2.90%
Map own	6.00%	8.00%	5.90%	2.80%
ODIFF	9.00%	4.00%	7.30%	3.50%
Gener stat	14.00%	10.00%	9.60%	4.20%
Organiz stat	14.00%	10.00%	11.70%	3.00%

Linking				
Reaching Consensus	High		Low	
	Average	Std Dev.	Average	Std Dev.
Talk	29.00%	4.00%	21.00%	9.00%
KC	0.00%	0.00%	0.00%	0.00%
OSAME	17.00%	2.00%	19.00%	3.00%
Map share	2.00%	2.00%	2.00%	4.00%
Map own	7.00%	8.00%	7.00%	8.00%
Link	17.00%	4.00%	21.00%	9.00%
ODIFF	12.00%	2.00%	15.00%	2.00%
Gener stat	1.00%	1.00%	1.00%	1.00%
Organiz stat	0.00%	0.00%	2.00%	3.00%
Revise	14.00%	5.00%	12.00%	5.00%

Next I performed a regression analysis for the 4 dimensions, using the counts of sub processes as independent variables and the collaboration category (high or low) as dependent variable using SPSS tool. This analysis gives a “B” value (or slope), which is a measure of how the sub-process count and collaboration levels are related. “B” gives the magnitude of the change that will take place in the dependent variable per unit change in an independent variable with all other variables as constant. A positive “B” signifies a positive relation and vice versa. However, this merely gives the closest representation of the two variables on a straight line. It may or may not be an actual predictor of the observed data. I examined whether there was a significant relationship between the collaboration scores and the aggregated count of sub-processes using a stepwise regression. Using the forward-LR method, each variable was entered into the process if its contribution to the score statistic was more than 0.05, which is a default entrance criterion for this model. Therefore, variables with the most statistical significance were added first, followed by the second and so on, till they did not have a significant impact on the model fit. This stepwise technique starts with step 0, which creates a model only using an intercept. It then adds the statistically significant variables, one at a time, to the regression using the entry probability as previously described. As seen in Table 4.1.5, for sustaining mutual understanding dimension, “link” was the only sub process that satisfies the criteria to enter step 1. The “B” value in this table is negative, which implies that with an increase in count of “link”, the probability of collaboration reduces. The extremely low “sig.” value reaffirms that this parameter is statistically significant in the regression equation. However, since a stepwise model already takes into account the score statistics while entering a variable into the equation, the “sig.” value is irrelevant and is not used further to draw any conclusions. There were no other independent variables that were statistically significant and thus the regression finished at step 1. From step 1(Table 4.1.6) it could be seen that the adjusted R square value for this regression was 0.398. The R square value is a measure of how well the predicted model from a regression fits the actual data. The higher the R square score, the better is the accuracy of predictions. Thus if a model were built using the regression values for a low r-square model, the accuracy of predictions from the model would also be low. As we can see from Table 4.1.6, this is relatively low which also implies that this model built upon the count of the sub-processes plotted against the collaboration score is not an effective model to study and predict group collaboration.

Table 4.1.5 Regression Results for Sustaining Mutual Understanding

Variables in the Equation

		B	Sig.
Step 1[a]	Link	-.131	.028
	Constant	4.243	.018

a. Variable(s) entered on step 1: Link.

Table 4.1.6 Link vs. Score Regression Model Summary- Sustaining Mutual Understanding

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	19.075[a]	.289	.398

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Similar regressions were performed for “Dialogue Management”, and “Joint Information Pooling” and “Reaching Consensus” in brainstorming and linking phases. For all of these activities, the regression terminated at step 0. This implies that none of the independent variables were statistically significant to enable creation of a prediction model.

From the above analysis we can conclude that these aggregated statistical counts for each of the sub-processes were not enough to create a predictive model of high or low collaboration levels. Moreover one more limitation is that these counts are aggregated at the end of the session, which makes it unsuitable for a real time analysis as well. These averages and counts do not exploit the sequential nature of the interactions and thus in the next section, I present an approach which can be used to explore patterns of interactions that can help distinguish high vs. low collaborative groups.

