

Premchaiswadi, Wichian; Porouhan, Parham

## Article

# Process modeling and decision mining in a collaborative distance learning environment

Decision Analytics

## Provided in Cooperation with:

Springer Nature

*Suggested Citation:* Premchaiswadi, Wichian; Porouhan, Parham (2015) : Process modeling and decision mining in a collaborative distance learning environment, Decision Analytics, ISSN 2193-8636, Springer, Heidelberg, Vol. 2, Iss. 1, pp. 1-34, <https://doi.org/10.1186/s40165-015-0015-5>

This Version is available at:

<https://hdl.handle.net/10419/161802>

### Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

### Terms of use:

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<http://creativecommons.org/licenses/by/4.0/>

RESEARCH

Open Access



# Process modeling and decision mining in a collaborative distance learning environment

Wichian Premchaiswadi and Parham Porouhan\*

\*Correspondence:  
parham@siam.edu  
Graduate School  
of Information Technology,  
Siam University, 38 Petkasem  
Road, Bangkok 10163,  
Thailand

## Abstract

This paper is divided into four main parts. In the first part of the study, we identified the most significant factors that affect the performance of groups in collaborative learning situations. The results showed that the extent of communication, interactions and involvement/participation between students have crucial impacts on the performance of groups. In the second part of the study, we defined and explained specific alphabets and keywords derived from a collected event log during a distance learning activity using a real-time multi-user concept mapping service. Our aim was to interpret the data in such a way that eventually can increase the instructor's awareness about entire the collaborative process. In the third part of the study, we used several statistical and process mining techniques in order to discover and compare distinguished patterns of interaction and involvement between the groups with high and low performance. The results showed that the extent of students' interaction was four times greater in the high performance groups. Similarly, the extent of students' involvement and participation was three times greater in the high performance groups compared with the low performance groups. In the fourth part of the study, we analyzed the extent of communication with respect to textual and semantic contributions of the students written/typed and shared in the chat rooms during the online distance activity. The results showed that the level of students' communication was two times greater in the groups with high performance. And finally, we applied a Decision Tree/Rules technique to extract a model of decisions (as well as their possible consequences) about performance of groups in collaborative learning situations.

**Keywords:** Process mining, Fuzzy mining, Computer-supported collaborative learning, Distance learning, Concept mapping, Collaborative communication, Patterns of interaction process, Level of involvement, Decision Point Analysis

## Definitions and context

Although the majority of readers might be familiar with most of the issues and concepts discussed in this paper, we were wondering whether some of the abbreviations and terminology used in the present study make sense for few readers at the first glance. Therefore, before talking about benefits of the study we decided to completely (but in a few words) provide a list of the most important terms and contexts used in the upcoming sections as the following:

*Web 3.0.* In learning, this refers to application of Internet-based services—such as communication tools, wikis, social networking sites and folksonomies—which focus

more on online collaboration and sharing (contents and solutions) among students. (Coleman 2011; Spivack 2015).

*Flowchart.com<sup>a</sup>* This is a free online real-time, multi-user and collaborative concept mapping service which works on any Operating System. (Web 3.0 Software Service 2014). In this study we used flowchart.com as a platform for distance concept mapping activities between small groups of students at a private university in Thailand (Flowchart.com 2014).

*Concept map* This is a diagram that illustrates suggested relationships between concepts and their linking phrases and causes (Concept map 2015).

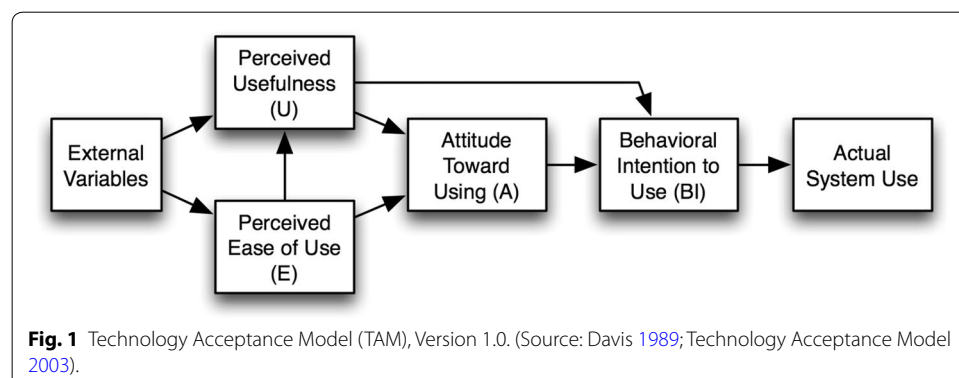
*Technology Acceptance Model (TAM)<sup>b</sup>* This is an information systems theory that explains how users may accept and use a new technology. (Davis 1989; Technology Acceptance Model 2003). In this study, small groups of students of Bachelor of Business Administration program were asked to build a TAM concept map model during a distance learning activity via flowchart.com (see Fig. 1).

*Process mining<sup>c</sup>* This is an almost new process management technique that provides process discovery and conformance checking based on event logs (Aalst 2009; Process mining 2011). In this study, we applied several process mining techniques in order to discover distinguished patterns of behavior and interaction between small groups of students in a distance concept mapping activity at a private university in Thailand.

*ProM* This is an Open Source framework for process mining algorithms (ProM 2011). In this study we applied ProM Fuzzy Mining algorithm, Social Network Miner and ProM Decision Tree/Rules J48 algorithm (Decision Point Analysis) in order to discover and analyze distinguished patterns of behavior and interaction between high and low achieving groups.

*Disco<sup>d</sup>* This is a process mining tool developed by Fluxicon Process Laboratories (Disco 2012). In this study we applied Disco Fuzzy Mining algorithm with respect to “absolute frequency” and “duration of activities” in order to discover distinguished patterns of behavior and interaction (as well as time intervals) between high and low achieving groups.

*Fuzzy Miner* This is one of the process mining algorithms. As we mentioned earlier, currently there are two popular types of fuzzy mining approaches: Fuzzy Mining by Disco Fluxicon and Fuzzy Mining by ProM (Fuzzy Miner 2009; Günther and Aalst 2007; ProM Tips 2010).



*Social Network Analysis (SNA)* This is a process mining technique used for investigating social structures with respect to networks and graph theories. In this study, we used SNA technique in order to study the handover of work/task (or behavior of collaborative interactions) between high and low achieving groups (Evelien and Ronald 2002).

*Social Network Miner* This is a process mining plugin that generates social networks from a process log (Social Network Miner 2012).

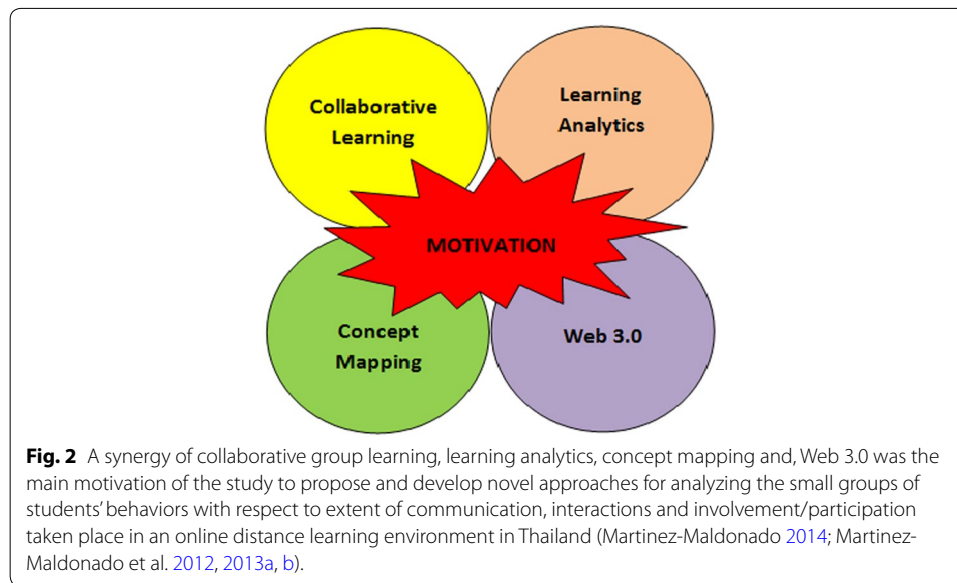
*Decision Point Analysis<sup>e</sup>* This is one of the ProM process mining plugin techniques used for decision/rule mining based on event logs (Process Mining Group 2009; Rozinat and Aalst 2006).

*MXML* This is a standard XML-based format supported in ProM 5.2 and ProM 6.4 frameworks (Aalst 2009).

### **Introduction and motivation of the study**

Collaborative group learning is an outcome of communication, interaction, participation and involvement. Based on the learning design and education settings, students might interact with instructors and trainers, with content, materials and/or with other classmates in the classroom. Many instructors spend considerable amounts of time and effort to developing their teaching style in such a way to increase the level of participation and interactions amongst students during the assigned activities and assignments (Elias 2011). In recent times, an increasing level of awareness and interest toward the way educational data can be applied to improve the quality of learning and teaching has led to extraordinary growth of an almost new field of study called “learning analytics” (Elias 2011). State-of-the-art analytics tools as well as new technologies make possible the statistical analysis of datasets (collected from learning situations) as well as the discovery of patterns and models within the event logs. These patterns and models can be used to enhance the prediction of future events in the learning by increasing the awareness of the instructors towards the interaction behaviors of the students during the group activities (Seven Things you should know about analytics 2010). On the other hand, concept mapping is a technique that can assist learners to construct and build visual demonstrations of the structure of their knowledge, information and comprehension about almost any topic or subject, founded on meaningful learning (Novak 1990). Concept maps are a good method of developing logical reasoning, methodical thinking and learning abilities by finding cause and effect relationships and by enabling learners to see how thoughts and ideas can create a bigger and more complete whole (Concept Mapping Fuels 2008).

This paper linked the concepts of collaborative learning and learning analytics (i.e., educational data mining in this study) with a concept mapping activity using a Web 3.0 service provider. Therefore, as shown in Fig. 2, a synergy and intersection of collaborative group learning, learning analytics, concept mapping and Web 3.0 was the main motivation of the study to propose and develop novel approaches for analyzing the students’ behaviors with respect to collaborative communication, interactions and participation taken place in an online distance learning environment in Thailand. Furthermore, the works conducted by Martinez-Maldonado (2014), Martinez-Maldonado et al. (2013b), Wang et al. (Wang et al. 2014) and Östlund (Östlund 2008) highly motivated us to think carefully (and by following similar ways and approaches) about possibility of applying Process Mining techniques in Distance Learning situations in Thailand.



In the first part of the study, we aimed to identify the most significant factors that influence the performance of the small groups of students in distance learning situations in Thailand. To do this, we conducted a quantitative survey in 3 private universities of Bangkok. In the second part of the study, we applied several process mining techniques using ProM 6.4.1 and ProM 5.2 (as Open Source frameworks for all of the process mining algorithms) as well as Disco Fluxicon (as an Open Source framework for couple of the process mining algorithms) to extract knowledge from the event logs collected and captured during a distance learning activity (i.e., online concept mapping) via *flowchart.com* as a free online real-time multi-user collaborative concept map maker service (Web 3.0 Software Service 2014). Using Disco Fluxicon, the data sets were initially converted into the *MXML* (XML-based) process mining standard formats. Accordingly, the datasets were divided into two main sets: datasets of the groups with high performance and datasets of the groups with low performance. We also inspected statistical and process map details about the actions occurred in each group by providing an overview of information about the number of cases and events in the datasets, level of communication, level of interactions, level of involvements/participation, duration of time spent (i.e., active versus idle intervals of time), total number of active students and a tree-like model of rules and decisions embedded in the datasets. To do this, we applied Fuzzy Mining (Disco) and Fuzzy Mining (ProM) techniques in order to discover and compare process maps between the High and Low Performance groups during the online distance activity in a private university in Bangkok (Thailand). We also applied Social Network Analysis technique (i.e., in terms of Handover of Work/Task) in order to investigate the extent of interactions between peers in each distance group of students. Consequently, using Decision Point Analysis technique (ProM) and by help of Decision Tree/Rules technique, we could analyze the behavior of the High and Low Performance groups in a more sophisticated and timely manner. Consequently, we analyzed the semantic and textual contributions of students shared and written in the chat rooms as well.

### Problems of the study

Though the intersection of Collaborative Learning and Learning Analytics with Concept Mapping and Web 3.0 sounded interesting, yet the group work in computer-based collaborative learning environments needs to be carefully addressed and monitored by lecturers in order to ensure collaborative group progress. In reality, lecturers mostly care (and are aware of) the *final artifacts* constructed by groups instead of the *whole collaborative process* in a detailed manner (Martinez-Maldonado 2014; Martinez-Maldonado et al. 2013b). Lecturers usually have *inadequate time* and *insufficient resources* to manage and supervise all group activities of students with regard to qualitative (e.g., movements, verbal and audio communications, gestures, body language, students' feelings and mood, and so on) or quantitative (e.g., number of words typed in the chat rooms, number of questions asked from each other, number of interactions executed by students, number of active versus passive students, duration of inactive time intervals, and so on) details. On the other hand, the final objects built by groups provide *incomplete insight and information* about collaborative processes as well (Martinez-Maldonado 2014; Martinez-Maldonado et al. 2012, 2013a, b). Therefore, there is a crucial need to new techniques and approaches that can assist the lecturers to increase their awareness toward the students' collaboration process holistically. To reach the above-mentioned goals, firstly we investigated the main factors that affect the students' group performance in Thailand. Secondly, we defined specific keywords and alphabets in order to extract the appropriate type of information and knowledge from the collected event log. Thirdly, we applied statistical and process mining techniques (by using a quantitative type of research) to analyze and study students' collaborative behavior in an online concept map activity (through a distance learning course) launched and run in a private university in Thailand. Fourthly, we analyzed the semantic and textual contributions of students using a qualitative approach.

### Questions of the study

The intersection of the above-mentioned issues rose the below main *questions*:

1. What are the most important factors that affect the performance of small groups in distance learning situations in Thailand? (Porouhan and Premchaiswadi 2011; Premchaiswadi et al. 2012)
2. How can the distance learning collected dataset be analyzed and interpreted in order to increase the lecturer's awareness toward the collaborative activity process in the groups of students? (Martinez-Maldonado 2014; Martinez-Maldonado et al. 2013b)
3. After the end of the class, can the lecturer discover and compare *statistically relevant relationships* on the distance learning collected dataset between high performance groups and low performance groups based? (Martinez-Maldonado 2014; Martinez-Maldonado et al. 2013b)
4. After the end of the class, can the lecturer discover and compare distinguished *patterns (graphs/models) of interaction* based on the distance learning collected dataset between high and low performance groups? (Martinez-Maldonado 2014; Martinez-Maldonado et al. 2013b)

5. After the end of the class, can the lecturer compare and study the contributions made by the students (textually and semantically) between high and low performance groups while writing and typing in chat rooms during the distance learning activity? (Östlund 2008)

### Statement of the study

The main objectives of the research were to address the five *questions* of the study. By considering the five main questions, a *single statement* that can contain all the approaches of the study was defined as follows: “To identify the most important factors that affect the performance of groups in Thailand and to conduct an empirical study to analyze student’s collaborative behavior as well as textual and semantic contributions using process mining and statistical techniques in order to provide support to lecturers by increasing their awareness toward students’ collaboration process during online distance learning activities and assignments” (Martinez-Maldonado 2014; Martinez-Maldonado et al. 2012, 2013a, b).

### Related works

In a research conducted by Berea et al. (2015), a data collected from a local survey—about students achievement in college—was studied by data analytics methods. Their results showed that academic performance of a student in a college depends on many factors/patterns and variables. Therefore, in a diverse population of students and in a mixed set of colleges, not all of the environments can perfectly lead to success of a student in a college. In their proposed conceptual framework, two of the important factors that may affect the success of a student in a college were “personal traits” and “college environment”. They proposed an algorithm that predicts which “college environment” might be the best place for a college student. Moreover, using the same algorithm they could predict which “personal traits” of a college student may lead to success in study and learning environments.

Jishan et al. (2015) presented a decision analytics approach in which how data can be preprocessed using Optimal Equal Width Binning (i.e., which is a discretization method) and Synthetic Minority Over-Sampling (i.e., which is an over-sampling method) techniques. In their work, they used a dataset collected from a course offered at a university in Bangladesh. Their main goal was to create a model that can accurately predict the students’ final grades. Using a Decision Tree approach, they could build a model that could indicate and predict when the discretization and over-sampling methods should be applied. In order to improve the accuracy of their model, they applied Neural Network and Naive Bayes classifier models as well. Eventually, the Naive Bayes classifier could 14% increase the accuracy of final grades predictions in such a way that significantly could reduce the level of misclassification error.

In another research conducted by Agarwal et al. (2012), they initially collected a database including the data about students of a college and they secondly applied several classification methods based on the dataset. Their results showed that Support Vector Machines (SVM) technique could result in the maximum amount of accuracy with minimum error. Moreover, they proposed a Decision Tree and Decision Rule Mining approach which can be a useful basis for admission of the best novice students (with

respect academic performance and subsequent requirements) for any course and any program. The Decision Tree in their study initially uses many parameters and can predict which might be considered as the most important one during the decision making process for selecting new students to the college.

In a research done by Al-Barrak and Al-Razgan (2015), they used a Decision Tree/Rules analysis and Classification Rules based on the J48 algorithm for predicting final GPAs of students for the fourth semester of an academic curriculum in at a university. Their proposed decision analytics approach was based on the grades of students' courses in previous semesters of the study. Accordingly, they could predict that the most important courses in the study plan that have a substantial influence on the students' final GPA.

Martinez-Maldonado et al. (2013b) proposed and developed a new and interesting approach to analyze collaborative traces of students on a concept map building activity in an authentic classroom and by means of several multi-user and multi-touch tabletops. Their research contained technological infrastructure as well as empirical results with respect to two educational data mining techniques (i.e., sequential pattern mining and process mining). Their main goal was to study and investigate the actions that differentiate high achieving groups from low achieving groups of students. To be able to analyze the collected interaction datasets, they defined three alphabets with specific keywords related to each category of alphabet. Their results showed that the keywords *Parallel* (i.e., students executed an action together and simultaneously) and *Other* (i.e., an action was executed by another student alternatively) appeared much more in top-3 frequent patterns of high achieving groups compared with low achieving groups. Similarly, the keyword *NoOwn* (i.e., a student executed on an object that previously was created by someone else) appeared two times more in high achieving groups. In our work we followed a similar approach and we compared our results with the findings of Martinez-Maldonado et al. (2013b). However, in their work only a Fuzzy Mining (Disco) technique (from process mining analysis tools) was applied on the dataset, and some other techniques such as Fuzzy Mining (ProM), Social Network Miner and Decision Point Analysis techniques were not applied on the collected data.

In another comprehensive research (doctoral dissertation) conducted by Martinez-Maldonado (2014), several new approaches to analyze student's interactions data collected from interactive tabletops were discussed and elaborated. However, the main emphasis of the thesis was on technical infrastructure development and statistical analysis (or data mining methods) of the data. From process mining analysis tools, only Fuzzy Mining (Disco) technique was applied and some other techniques such as Fuzzy Mining (ProM), Social Network Miner and Decision Point Analysis techniques were not applied on the authentic datasets.

## **Methodology and conceptual framework**

### **Participants and tutorial design**

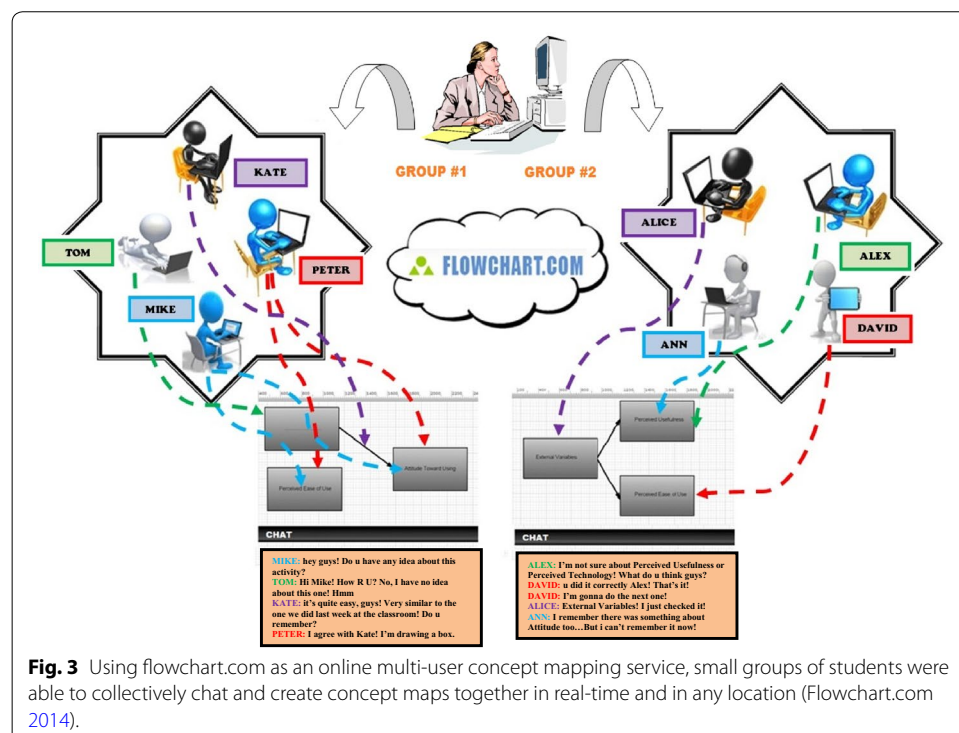
In general, two tutorial sessions were organized for the Bachelor of Business Administration (BBA) program of a private university in Bangkok (Thailand) during the 8th week of Semester 1, 2015. A total of 120 students aged between 20 and 23 attended the tutorial sessions (via distance learning) designed for the course: BPS207: Marketing Psychology.