4.2 Sequential Pattern Mining Results

I executed DSM algorithm for the four dimensions of collaboration with different alphabets as discussed and only selected the s-frequent patterns comprising at least two blocks of processes ({Talk-Link-} -> {Talk-Revise-}...) and having a confidence factor of at least 90% (p value<=0.1) for i-support (instance support) values that distinguish high from low groups for the different dimensions of collaboration. The p values here are not used as a confidence factor to differentiate between low and high groups. It signifies an important measure to limit patterns that are most likely to be used differentially in the groups. The top differential patterns were chosen based on comparing the mean i-support values between the high and the low groups for the frequent patterns (diff= i-value (high)- i-value (low)). The frequent patterns with the maximum "diff" value were selected as the top patterns, which provides a useful heuristic for most likely differential patterns.

For the first research question, targeting the first two dimensions of collaboration, the dataset wasn't partitioned into the brainstorming and linking phases. The reason behind taking this decision was that these two dimensions are concerned with mutual knowledge sharing and turn taking during the collaborative process, for which it makes more sense to consider the dataset collectively for both the phases. The differentiation for these two dimensions is not intuitive across the two phases as the brainstorming phase is marked by only addition of most general concepts.

For the next research question, which seeks to address the "information pooling" and "reaching consensus" dimensions of collaboration, the analysis of the frequent sequences was carried out separately for the brainstorming and linking phase of the collaborative concept mapping activity.

Q1: Is the differentiation between high vs. low groups with respect to the communication aspect of collaboration characterized by more equality of discussion and conformance to presentation and acceptance phase in the frequent patterns of interaction?

After applying the DSM algorithm on Alphabet 1 for the first two dimensions of collaboration, I got the frequent patterns given in Table 4.2.1 for both high and low collaborative groups. I obtained 30 differential patterns for sustaining mutual understanding and 22 for dialogue management. The top-3 most differential patterns that were found in each subset for each dimension are shown in the table. As discussed in section 3.2 and as expected, the frequent patterns for the high groups for both "Sustaining Mutual Understanding" and "Dialogue Management" are marked by higher percentage of "Talk" process occurring in parallel with other sub-processes. The frequent patterns for high groups have "Talk" for each process block in the sequence. In contrast, the low groups for these two dimensions have processes occurring without any speech in parallel in some of the sequences.

Table 4.2.1 Top Sequential Patterns for Sustaining Mutual Understanding for high and low groups

Pattern	I Freq (|Hi-Low|)	S-support Diff (|Hi-Low|)	S-frequent group	P-value	I-Support (Hi)	I-Support (Low)
{Talk-Map share-} -> {Talk-} -> {Talk-}	1.208	0.53	High	0.02	1.92	0.71
{Talk-} -> {Talk-Revise-} -> {Talk-} -> {Talk-}	1.24	0.53	High	0.01	1.38	0.14
{Talk-Link-} -> {Talk-} -> {Talk-Revise-} -> {Talk-}	0.85	0.53	High	0.02	1	0.14
{Talk-Link-Map own-} -> {Talk-Link-Map share-}	0.71	0.71	Low	0.008	0	0.71
{Link-Revise-} -> {Link-}	2.06	0.57	Low	0.04	0.08	2.14
{Talk-Link-Map own-} -> {Talk-Link-} -> {Link-}	0.76	0.57	Low	0.04	0.23	1

Table 4.2.2 Top Sequential Patterns for Dialogue Management for high and low groups