Each tutorial session included 2 activities set up in English language. In this study, we only focused on the second activity, as the first activity was given just for warm up and practice only. In total, 57 of the total 120 students (47.5%) were female while 63 of the total 120 students (52.5%) were male. Altogether, 68% of the participants were native English speakers and the rest of them were native Thai. Each tutorial session was organized in groups of 4 students, which means 30 groups with 4 members in total. The lecturer chose a *concept mapping* activity so as to elaborate the topic of the “*Technology Acceptance Model (TAM)*” in proportion to the eighth week of the study in Marketing Psychology (via a distance learning activity). Using *flowchart.com* as a free online real-time multi-user collaborative concept map maker service (Web 3.0 Software Service 2014), students were asked to build/construct a *Technology Acceptance Model (TAM)* concept map regarding to how users come to accept and use a new technology in their environment. The final artifact of the concept mapping activity (i.e., the master model or the correct concept map) needed to be a model consisted of 6 *Components* and 8 *Arrows* in total. Using the *flowchart.com* for the online distance learning, students were able to collaborate with their peer group members in real-time and by typing text in the private chat rooms (provided for each group) as well. As shown in Fig. 3, all of the collaborating groups of students were empowered to collectively chat and create concept maps at the same time (Web 3.0 Software Service 2014).

### Conceptual framework

Consistent with Question 1 of the study and in order to identify the most significant factors affecting the performance of the students in small groups, a quantitative survey (using online questionnaires) was conducted. The questionnaires were distributed to 86



Bachelor students of 3 private universities in different metropolitan areas of Bangkok (Thailand). In general, two types of questionnaires were provided; English Version and Thai Version. The first type of the questionnaires (i.e., English Version) was distributed to English speaker students only while the latter type of questionnaires (i.e., Thai Version) was distributed to local Thai students. Overall, the questionnaires included 25 questions and a “*Likert Five Point Scale*” format was used as a basis of the structured questions (Porouhan and Premchaiswadi 2011). After reviewing the secondary data related to the “Theories of Groups” (Mcgrath 1991), “Theories of Groups Performance and Interaction” (McGrath 1984) and after careful investigation of the works written by Martinez-Maldonado (2014) and Martinez-Maldonado et al. (2013b); eight independent variables and one dependent variable were selected and defined for the *Conceptual Framework* of the survey as the following:

- *Facilitating conditions (independent variable)* is defined as the degree to which a student receives support (i.e., mentally or technically) during the distance learning activity via *flowchart.com*.
- *Level of involvement (independent variable)* is defined as the degree to which a student participates in the distance learning activity (via *flowchart.com*) by creating objects or performing activities (or actions).
- *Level of interaction (independent variable)* is defined as the degree to which a student works with a concept map object created by another fellow group member during the distance learning activity via *flowchart.com*.
- *Level of communication (independent variable)* is defined as the degree to which a student writes (or types) a text in the *Chat Box* in order to communicate with (a) fellow group member(s) in real-time during the distance learning activity via *flowchart.com*.
- *Degree of difficulty (independent variable)* is defined as the degree to which a student who perceives the concept map activity (during the distance learning activity via *flowchart.com*) is hard and difficult.
- *Group size (independent variable)* is defined as the total number of individuals in small groups during the distance learning activity via *flowchart.com*.
- *Prior experience (independent variable)* is defined as a student’s past participation in concept map creation activities through a distance learning course.
- *Gender (independent variable)* refers to the sexual identity of the students (during the distance learning activity via *flowchart.com*) in terms of male or female.
- *Performance of group (dependent variable)* is defined as the degree to which a final artifact created by small groups of students (during the distance learning activity via *flowchart.com*) is correct and compatible with the lecturer’s master concept map.

Table 1 shows the reliability analysis of the proposed conceptual framework questions (including eight independent variables and one dependent variable) with regard to the Cronbach’s Alpha ( $\alpha$ ) which is commonly used as a measure for reliability analysis of data. By contemplating on the reliability results of Table 1, we can see a fair reliability (almost high) for every component of the proposed conceptual framework.

**Table 1 Results of reliability analysis for components of the initial conceptual framework (eight independent and one dependent variables) based on the Cronbach's ( $\alpha$ )**

Components	Cronbach's alpha	No. of items
Facilitating conditions (independent variable)	0.953	5
Level of interaction (independent variable)	0.883	3
Degree of difficulty (independent variable)	0.812	4
Group size (independent variable)	0.831	3
Level of involvement (independent variable)	0.905	4
Prior experience (independent variable)	0.736	4
Gender (independent variable)	0.814	4
Level of communication (independent variable)	0.790	6
Performance of group (dependent variable)	0.811	2
Total (all independent and dependent variables)	0.783	35

**Table 2 Results of the Pearson correlation (2-tailed) analysis for seven independent variables and one dependent variable**

Components	Cronbach's alpha	No. of items
Performance_of_group_Y	pearson correlation Sig. (2-tailed)	1
Level_of_communication_X1	Pearson correlation Sig. (2-tailed)	0.475** 0.000
Level_of_interaction_X2	Pearson correlation Sig. (2-tailed)	0.291** 0.007
Level_of_involvement_X3	Pearson correlation Sig. (2-tailed)	0.368** 0.000
Degree_of_difficulty_X4	Pearson correlation Sig. (2-tailed)	0.442** 0.000
Prior_experience_X5	Pearson correlation Sig. (2-tailed)	0.460** 0.000
Facilitating_conditions_X6	Pearson correlation Sig. (2-tailed)	0.231* 0.032
Group_size_X7	Pearson correlation Sig. (2-tailed)	0.211 0.052
Gender_X8	Pearson correlation Sig. (2-tailed)	-0.076 0.486

\* Correlation is significant at the 0.05 level (2-tailed).

\*\* Correlation is significant at the 0.01 level (2-tailed).

Most importantly, the total reliability of the conducted survey was 78.3% which is quite acceptable for this survey.

In order to measure the linear correlation (i.e., the level of dependency) between variables, we applied the Pearson product-moment correlation coefficient (simply called as Pearson coefficient). Table 2 illustrates the correlation between 8 independent variables and 1 dependent variable (performance of group) in more detail. Considering the results of *Pearson Correlation (2-tailed)*, two components of "Group Size" and "Gender" were eliminated from the initial conceptual framework (i.e., their significance level was not adequate). Therefore the number of the independent variables was reduced from 8 to

**Table 3 Results of the ridge regression analysis for six independent variables and one dependent variable**

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-20.106	1.933		-10.401	0.000
	Level_of_communication_X1	1.896	0.284	0.445	6.668	0.000
	Level_of_interaction_X2	1.209	0.221	0.378	5.463	0.000
	Level_of_involvement_X3	1.097	0.269	0.340	4.085	0.000
	Degree_of_difficulty_X4	1.079	0.265	0.323	4.066	0.000
	Prior_experience_X5	0.826	0.224	0.249	3.686	0.000
	Facilitating_conditions_X6	0.157	0.335	0.042	0.468	0.641

6 independent variables in total and now we have a new conceptual framework with 6 independent variables and 1 dependent variable.

Based on the “*Ridge Regression Analysis*” (shown in Table 3), we realized that the component “Facilitating Conditions” was not supported by the results of the hypothesis testing as well (i.e., because the *t value* of 0.468 is less than 2.0 and the *significance level* of 0.641 is not less than 0.05). Therefore, the most significant factors/variables affecting the performance of groups in Thailand were listed/sorted down in sequence (and from top to down) as the following: (1) level of communication (significance level = 0.000 < 0.05 and *t value* = 6.668 > 2.0), (2) level of interaction (significance level = 0.000 < 0.05 and *t value* = 5.463 > 2.0), (3) level of Involvement (significance level = 0.000 < 0.05 and *t value* = 4.085 > 2.0), (4) degree of difficulty (significance level = 0.000 < 0.05 and *t value* = 4.066 > 2.0), and (5) prior experience (significance level = 0.000 < 0.05 and *t value* = 3.686 > 2.0).

#### **Level of communication**

In order to measure the level of communication, we decided to analyze the contributions made by the students in the chat environment (*Chat Box*) of the *flowchart.com* based on the observation and based on the following qualitative metrics: (1) Total number of the *words* typed during the activity, (2) average number of the *words* typed (per group), (3) total number of the *questions* asked during the activity, (4) average number of *questions* asked during the activity (per group), (5) total number of the written sentences addressed to the *whole group*, (6) total number of the written sentences addressed to a *specific person*, (7) total number of the *encouragements* (yeah, good-job, well-done, etc.) used in the sentences, (8) total number of the *acknowledgements* (such as that’s right, correct, etc.) in the sentences, (9) total number of the *uncertainty expressions* (due to lack of experience) used in the sentences, (10) total number of *uncertain statements* (such as perhaps, maybe, not sure, etc.) used in the sentences, and (11) total number of *certain statements* (such as It is, I believe, I’m sure, etc.) used in the sentences (Östlund 2008).

#### **Data description and definition of keywords and alphabets**

The preliminary raw data (viewed by the lecturer or administrator) in *flowchart.com* was consisted of a lengthy sequence of actions labeled as: {ActionType, Construct, Subject, TimeStamp}, where (1) *ActionType* can be: a *Create* (i.e., build/create a Rectangle, Simple

Line or a Text Object), a *Delete* (i.e., remove/delete a Rectangle, Simple Line or a Text Object), a *Move* (i.e., move/shift a Rectangle, Simple Line or a Text Object), an *Edit* (i.e., add/edit a Text Object in a component or an arrow), a *Scroll* (i.e., scroll up or down the list of suggested components through the menu window), an *Open* (i.e., start the menu window in flowchart.com), or a *Close* (i.e., finish the menu window in flowchart.com). (2) *Construct* can be: a *Rectangle* (component), a *Simple Line* (arrow), a *Text Object* or a *Menu* (window). (3) *Subject* is the student who executes the action (such as Me, Mike, Jack, Kate, etc.) (4) *TimeStamp* is the time when the action takes place (such as 10:12:07 or 23:14:15) (Martinez-Maldonado 2014; Martinez-Maldonado et al. 2013b).

One example of the *lecturer/administrator's view* in flowchart.com includes: {"Create", "Simple Line", "Mike", "18:02:00"}, when Mike creates (or adds) a Simple Line (or an arrow) at 18:02:00 o'clock. Another example includes {"Move", "Rectangle 2", "Jennifer", "18:04:12"}, when Jennifer moves (or shifts) the second Rectangle (or Component) at 18:04:12 o'clock. Similarly, the sequence {"Open", "Menu", "Ted", "15:16:02"} is shown when Ted opens the menu window at 15:16:02 o'clock (Martinez-Maldonado 2014; Martinez-Maldonado et al. 2013b).

#### **Level of interaction**

In order to measure the level of interaction, and in order to follow a similar approach previously developed by Martinez-Maldonado (2014) and Martinez-Maldonado et al. (2013b); we decided to identify the *ownership* of the actions with respect to: (1) the actions that students perform on the objects constructed by themselves (keyword: *Possess*), or (2) the actions that students perform on the objects constructed by their other fellow group members (see alphabet 1 in Table 4).

#### **Level of involvement**

In order to measure the level of involvement, and in order to follow a similar approach previously developed by Martinez-Maldonado (2014) and Martinez-Maldonado et al. (2013b); we decided to identify the *sequence or order* of the actions executed by the students during the distance learning activity (i.e., online concept mapping) via *flowchart.com*. As shown in Alphabet 2 in Table 4, the students' actions (or activities) can take place: (1) simultaneously (or at the same time) with other students' actions (keyword: *Simultaneous*), (2) in-turns when the prior action is performed by another fellow group member (keyword: *Another*), or (3) as a series of actions performed by the same student in a row (keyword: *Same*).

**Table 4 Specific alphabets and keywords defined for better analysis and interpretation of the distance learning collected event log**

Object	Action type		Alphabet 1	Alphabet 2
Component-C	Add-C,A	Del-C,A	Possess	Simultaneous
Arrow-A	Edit-C,A	Open-M	NoPossess	Another
Menu-M	Shift-C,A,M	Close-M		Same
Inactivity Interval-Idle	Short Inactivity-IdleShort	Long Inactivity-IdleLong		

Sources: (Section 8.7.2, Page 179, Martinez-Maldonado 2014; Section 4, Martinez-Maldonado et al. 2013b).

### ***Categorization of time intervals***

In order to further investigate the students' actions with respect to the intervals of time and similar to the previous studies conducted by Martinez-Maldonado (2014) and Martinez-Maldonado et al. (2013b); we decided to define 2 new categories of intervals of time as follows: (1) *IdleShort* (i.e., refers to the time when the gap between two actions executed during the distance concept mapping activity via *flowchart.com* is between 30 to 60 s) and (2) *IdleLong* (i.e., refers to the time when the gap between two actions executed during the distance concept mapping activity via *flowchart.com* is greater than 60 s).

### ***Categorization of actions***

In order to further analyze the students' actions with respect to the level of *influence* or *impact* on the concept mapping assignment, and in order to follow a similar approach developed by Martinez-Maldonado (2014) and Martinez-Maldonado et al. (2013b); we decided to categorize the actions as follows: (1) *high-impact* actions (i.e., refers to those types of actions that can substantially change the content or structure of the online concept map, such as: *Add* a component or arrow, *Delete* a component or arrow, *Edit* text a component or arrow), (2) *low-impact* actions (i.e., refers to those types of actions that only can change the layout or formation of the concept map, such as: *Shift* a component or arrow), and (3) *no-impact* actions (i.e., refers to those types of actions that have no influence on the contents or formation of the concept map, such as: opening and closing the main menu window, or scrolling up and down through the main menu in *flowchart.com* environment).

### ***Grouping of actions***

Subsequent to categorization of actions mentioned above and in order to follow a similar approach developed by Martinez-Maldonado (2014) and Martinez-Maldonado et al. (2013b); we decided to group the actions of students as follows: (1) *HighOnly* groups of actions are those types of actions consisted of *high-impact* action(s); (2) *LowOnly* types of actions are those types of actions consisted of *low-impact* action(s); and (3) *NoImpact* types of actions are those types of actions consisted of *no-impact* action(s).

### ***Degree of difficulty***

The difficulty level of the assigned online distance learning activity to build a concept map of *Technology Acceptance Model (TAM)* using *flowchart.com* was chosen in such a way to be neither too difficult nor too easy for all of 120 students in small groups.

### ***Categorization of groups of students***

In order to extract knowledge from the event logs collected and captured during the distance learning activity (i.e., *Technology Acceptance Model or TAM*) using *flowchart.com*, and in order to follow a similar approach developed by Martinez-Maldonado (2014) and Martinez-Maldonado et al. (2013b); we decided to divide the students into 2 main groups as follows: (1) *High Performance groups* (with greater or equal to 90% accuracy in creating the final artifacts), and (2) *Low Performance groups* (with less or equal to 89% accuracy in building the final artifact). We selected 90% accuracy as a threshold for high performance as this was an online concept map creation activity (rather than an essay/

**Table 5 Detailed demographic data about the high and low performance groups participating in distance concept mapping activity in Thailand**

Demographic data	High performance	Low performance
Number of groups	20	10
Number of students	80	40
Age	(20) 5% (21) 16% (22) 32% (23) 47%	(20) 12% (21) 8% (22) 17% (23) 63%
Gender	(Female) 46.25% (Male) 53.75%	(Female) 50% (Male) 50%
Profession	(Out of work) 82% (In Work) 18%	(Out of work) 79% (In Work) 21%
Prior computer experience (%)	100%	100%
Prior distance study experience >2 (%)	93%	76%
Online concept mapping experience (%)	82%	71%
Preferred learning style	(Individual) 3% (Group) 55% (Listen) 13% (Read) 3% (Write) 6% (Discuss) 20%	(Individual) 43% (Group) 12% (Listen) 14% (Read) 16% (Write) 7% (Discuss) 8%

written activity with strict course review structure and required discipline and parameters) and students were empowered by contacting each other during the activity run using *Chat Box*. Table 5 shows more demographic data about both of the High and Low groups (Östlund 2008).

## Results and findings

### Statistical results

As we mentioned earlier, the entire collected (and integrated) data were divided into two main sets of High Performance event logs and Low Performance event logs. Firstly, we compared the groups based on the *average time* taken to finish the assigned task. Out of maximum 30 min time for the Technology Acceptance Model (TAM) concept map

Events	545	Events	412
Cases	20	Cases	10
Activities	13	Activities	13
Median case duration	11.7 mins	Median case duration	18.8 mins
Mean case duration	12 mins	Mean case duration	18.6 mins

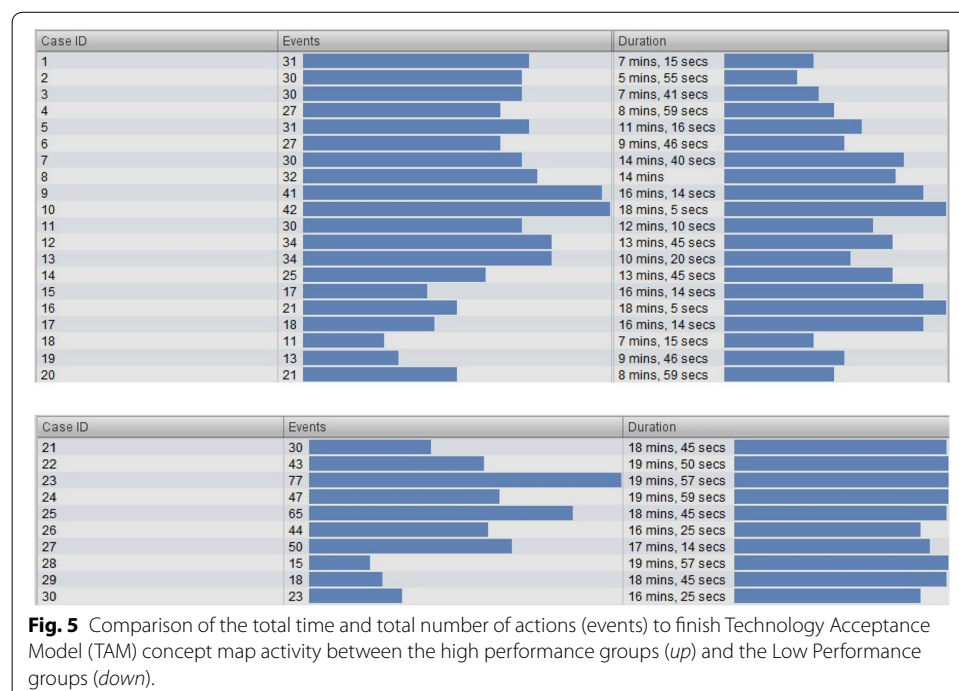
**Fig. 4** Comparison of the median and mean (average) time to accomplish the assigned distance activity between the high and low performance groups.

activity, it took 12 min *in average* for the High Performance groups to finish the concept map creation task. However, for the Low Performance groups, the *average time* spent to finish the same task was 18.6 min. Therefore, as shown in Fig. 4, none of the groups consumed whole of the 30 min allowed time to accomplish the Technology Acceptance Model (TAM) concept mapping task, though, Low Performance groups spent more time to finish the TAM concept map activity.

Secondly, investigating the details of the *total time* and *total number of actions*, we realized that in the High Performance groups (i.e., 20 groups), the maximum *duration of time* spent to finish Technology Acceptance Model (TAM) concept map activity was 18 min and 5 s in Group #10 whereas the minimum *duration of time* consumed to finish the same tasks was 5 min and 55 s in Group #2. On the other hand, as shown in Fig. 5 (up), the maximum and minimum *numbers of students' total actions* (called *events*) were 42 (Group #10) and 11 (Group #18), respectively, in the High Performance groups.