Pattern	I-Freq (|Hi-Low|)	S-Support Diff(|High-low|)	S-frequent Group	P-value	I-Support(Hi)	I-Support (Low)
{Talk-Revise-} -> {Talk-Map share-}	0.71	0.5	High	0.006	0.71	0
{Talk-} -> {Talk-Revise-} -> {Talk-Link-Revise-}	0.83	0.5	High	. 019	1	0.17
{Talk-Map share-} -> {Talk-} -> {Talk-}	1.19	0.5	High	0.02	1.86	0.67
{Talk-Revise-} -> {Talk-Link-} -> {Talk-}	1.35	0.64	High	0.028	1.86	0.5
{Talk-Link-Revise-} -> {Link-} -> {Link-}	0.5	0.5	Low	0.07	0	0.5
{Talk-Link-Revise-} -> {Link-Revise-} -> {Link-}	0.5	0.5	Low	0.07	0	0.5
{Talk-} -> {Talk-Revise-} -> {Talk-Gener stat-Revise-}	0.5	0.5	Low	0.07	0	0.5

For both the dimensions the high groups conform to the processes in the presentation phase followed by the acceptance phase (refer Table 3.2.1). For instance, the frequent patterns *{Talk-Map share-} -> {Talk-} -> {Talk-}*, *{Talk-} -> {Talk-Revise-} -> {Talk-} -> {Talk-}*, *{Talk-Link-} -> {Talk-} -> {Talk-Revise-} -> {Talk-}*, *{Talk-} -> {Talk-Revise-} -> {Talk-Link-Revise}* reflect the presentation phase with processes like "Talk", "Map share", "Link" occurring independently or concurrently, followed by the acceptance phase marked by "Talk" representing short utterances of approval or "Revise" and "Talk" representing some clarifications or negotiation. Delving in one more layer deeper into the specific dimensions of collaboration and comparing the frequent patterns that characterize them, it is evident that for “Sustaining Mutual Understanding” the high groups have all the frequent sequences with well-defined presentation phases which start with processes “Talk”, “Map share”, “Link” followed by an acceptance phase with independent “Talk” process reflecting evidence of approvals with short utterances. On the other hand, the high groups for “Dialogue Management” have some frequent sequences having “Talk” and “Revise” in the beginning which is then followed by “Talk”, “Link”, “Revise” or “Map share” (independently or concurrently). This represents a discussion ensuing as part of a justification or clarifications in which different learners indulge in making the point clear, possibly taking turns to present different point of views.

For the low groups, the frequent patterns do not seem to conform to the order of presentation phase followed by the acceptance phase. For instance, the frequent pattern “{Link-Revise-} -> {Link}” for the “Sustaining Mutual Understanding” dimension doesn’t comprise of processes which reflect any discussion or actions pertaining to presentation and then acceptance phases. Similarly most of the frequent patterns for both the groups are in some cases marked by a presentation phase but it is followed by independent “Link” process in most of the cases which doesn’t reflect an acceptance phase and as a whole, the patterns don’t reflect mutual knowledge building or turn taking. In terms of the s-support and i-support values, it has already been discussed that the s-support values indicate the percentage of the high groups (or low groups) the particular sequence occurs in. For instance, we have 13 high groups and 7 low groups for the dimension “Sustaining mutual understanding”. Then a score of .53 for the frequent pattern “{Talk-Map share-} -> {Talk-} -> {Talk-}” means that it occurs in 7 of the 13 high groups. Similarly these numbers can be easily interpreted for all the other frequent patterns. For the i-support values, which signify how frequently the patterns occur within the groups, the i-support (high) value 0.71 for the frequent pattern “{Talk-Revise-} -> {Talk-Map share-}” (refer Table 4.2.2) means that it occurs 0.71 times within all the high groups on an average. The i-support (low) value 0.00 for the same pattern indicates that it doesn’t occur at all in those 7 low groups for the specified dimension. The i-support (high-low) scores, which we have used for getting the top-k differential patterns, indicate the magnitude of differentiation with which these patterns occur in high vs. low groups.