By the same token, as shown in Fig. 5 (down), the maximum and minimum *duration of time* to finish Technology Acceptance Model (TAM) concept map activity in the Low Performance groups (i.e., 10 groups) were 19 min and 57 s (in Groups #23 and #29) and 16 min and 25 s (in Groups #26 and 30), respectively. The maximum *number of students' total actions* during the Technology Acceptance Model (TAM) concept map activity was 77 actions (or *events*) in Group #23 whereas the *minimum number of students' total actions* was 15 in Group #29 of the Low Performance groups.

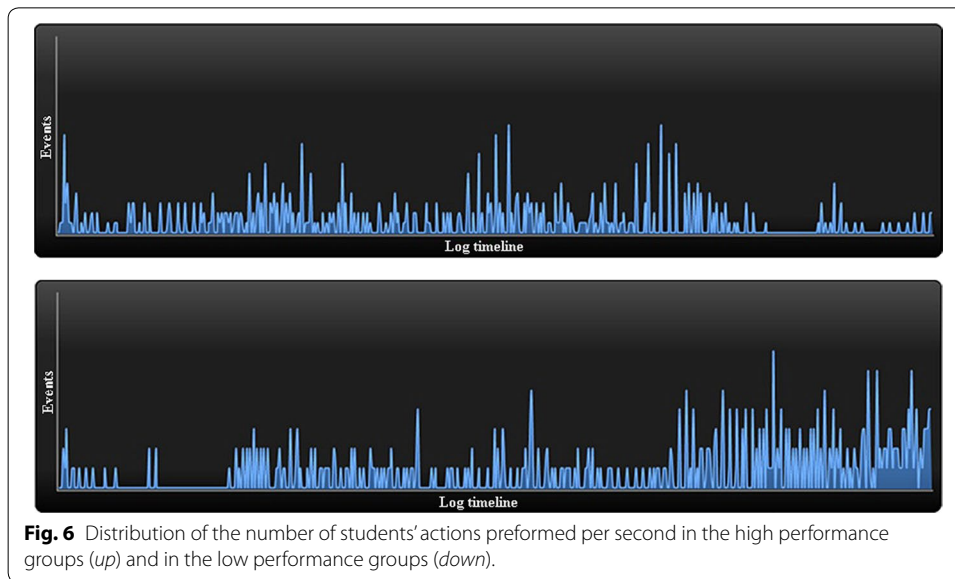
In addition, as shown in Table 6, the *average number of actions (events)* executed in the High Performance groups was 27.25 (i.e., 545 total number of events divided by 20 total number of High Performance groups is equal with 27.25) actions whereas the *average number of actions (events)* executed in the Low Performance groups was 41.2 (i.e., 412 total number of events divided by 10 total number of Low Performance groups is equal





**Table 6 Comparison of the median and mean frequency of actions (events) to accomplish the assigned distance activity between the high and low performance groups**

	Median frequency	Mean frequency
High performance groups	22	27.25
Low performance groups	22.89	41.2

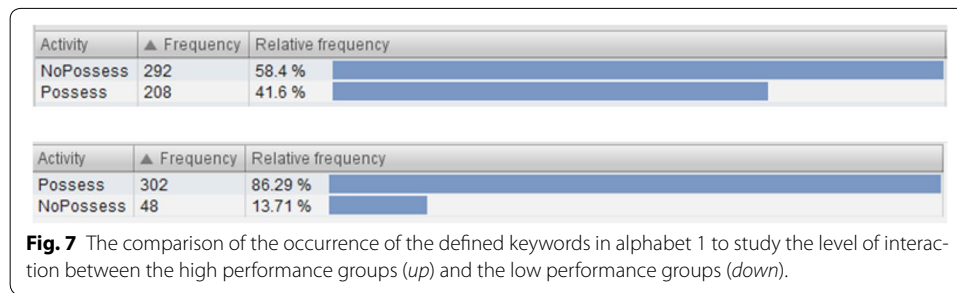


with 41.2). This means that the students in the Low Performance groups performed more actions (in average) and created more events (in average).

Thirdly, we illustrated the *number of students' actions* performed over the *time*. The Y coordinate in Fig. 6 represents the *frequency (number of actions)* while the X coordinate illustrates the *time* of the tutorial session in Technology Acceptance Model (TAM) concept map activity. The distribution diagram of the High Performance groups significantly demonstrates a very high *ratio* of actions performed per second in the middle of the tutorial sessions. On the contrary, the distribution diagram of the Low Performance groups significantly demonstrates a very high *ratio* of actions performed per second a moment just *before the end* of the Technology Acceptance Model (TAM) concept map activity. Furthermore, in the High Performance groups, the *maximum number of the actions per second (ratio)* occurred at 08:11:11 o'clock and 10:12:03 o'clock (with 5.11 and 5.09 actions per second) while in the Low Performance groups, the *maximum number of the actions per second (ratio)* occurred at 12:07:12 o'clock (with 8.6 actions per second).

#### Level of interaction results

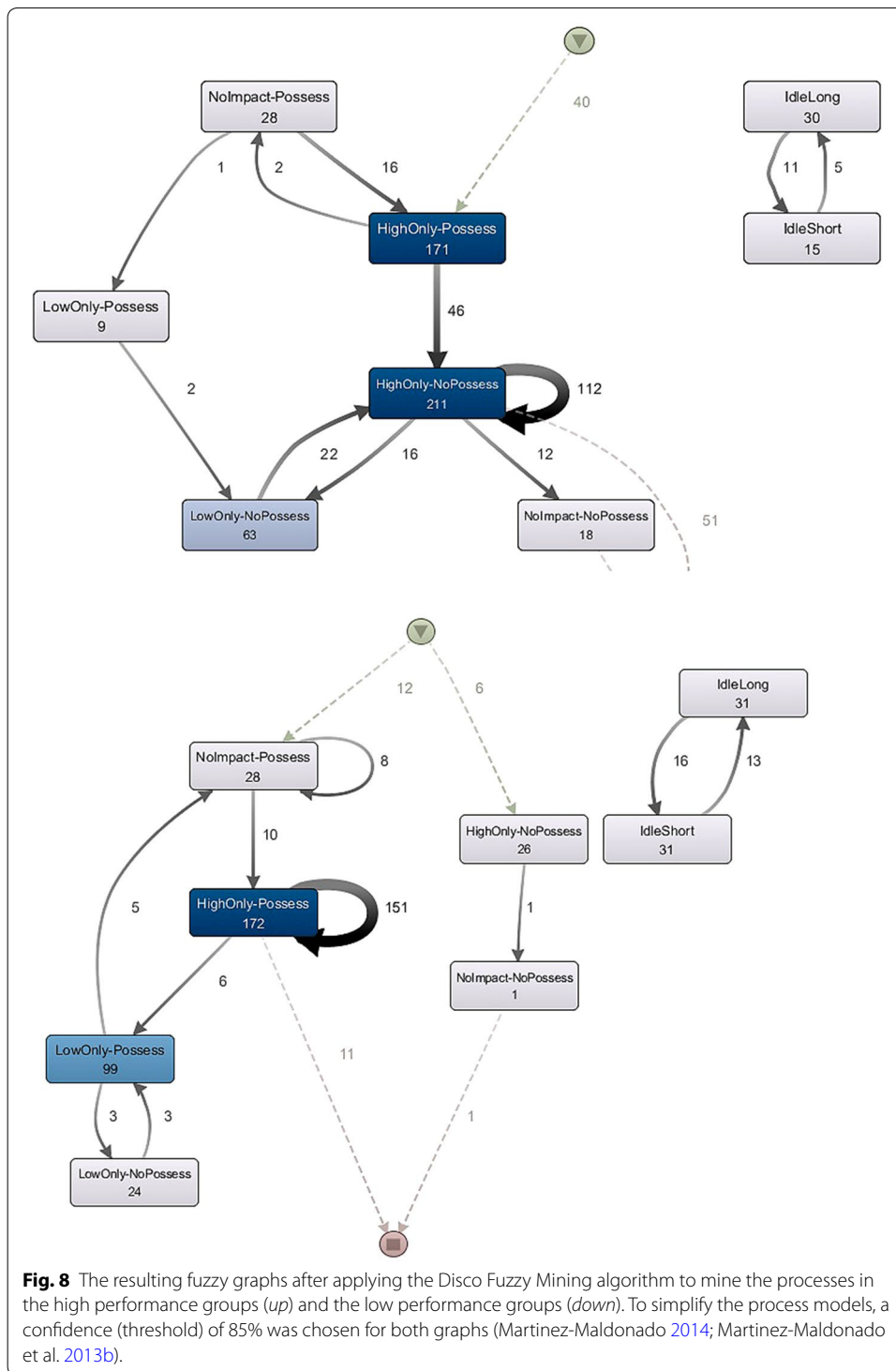
Firstly and after careful investigation of the High and Low Performance groups (using Statistics Overview of Disco Fluxicon) in online distance learning activity using *flow-chart.com*, we realized that the keywords *Possess* and *NoPossess* appeared in 41.06 and 58.04% of the whole dataset in the High Performance groups, while the same keywords appeared in 86.29 and 13.71% of the whole dataset in the Low Performance groups,



respectively (see Fig. 7). Therefore, the occurrence of the keyword *NoPossess* was 4 times greater in the High Performance groups compared with the Low Performance groups. Consequently, the level of *students' interaction* was almost 4 times (i.e., 58.04 divided by 13.71 is equal with 4) greater in the High Performance groups. These results were consistent with the findings of Martinez-Maldonado (2014) and Martinez-Maldonado et al. (2013b) as in their works the occurrence of the keyword *NoOwn* (i.e., in our study we changed the same keyword to *NoPossess*) was two times greater in the top-4 frequent sequences (not whole the dataset) of the high achieving groups compared with the low achieving groups.

Secondly, in order to further investigate the level of interaction between small groups of students, we used *Disco Fuzzy Mining algorithm* in order to mine the interaction processes in both of the High and Low Performance groups. By visually comparing the graphs we highlighted that they both share identical *core blocks of activity*. This was not compatible with the findings of Martinez-Maldonado (2014) and Martinez-Maldonado et al. (2013b) as in their works some blocks of activities in the resulting fuzzy models (highlighted with a yellow star) appeared in the high achieving groups but not in the low achieving groups.

By contemplating on the resulting fuzzy graphs (Disco) for both of the High and Low Performance groups in our study (see Fig. 8), we realized that the block named "*HighOnly-NoPossess*" was the *most significant* trace of the interaction (with absolute frequency of 211 times) in the High Performance groups. This was not compatible with the findings of Martinez-Maldonado (2014) and Martinez-Maldonado et al. (2013b) as in their works the block named "*HighLow-NoOwner*" allocated the highest level of significance to itself. Quite the opposite, the block named "*HighOnly-Possess*" in our study was the *most significant* trace of the interaction (with absolute frequency of 172 times) in the Low Performance groups. This also was not compatible with the findings of Martinez-Maldonado (2014) and Martinez-Maldonado et al. (2013b) as in their works the block "*HighLow-NoOwner*" (similar to high achieving groups) allocated the highest level of significance to itself. Therefore, based on the findings of our study, we can conclude that students in the Low Performance groups showed more tendencies to execute *high-impact actions* on the objects created and possessed by themselves while students in the High Performance groups showed more tendencies to execute *high-impact actions* (such as adding, deleting, or editing text in a component or arrow) on the objects previously created and possessed by their other fellow group members.



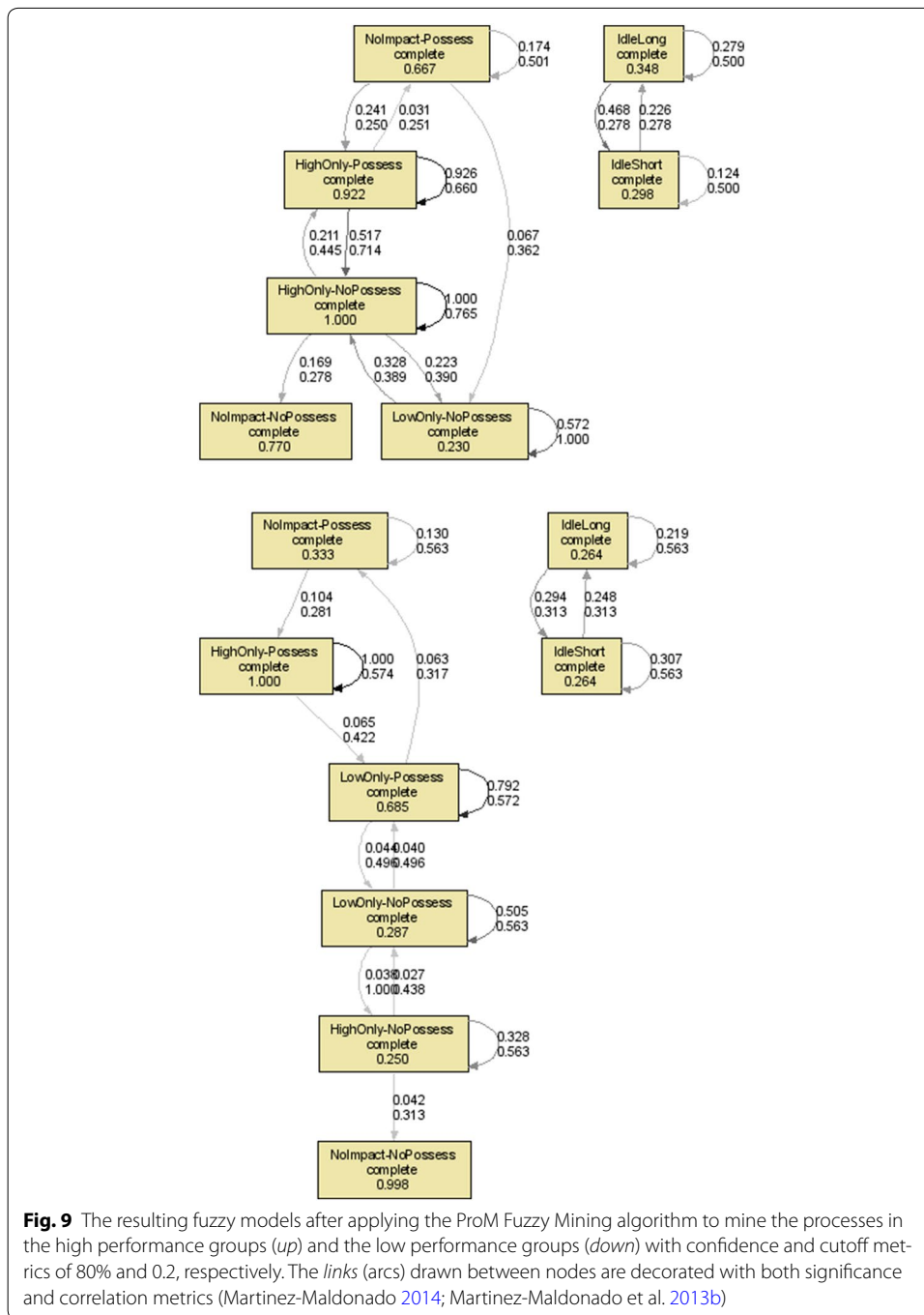
**Fig. 8** The resulting fuzzy graphs after applying the Disco Fuzzy Mining algorithm to mine the processes in the high performance groups (*up*) and the low performance groups (*down*). To simplify the process models, a confidence (threshold) of 85% was chosen for both graphs (Martinez-Maldonado 2014; Martinez-Maldonado et al. 2013b).

In addition, the Low Performance groups in average executed more blocks of actions with *no-impact* such as opening or shifting the main menu window in *flowchart.com* free online concept making environment (see Fig. 8). The average frequencies of the “*NoImpact-Possess*” and “*NoImpact-NoPossess*” blocks were 4.6 in High Performance groups

(i.e., 46 divided by 20 is equal with 2.3) versus 3 in Low Performance groups (i.e., 30 divided by 10 is equal with 3). Therefore, the average frequencies of the “*NoImpact-Possess*” and “*NoImpact-NoPossess*” blocks were 1.3 times more in the Low Performance groups compared with the High Performance groups. This was not compatible with the findings of Martinez-Maldonado (2014) and Martinez-Maldonado et al. (2013b) as in their works the blocks “*NoImpact-NoOwner*” and “*NoImpact-Owner*” only appeared in the resulting fuzzy model of the high achieving groups, while the resulting fuzzy model of the low achieving groups did not contain the blocks of “*NoImpact-NoOwner*” and “*NoImpact-Owner*”. However, we need to consider an important issue that in the studies conducted by Martinez-Maldonado (2014) and Martinez-Maldonado et al. (2013b); a “*High Only*” block was defined and introduced as a mixture of both *high impact actions* and *no impact actions*; and a “*LowOnly*” block was defined and introduced as a mixture of both *low impact actions* and *no impact actions*; and a “*NoImpact*” block was defined and introduced as a block which only includes *no impact actions*.

In the same way, in our study the Low Performance groups on average (i.e., divided by 10), performed more blocks of actions with *low-impact* (i.e., 123 divided by 10 is equal with 12.3) compared with the High Performance groups (i.e., 72 divided by 20 is equal with 3.6). The average frequencies of the blocks including “*LowOnly-Possess*” and “*LowOnly-NoPossess*” were almost 3.5 times more in the Low Performance groups (see Fig. 8). Compatible with the findings of Martinez-Maldonado (2014) and Martinez-Maldonado et al. (2013b), the occurrence of the block “*LowOnly-NoPossess*” (i.e., in our study we changed the same keyword to *LowOnly-NoPossess*) was greater in the *low achieving groups* (i.e., in our study we changed it to the *Low Performance groups*). Nevertheless, based on the findings of Martinez-Maldonado (2014) and Martinez-Maldonado et al. (2013b) in their works, the block “*LowOnly-Owner*” (i.e., in our study we changed the same keyword to *LowOnly-Possess*) appeared only in the low achieving groups (but in our study the block “*LowOnly-Possess*” appeared in both of the High and Low Performance groups).

Thirdly, similar to *Fuzzy Mining (Disco)*, we also applied *Fuzzy Mining algorithm (ProM)* in order to mine the interaction processes of both of the High and Low Performance groups. Quite different with Fuzzy Mining in disco, the ProM Fuzzy Miner deals with two fundamental metrics: (1) Significance and (2) Correlation. “Significance” deals with the relative *importance of behavior* while “Correlation” deals with *how closely related* two events following one another are (Günther and Aalst 2007). Figure 9 shows the resulting fuzzy models (ProM) in order to mine the interaction processes in both of the High and Low Performance groups with the overall *conformance* and *cutoff* metrics of 80% and 0.2, respectively. As illustrated, the *most significant blocks of activity* in the High Performance groups (with regard to the “significance” metric) were as follows: (1) *HighOnly-NoPossess* (with the highest significance of 1.000), (2) *HighOnly-NoPossess* (with significance of 0.922), (3) *NoImpact-NoPossess* (with significance of 0.770), (4) *NoImpact-Possess* (with significance of 0.667), (5) *IdleLong* (with significance of 0.348), (6) *IdleShort* (with significance of 0.298), and (7) *LowOnly-NoPossess* (with the lowest significance of 0.230). Therefore, similar to Disco fuzzy models, the resulting block of “*HighOnly-NoPossess*” was *the most significant behavior* in the High Performance groups. These were not compatible with the findings of Martinez-Maldonado (2014)



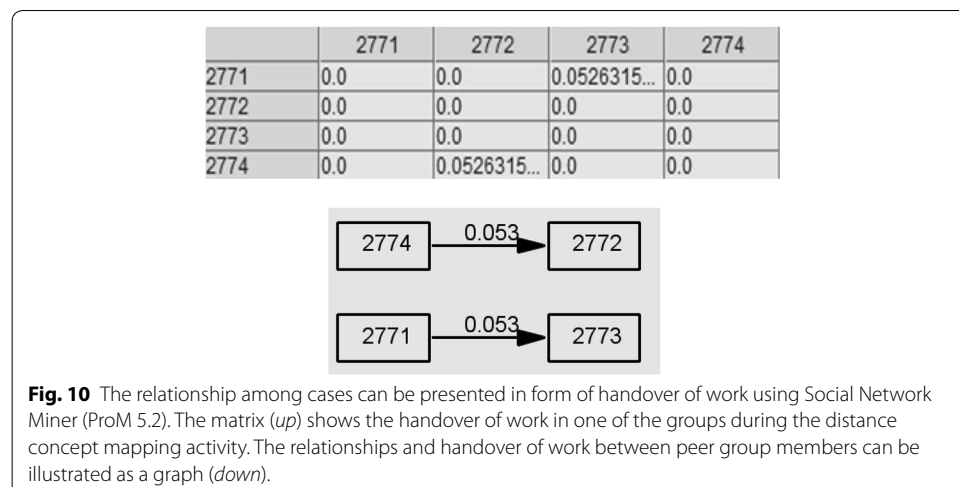
**Fig. 9** The resulting fuzzy models after applying the ProM Fuzzy Mining algorithm to mine the processes in the high performance groups (*up*) and the low performance groups (*down*) with confidence and cutoff metrics of 80% and 0.2, respectively. The *links* (arcs) drawn between nodes are decorated with both significance and correlation metrics (Martinez-Maldonado 2014; Martinez-Maldonado et al. 2013b)

and Martinez-Maldonado et al. (2013b), as in their works the *most significant blocks of activity* in the high achieving groups (with regard to the “significance” metric) were as follows: (1) *HighLow-NoOwner*, (2) *Inact-Short* (we changed the same keyword in our work to *IdleShort*), (3) *Inact-Long* (we changed the same keyword in our work to *Idle-Long*), (4) *LowOnly-NoOwner* (we changed the same keyword in our work to *LowOnly-NoPossess*), (5) *HighLow-Owner*, (6) *NoImpact-NoOwner* (we changed the same keyword

in our work to *NoImpact-NoPossess*), and (7) *NoImpact-Owner* (we changed the same keyword in our work to *NoImpact-Possess*).