Frequent patterns, that characterize a high or a low group and don’t occur at all in the other group, can be more meaningful. It is also interesting to note that the i-support values for patterns that characterize the high or low groups are not that high. For “sustaining mutual understanding” the values lie in the range 1.00 – 1.92 for the high groups and between .71-2.04 for the low groups. For “dialogue management” it lies between .71 – 1.86 for high groups and .50 for all the low groups. This indicates that there would be 1-2 episodes of around 2-5 seconds each of these frequent sequences (for low and high groups) in the whole group activity. These might be just because of the limitation of the dataset, as it is too small but we can still draw one general inferences that it is important to consider more number of frequent patterns for both the high and low groups to get a clear differentiation between the two in terms of these interaction patterns.

Q2: Can we differentiate the high and low groups with respect to the joint information processing aspect of collaboration based on the level of transactivity observed in the frequent patterns?

After running DSM on the sequence of processes with authorship information (Alphabet 2), I obtained the frequent patterns given in Table 4.2.3 and Table 4.2.4 for both "information pooling" and "reaching consensus" (brainstorming and linking phase), for high and low collaborative groups. I obtained 47 differential patterns for information pooling dimension and 218 differential patterns for reaching consensus. I present the top 3-4 most differential patterns. As expected, the brainstorming phase is characterized by majority of the interactions on own objects rather than other learners' objects for both high and low collaborative groups across both dimensions (information pooling and reaching consensus). This can be attributed to the fact that during the brainstorming phase, learners are mostly involved in adding the concepts independently and do not interact with each other's objects. Interestingly, the high and low collaborative groups are differentiated in terms of the processes in the frequent patterns with respect to the joint information-pooling dimension. The high collaborative groups are more focused towards adding propositions to the group map while the low collaborative groups tend more towards organizing the map. For the reaching consensus dimension, the results obtained weren't significant and meaningful.

Table 4.2.3 Top Sequential Patterns for Joint Information Pooling for high and low groups

	Pattern	I-Freq (|Hi-Low|)	S-Support Diff (|High-low|)	S-frequent Group	P-value	I-Support (Hi)	I-Support (Low)
Brainstorming	{Talk-Gener stat-OSAME-} -> {Talk-Gener stat-OSAME-}	0.9	0.5	Hi	0.03	1.07	0.17
	{Talk-Gener stat-Map share-OSAME-}	1.5	0.5	Hi	0.02	1.5	0
	{Talk-Gener stat-OSAME-} -> {Talk-Gener stat-OSAME-} -> {Talk-}	1.43	0.5	Hi	0.07	1.93	0.5
	{Talk-Gener stat-Map own-OSAME-} -> {Talk-Gener stat-Organiz stat-Map own-OSAME-}	0.83	-0.5	Lo	0.09	0	0.83
	{Gener stat-Organiz stat-OSAME-}	1.38	-0.7	Lo	0.08	0.29	1.67
	{Gener stat-Organiz stat-Map own-OSAME-}	0.5	-0.5	Lo	0.08	0	0.5
Linking	{Talk-Revise-Map own-OSAME-} -> {Talk-Revise-Map own-ODIFF-}	3.66	0.64	Hi	0.01	3.86	0.2
	{Talk-Revise-Map own-OSAME-} -> {Talk-Revise-Map own-ODIFF-} -> {Talk-Revise-Map own-OSAME-}	2.07	0.64	Hi	0.00	2.07	0
	{Talk-Revise-Map own-ODIFF-} -> {Talk-Revise-Map own-OSAME-}	3.53	0.64	Hi	0.01	3.93	0.4
	{Link-Revise-ODIFF-} -> {Link-Revise-OSAME-} -> {Link-OSAME-}	0.8	-0.6	Lo	0.10	0	0.8
	{Link-Revise-OSAME-} -> {Link-Revise-ODIFF-} -> {Link-OSAME-}	0.6	-0.6	Lo	0.07	0	0.6
	{Link-Revise-ODIFF-} -> {Link-OSAME-} -> {Link-OSAME-}	0.8	-0.6	Lo	0.10	0	0.8