In the same way, the *most significant blocks of activity* in the Low Performance groups with regard to the “significance” metric (see Fig. 9) were as follows: (1) *HighOnly-Possess* (with the highest significance of 1.000), (2) *NoImpact-NoPossess* (with the significance of 0.998), (3) *LowOnly-Possess* (with the significance of 0.685), (4) *NoImpact-Possess* (with the significance of 0.333), (5) *IdleShort* (with the significance of 0.264 and better correlation metrics of 0.307), and (6) *IdleLong* (with significance of 0.264 and less correlation metrics of 0.219). Therefore, the resulting blocks of “*HighOnly-Possess*” (with the highest significance of 1.000) and “*IdleLong*” (with the lowest significance of 0.264) were *the most and the least significant behaviors* in the Low Performance groups, respectively. These were not compatible with the findings of Martinez-Maldonado (2014) and Martinez-Maldonado et al. (2013b), as in their works the *most significant blocks of activity* in the low achieving groups (with regard to the “significance” metric) were as follows: (1) *High-Low-NoOwner*, (2) *Inact-Long* (or *IdleLong*), (3) *Inact-Short* (or *IdleShort*), (4) *HighOnly-NoOwner*, (5) *LowOnly-NoOwner* (or *LowOnly-NoPossess*), (6) *HighLow-Owner*, (7) *HighOnly-Owner* (we changed the same keyword in our work to *HighOnly-Possess*), and (8) *LowOnly-Owner* (we changed the same keyword in our work to *LowOnly-Possess*). However, as mentioned earlier, in their work (and a bit different with our approach in this study) a “*High Only*” block was defined as a combination of *high impact actions* and *no impact actions*; and a “*LowOnly*” block was defined as a combination of *low impact actions* and *no impact actions*; a “*NoImpact*” block only included *no impact actions*.

Fourthly, we used *Social Network Miner* technique to more investigate the students’ interaction processes with respect to *handover of work* (i.e., traces of interaction with others’ objects) during the distance learning activity, and during the Technology Acceptance Model (TAM) concept map assignment, between the High and Low Performance groups. The technique allowed us to visualize the handover of work from *Student A* to *Student B* if there are two subsequent activities where the first is completed by *Student A* and the second by *Student B*. To better understand the technique (see Fig. 10), the results of applying *Social Network Miner* on Group#13 as follows:



- *Student 2774* has executed at least one action on an object previously created by *Student 2772*.
- *Student 2771* has executed at least one action on an object previously created by *Student 2773*.
- *Student 2774* has never executed an action on an object previously created by *himself* (i.e., *Student 2774*).
- *Student 2772* has never executed an action on an object previously created by *himself* (i.e., *Student 2772*).
- *Student 2771* has never executed an action on an object previously created by *himself* (i.e., *Student 2771*).
- *Student 2773* has never executed an action on an object previously created by *himself* (i.e., *Student 2773*).
- *Student 2771* has never executed an action on an object previously created by *Student 2774* and vice versa.
- *Student 2771* has never executed an action on an object previously created by *Student 2772* and vice versa.
- *Student 2773* has never executed an action on an object previously created by *Student 2774* and vice versa.
- *Student 2773* has never executed an action on an object previously created by *Student 2772* and vice versa.

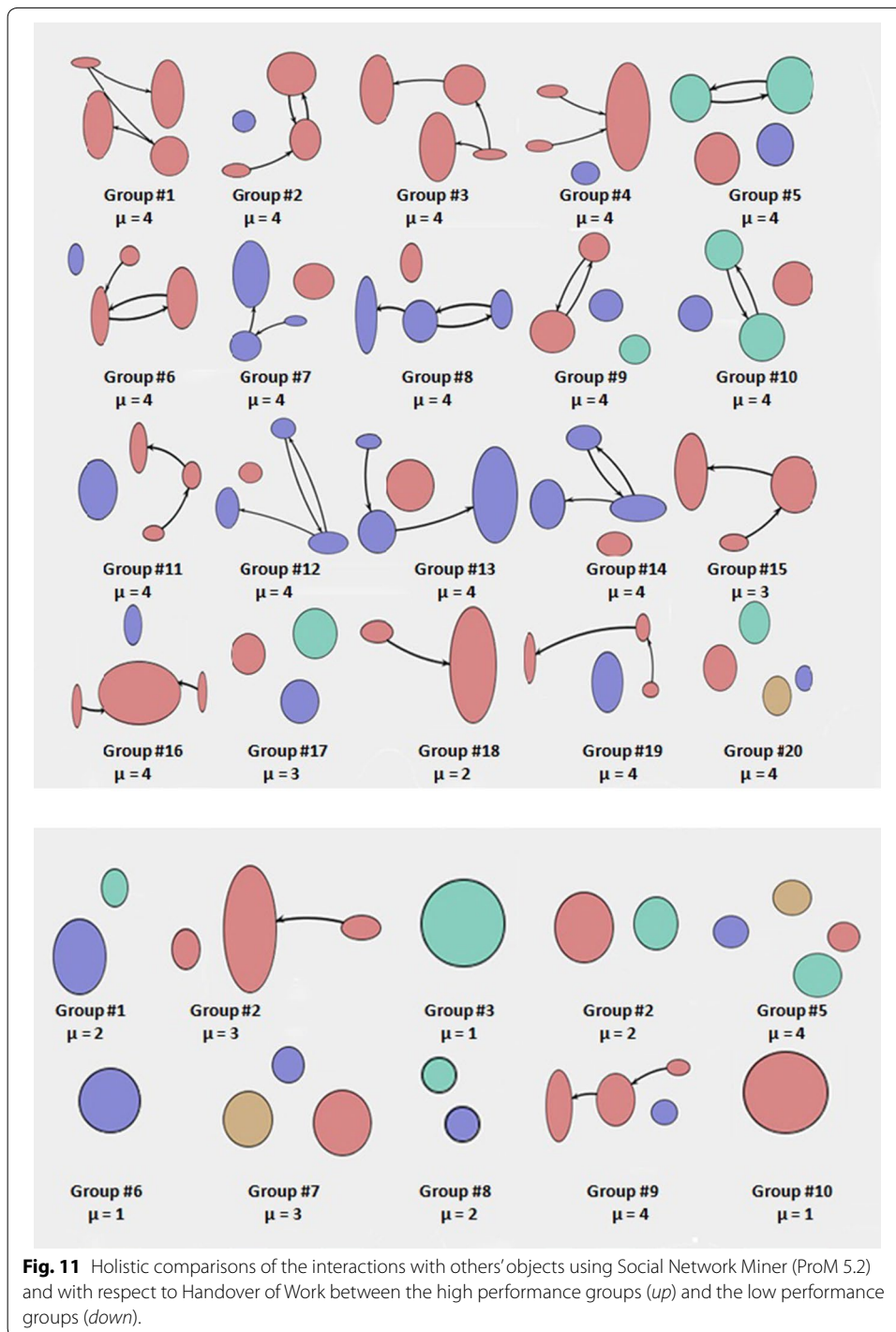
The matrix in Fig. 10 (up) simply shows the handover of work situation in Group #13. The main idea was to firstly count the number of times *Student 2774* has executed an activity on an object previously created by *Student 2772*, and secondly, divide the obtained number by the total number of handover of works taken place in Group #13. Finally, as shown in Fig. 10 (down), these relationships can be finally illustrated as a graph.

Based on the above-mentioned approach, a *holistic comparison* of the *handover of work* (i.e., interactions with others' objects) between the High and the Low Performance groups was illustrated in Fig. 11. More proportions of *interaction with others' objects* lead to illustration of more horizontal oval shapes. Quite the reverse, more proportions of *creating objects that others use* lead to illustration of more vertical oval shapes.

By comparing the Social Network Miner graphs shown in Fig. 11, we realized that the High Performance groups were obviously more involved in production of more collaborative processes (with higher level of interaction) expressing more complex handover of tasks from one student to another student.