Table 4.2.4 Top Sequential Patterns for Reaching Consensus for high and low groups

	Pattern	I-Freq (|Hi-Low|)	S-Support Diff(|High-low|)	S-frequent Group	P-value	I-Support(Hi)	I-Support (Low)
Brainstorming							
	{Talk-} -> {Talk-Gener stat-Map share-OSAME-}	0.4	0.5	Hi	0.05773	0.5	0.1
Linking	{Talk-Revise-ODIFF-} -> {Talk-Revise-OSAME-} -> {Talk-Link-OSAME-} -> {Talk-}	0.767	0.800	Hi	0.0167	1.10	0.33
	{Talk-Revise-ODIFF-} -> {Talk-} -> {Talk-}	1.589	0.500	Hi	0.0458	2.70	1.11
	{Talk-Revise-ODIFF-} -> {Talk-Revise-OSAME-} -> {Talk-} -> {Talk-}	1.600	0.500	Hi	0.0392	2.60	1.00
	{Link-OSAME-} -> {Link-ODIFF-} -> {Talk-Link-OSAME-}	0.900	0.556	Lo	0.0441	0.10	1.00
	{Link-OSAME-} -> {Link-Map own-OSAME-}	0.456	0.556	Lo	0.0424	0.10	0.56

Moving to the linking phase, for the high collaborative groups, as expected, students interact more with others' object and also communicate more while engaged in interacting with other learners' objects for both the dimensions. The low collaborative groups on the other hand have lower interactions involving different owners of objects and almost negligible communication. As for the collaborative concept mapping processes, the high and low collaborative groups have interesting trends in the frequent patterns that characterize the joint information-pooling dimension. The high collaborative groups are more inclined towards the revision of the group map and individual students consult their own maps while talking to each other and revising the map. As already pointed out students do interact with each other's objects while engaged in the revision activity. The low collaborative groups tend more towards adding links to the group map and revising the group map but students communicate less and act more on their own objects. For the "reaching consensus" dimension, the high groups are more focused towards revising their group map and they also interact with each other's' objects during this revision process. In comparison to the "joint information pooling" dimension, the high groups for this dimension do not consult their map and the frequent patterns are marked by independent "Talk" process towards the end of the sequence. This behavior can be explained by the fact that for information pooling, students have to share or refer to their knowledge base and then pool the shared or unshared information they possess. The independent "Talk" process in the end of sequence of frequent patterns for "reaching consensus" marks the joint decision making that students achieve by communicating with each other. For the low collaborative groups, the trends are quite similar as for the "information pooling" dimension, adding links to the group map is more prominent, students communicate less and act more on their own objects.

5. Conclusion and Future Work

This thesis was aimed at exploring whether we can associate frequent patterns of students' interactions with the dimensions of assessing the quality of collaboration. My work was based upon the study conducted by Martinez et al. (2013a) and dataset from that study. The students' interactions were represented in terms of sequences of high-level concept mapping processes encoded for each of the 20 groups of the study. I used an exploratory data mining technique called the sequence pattern mining to extract frequent patterns that can characterize strong and weak collaborative groups across dimensions (first four) of collaboration quality defined by Meier. The results although preliminary, were effective in addressing the research questions that were concerned with characterizing the different groups across the dimension of collaboration.

First up, the statistical analysis revealed that there were some differences in high vs. low groups for the four dimensions in terms of the aggregated averages of the concept mapping sub-processes. These differences were not meaningful considering the standard deviation values. I also performed regression analysis for the four dimensions for high and low collaboration groups to find if the aggregated counts of the sub-processes are predictive of the high vs. low collaboration. The regression results were not significant for any of the dimension. Moreover these statistical aggregations of the discrete events/interactions cannot be effectively used to evaluate collaborative learning in general and to overcome this limitation, I used differential sequence mining (DSM) which takes into account the sequential nature of these interactions.

The analysis was based on constructing alphabets for targeting each research question and preparing the data to reveal the succession of the sub-processes for all the groups. The alphabets were designed to capture the interactions of students with each other's objects, parallelism along with the concept mapping sub-processes in the patterns.

The frequent patterns for the "sustaining mutual understanding" and "dialogue management" dimensions showed tendency of the better groups to talk more as compared to the weaker ones. The high groups for both the dimensions conformed to presentation phase followed by the acceptance phase in terms of the concept mapping sub-processes. The frequent sequence of concept mapping sub-processes for "sustaining mutual understanding" (for high groups) represented establishing common ground by students. For the "dialogue management" dimension patterns of sub-processes representing turn taking, clarifications and justifications were evident.

For the "joint information pooling" and "reaching consensus' dimension, I used Alphabet 2, which had the authorship information of the objects and also partitioned the dataset into the brainstorming and linking phase. The frequent patterns of the sub-processes differentiated the high vs. low groups for both these dimensions as well. For the brainstorming phase, the interactions were majorly with own objects and for "joint information pooling", the high groups were mostly involved with generating propositions where the low groups were more inclined towards organizing concepts in the group map. The results obtained for "reaching consensus" weren't significant. For the linking phase, the high groups for both the dimensions interacted with each other's' objects more and the low groups didn't. For the "joint information pooling", the high groups were characterized by processes representing accessing individual knowledge bases and pooling information together. These processes were missing for the low groups. For the "reaching consensus' dimension the frequent patterns for high groups evidently represented joint decision making.

The frequent patterns of the high level concept mapping sub-processes obtained are useful for multiple reasons. These patterns represent students' interactions in terms of the collaborative concept mapping processes, which in itself is a novel formulation. The associations of these frequent patterns with the four dimensions of assessing the collaboration quality takes the work beyond what has been done before and presents students' collaborative strategies at a more granular level. It has the potential to enable teachers to orchestrate more efficiently, the collaborative processes of multiple groups in classrooms. The different alphabets constructed for targeting the two research questions by adding selective contextual information builds upon previous work and provides evidence that there is a need to design these alphabets carefully in order to get precise information and insights for different scenarios. This work can also serve as a platform for providing support to monitor the collaborative process for the students and teachers by creating real time visualizations of the group activities and even assist teachers to provide appropriate feedback and interventions to the groups at the right moments. For instance, these frequent patterns found for the low and high groups for the four dimensions can be used as a benchmark to compare interaction patterns of other groups.

Teachers can be informed of the patterns occurring in real time for the low and high groups for them to intervene with particular groups to provide help. One future step for this work is also analyzing the frequent patterns for the high and low groups in the recorded videos and gain awareness about the different strategies followed by different groups. This high level information about the process and strategies followed by high and low collaborating groups can be associated with respective frequent patterns and presented to the teachers.

6. References

Agrawal, R. & Srikant, R. (1995). Mining sequential patterns. In Proceedings of the Eleventh IEEE International Conference on Data Engineering (ICDE). 3–14.

Alavi,M. (1994). Computer-mediated collaborative learning: An empirical evaluation. In MIS quarterly Volume 18 Issue 2 pp. 159-174.

Anaya, A. R., & Boticario, J. G. (2009). A Data Mining Approach to Reveal Representative Collaboration Indicators in Open Collaboration Frameworks. Educational Data Mining 2009: Proceedings of the 2nd International Conference on Educational Data Mining, 210–219.

Basheri, M., Burd, L., & Baghaei, N. (2012). A multi-touch interface for enhancing collaborative UML diagramming, in Proc. OZCHI pp. 30-33.

Cañas, A. J., & Novak, J. D. (2008). Next step: consolidating the cmappers community. In Proceedings of the 3rd International Conference on Concept Mapping (pp. 532-539). OÜ Vali Press.

Chaka, C. (2010). Collaborative Learning: Leveraging Concept Mapping and Cognitive Flexibility Theory. In Handbook of Research on Collaborative Learning Using Concept Mapping (pp. 152-170). Birmingham, UK.

Cindy E. Hmelo-Silver, Clark A. Chinn, Carol Chan, & Angela M. O'Donnell. (2013a). Chapter 4. Theories of cognition in collaborative learning. In The International Handbook of Collaborative Learning (pp74–90). Abingdon, Oxon: Routledge.

Clark, H. H. (1996).Using language. Cambridge MA: Cambridge University Press.

Collins, A., & Kay, J. (2011). Who did what? Who said that? : Collaid: an environment for capturing traces of collaborative learning at the tabletop. Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces (ITS'11), 172–181.

Dillenbourg, P. (1999). What do you mean by' collaborative learning '? Collaborative Learning Cognitive and Computational Approaches, 1(6), 1–15.

Dillenbourg, P., & Fischer, F. (2009). Chapter 1 The Evolution of Research on Computer-Supported Collaborative Learning: From Design to Orchestration, Technology-Enhanced Learning: Principles and Products pp.3–20, Springer, 2009.

Dillenbourg, P., Zufferey, G., Alavi, H., Jermann, P., Do-Lenh, S., & Bonnard, Q. (2011). Classroom orchestration: The third circle of usability. In Proceedings of the International Conference on Computer Supported Collaborative Learning (CSCL 2011) (pp. 510-517).

Dimitracopoulou, A., Mones, A., Dimitriadis, Y., Morch, A., Ludvigsen, S., & Harrer, A. (2004). State of the art on interaction analysis: Interaction analysis indicators.

Do-Lenh, S. (2012). Supporting Reflection and Classroom Orchestration with Tangible Tabletops". PhD Thesis In:École Polytechnique Fédérale de Lausanne.

D'Mello, S., Olney, A., & Person, N. (2011). Mining Collaborative Patterns in Tutorial Dialogues. JEDM - Journal of Educational Data Mining, 2(1), 1-37

Jermann, P., Soller, A., & Muehlenbrock, M. (2000). From Mirroring to Guiding : A Review of State of the Art Technology for Supporting Collaborative Learning.

Kahrimanis, G., Avouris, N., & Komis, V. (2011a). Interaction analysis as a tool for supporting collaboration: An overview. In T. Daradoumis, S. Caballé, A. A. Juan, & F. Xhafa (Eds.), Technology-enhanced systems and tools for collaborative learning scaffolding (pp. 93–114). Berlin Heidelberg: Springer.

Kinnebrew, J. S., Loretz, K. M., & Biswas, G. (2012). A Contextualized,Differential Sequence Mining Method to Derive Students' Learning Behavior Patterns. In Journal of Educational Data Mining (JEDM).

Kreijns, K., Kirschner, P. A., & Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: a review of the research. In Computers in Human Behavior, 19(3), 335-353.

Lemonier, C., Martinez-Maldonado, R., & Yacef, K. (2014a). Tell me about the process: Exploring Visual Analytics to unveil Collaboration and Concept Mapping Processes. In Workshop on Intelligent Support for Learning, ITS.

Martinez-Maldonado, R. (2014b). Analyzing, visualizing and supporting collaborative learning using interactive tabletops”. PhD Thesis, University of Sydney, Australia.

Martinez-Maldonado, R., Dimitriadis, Y., Kay, J., Yacef, K., & Edbauer, M.-T. (2013b). MTClassroom and MTDashboard: supporting analysis of teacher attention in an orchestrated multi-tabletop classroom, in Proc. CSCL. pp. 119-128.

Martinez-Maldonado, R., Dimitriadis, Y., Martinez-Mones, A., & Yacef, K. (2013a). Capturing and analyzing verbal and physical collaborative learning interactions at an enriched interactive tabletop. In IJCSCL 8(4): pp. 455-485.

Martinez-Maldonado, R., Kay, J., & Yacef, K. (2012b). Analysing knowledge generation and acquisition from individual and face-to-face collaborative concept mapping. In: Proc. CMC. pp. 17-24.

Martinez-Maldonado, R., Yacef, K., Kay, J., Kharrufa, A., & Al-Qaraghuli, A. (2011). Analysing frequent sequential patterns of collaborative learning activity around an interactive tabletop, in Proc. EDM. pp. 111-120.

Martinez-Maldonado, R., Yacef, K., Kay, J., & Schwendimann, B. (2012a) .An interactive teacher’s dashboard for monitoring multiple groups in a multi-tabletop learning environment. In Proc. ITS. pp. 482-492.

Meier, A., Spada, H., & Rummel, N. (2007) .A rating scheme for assessing the quality of computer-supported collaboration processes. In International Journal of Computer-Supported Collaborative Learning (ijCSCL), 2(1), 63-86.

Mercier, E. (2015). Researching and Designing for the Orchestration of Learning in the CSCL Classroom.

Molinari, G., Sangin, M., Nüssli, M.A., & Dillenbourg, P. (2008). Effects of knowledge interdependence with the partner on visual and action transactivity in collaborative concept mapping. In Proceedings of the Proceedings of the 8th international conference on International conference for the learning sciences - Volume 2 (pp. 91-98). 1599883: International Society of the Learning Sciences

Novak, J. (1995). Concept mapping to facilitate teaching and learning. Prospects, 25(1), 79-86.

Novak, J., & Cañas, A. (2008). The Theory Underlying Concept Maps and How to Construct and Use Them, Technical Report IHMC CmapTools 2006-01.

Perera, D., Kay, J., Koprinska, I., Yacef, K., & Zaiane, O. (2009). Clustering and Sequential Pattern Mining of Online Collaborative Learning Data. IEEE Transactions on Knowledge and Data Engineering 21(6), 759-772.

Race, P. (2001). A briefing on self, peer & group assessment. Learning and Teaching Support Network.

Roschelle, J., & Teasley, S. D. (1995). The Construction of Shared Knowledge in Collaborative Problem Solving. In Computer Supported Collaborative Learning. (pp. 69-97): Springer

Soller, A., Martinez, A., Jermann, P., Muehlenbrock, M. (2005). From Mirroring to Guiding: A Review of State of the Art Technology for Supporting Collaborative Learning. IJAIED 15(4): pp. 261-290.

Soller, A., Wiebe, J., & Lesgold, A. (2002). A machine learning approach to assessing knowledge sharing during collaborative learning activities. In Proceedings of the Proceedings of the Conference on Computer Support for Collaborative Learning: Foundations for a CSCL Community (pp. 128- 137). 1658635: International Society of the Learning Sciences.

Stahl, G. (2006). Group Cognition: Computer Support for Building Collaborative Knowledge: MIT Press.

Stahl, G. (2013). Theories of Cognition in Collaborative Learning. In C. E. Hmelo-Silver, A. Chinn, C. K. K. Chan & A. M. O'Donnel (Eds.). The International Handbook of Collaborative Learning (74-90). New York, NY & London, UK: Routledge.

Van Boxtel, C., Van Der Linden, J., & Kanselaar, G. (2000). Collaborative learning tasks and the elaboration of conceptual knowledge. Learning and Instruction, 10, 311- 330

Vygotsky, L. (1978). Mind in society: Cambridge, MA: Harvard University Press
Johnson, R. T., & Johnson, D. W. (1986). Action Research: Cooperative Learning in the Science Classroom. Science and Children, 24(2), 31-32.

Webb, N.M., & Palincsar, A.S. (1996). Group processes in the classroom. D. C. Berliner & R. C. Calfee (Eds.), Handbook of educational psychology (pp. 841–873). New York: Simon & Schuster/Macmillan.