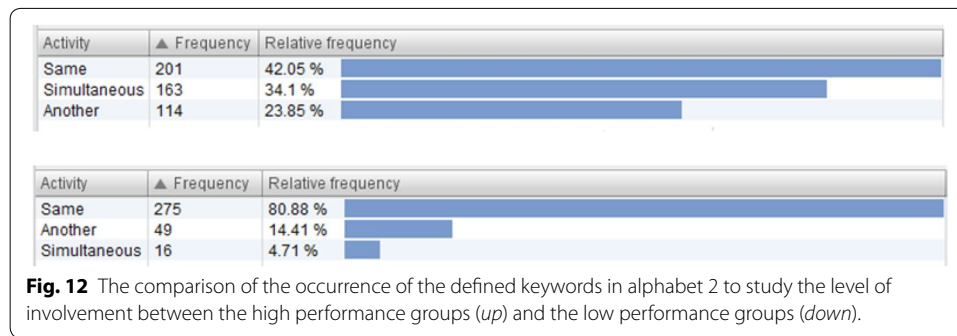
#### Level of involvement results

Firstly and after careful study of the both High and Low Performance groups (using Statistics Overview of Disco Fluxicon), we realized that the keywords *Simultaneous* and *Another* appeared in 34.1 and 23.85% of the whole dataset in the High Performance groups, while the same keywords appeared only in 4.71 and 14.41% of the whole dataset in the Low Performance groups, respectively (see Fig. 12). The occurrence of the keywords *Simultaneous* and *Another* was almost 3 times greater in the groups with high performance. Therefore, the level of *students' involvement* (with respect to the keywords



*Simultaneous* and *Another*) was almost 3 times greater—i.e., (34.1 + 23.85) divided by (4.71 + 14.41) was equal with 3.03—in the High Performance groups compared with the Low Performance groups. These results were consistent with the findings of Martinez-Maldonado (2014) and Martinez-Maldonado et al. (2013b) as in their works the occurrence of the keywords *Parallel* (i.e., in our study we changed the same keyword to

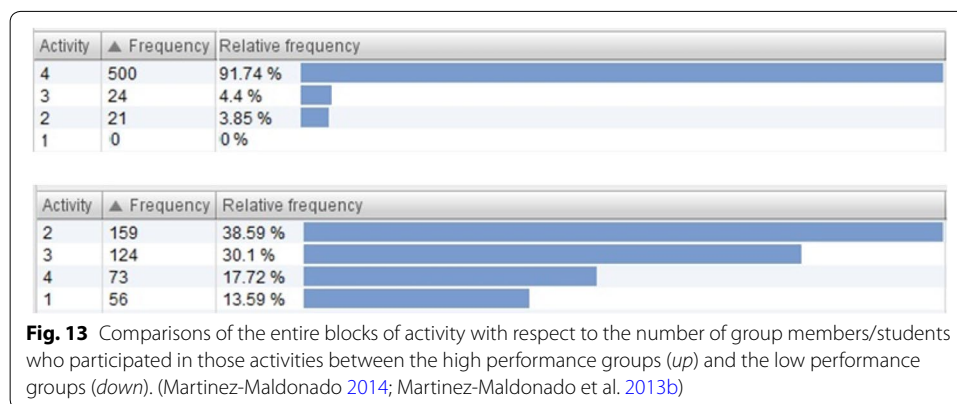




*Simultaneous*) and *Other* (i.e., in our study we changed the same keyword to *Another*) were almost 2.6 times greater in the top-4 frequent sequences (not whole the dataset) of the high achieving groups compared with the low achieving groups.

Secondly and by comparing the Social Network Miner graphs shown in Fig. 11 for a second time, we realized that out of the total 80 students (i.e., 20 high achieving groups); (1) 76 students *actively* engaged in the Technology Acceptance Model (TAM) concept map activity, while (2) 4 students did not engage in any activity (i.e., playing absolutely an idle role). Therefore, the total *participation rate* in the High Performance groups was 95%. On the other hand, out of the total 40 students (i.e., 10 low achieving groups) in the Low Performance groups; (1) 23 students *actively* engaged in the tutorial sessions, while (2) 17 students did not engage in any activity during the Technology Acceptance Model (TAM) concept map activity. Therefore, the total *participation rate* in the Low Performance groups was only 57.5%. In other words, the total number of students who *actively* engaged in the Technology Acceptance Model (TAM) concept map activity was almost 1.7 times more in the High Performance groups. In 17 groups with High Performance (85%) all of the *four group members/students* participated in the Technology Acceptance Model (TAM) concept map activity, while in 2 groups (10%) *three group members/students* participated in the activity, and in 1 group (5%) *two group members/students* participated in the activity, and in none of the groups (0%) only *one group member/student* participated in the activity. Quite the reverse, in 2 groups with Low Performance (20%) all of the *four group members/students* participated in the Technology Acceptance Model (TAM) concept map activity, and in 2 groups (20%) *three group members/students* participated in the activity, and in 3 groups (30%) *two group members/students* participated in the activity, and in 3 groups (30%) only *one group member/student* participated in the activity. Therefore, the High Performance groups were obviously more involved in production of more collaborative processes with higher level of involvement and participation.

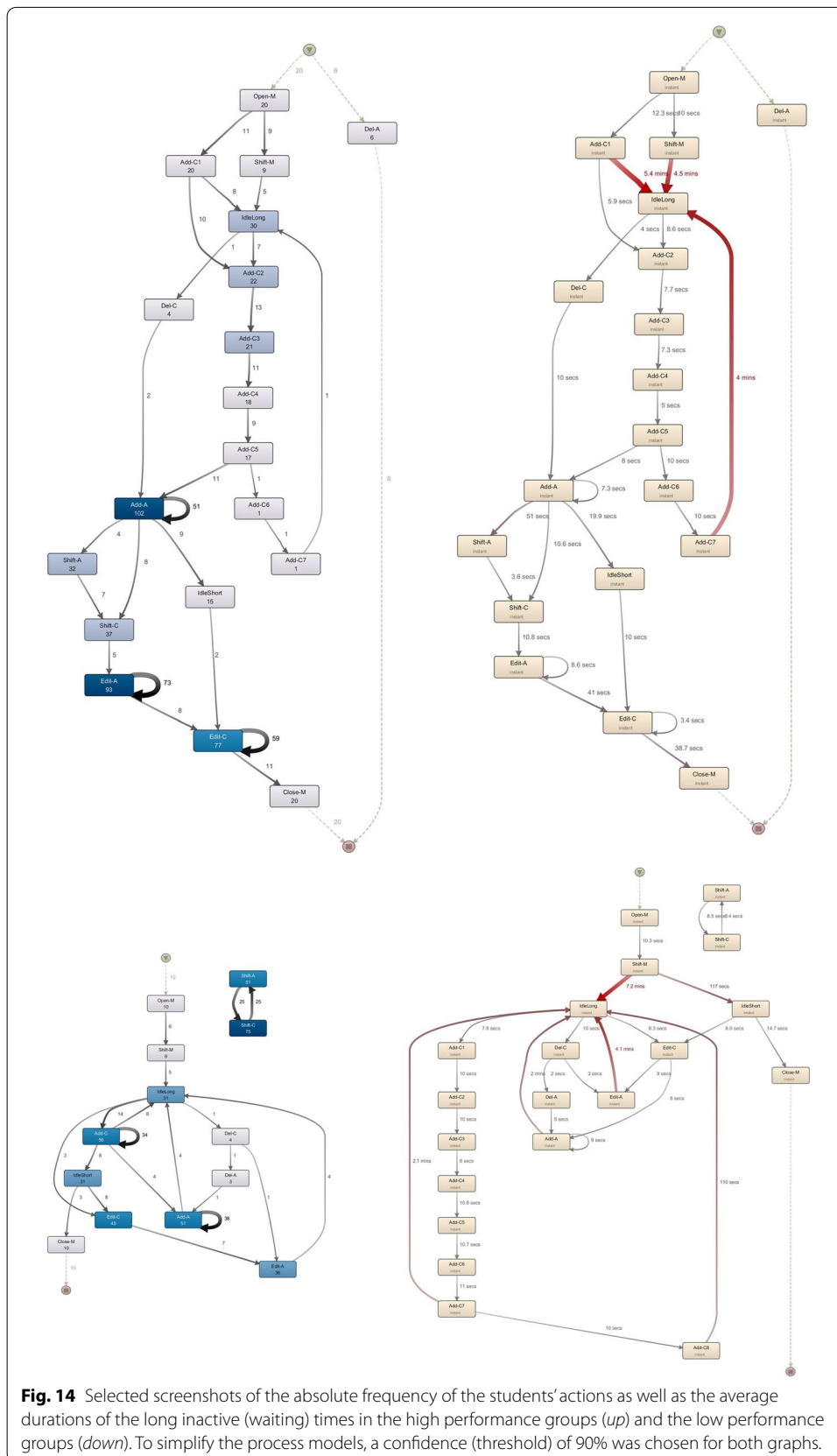
Thirdly and similar to an approach developed by Maldonado (2014) and Martinez-Maldonado et al. (2013b), we investigated the entire blocks of activity with respect to *1u* (i.e., the total number of actions performed by only one 1 student), *2u* (i.e., the total number of actions performed by only one 2 students), *3u* (i.e., the total number of actions performed by only one 3 students), and *4u* (i.e., the total number of actions performed by only one students). The results (see Fig. 13) showed that in the High Performance groups 91.74% of the activities (i.e., 500 activities) were executed by all of the



4 group members (i.e., *4u*). However, in the Low Performance groups, only 17.72% of the actions (i.e., 73 actions) were performed by 4 group members (i.e., *4u*), which was 5 times less than the High Performance groups. Similarly, in the High Performance groups none of the activities (i.e., 0 activities) was executed by 1 group member (i.e., *1u*), while in the Low Performance groups, only 13.59% of the actions (i.e., 56 activities) were performed by 1 group member (i.e., *1u*) which was extremely greater than the High Performance groups. Most of the actions (i.e., 159 actions) in the Low Performance groups were executed by 2 group members/students (i.e., *2u*) while most of the actions (i.e., 500 actions) in the High Performance groups were collectively performed by 4 group members. These results were not consistent with the findings of Martinez-Maldonado (2014) and Martinez-Maldonado et al. (2013b) as in their works most of the actions in both of the high and low achieving groups were performed by only 1 group member/student (i.e., *1u*).

#### Results of distinguished patterns of time intervals

Figure 14 shows the *mean (average) durations* for each activity and the critical paths (i.e., with long waiting times) in both of the High and Low Performance groups. Comparing the two maps we realized that both groups spent considerable *inactive (waiting) times* at the *beginning* of the Technology Acceptance Model (TAM) concept map activity and after shifting (or navigating through) the main menu window. Both of the High and Low Performance groups spent long waiting times (i.e., *IdleLong*) after instantly scrolling the Main Menu Window when the gap between *Shift-M* and *IdleLong* were 4.5 and 7.2 min in average, respectively. Therefore, the duration of the long waiting times (idle/inactivity time) consumed by the Low Performance groups at the beginning of the Concept Map Activity (distance learning) was 1.6 times greater compared with the High Performance groups. However, despite of the Low Performance groups, the High Performance groups showed long waiting times (i.e., *IdleLong*) instantly after creating the first component (i.e., *Add-C1*) when the gap between *Add-C1* and *IdleLong* was 5.4 min in average (i.e., the students were brainstorming together). However, the Low Performance groups did not show any long waiting times after creating the first component. Instead, the Low Performance groups spent long waiting times (i.e., *IdleLong*) only after editing the arrows (i.e., *Edit-A*) when the gap between *Edit-A* and *IdleLong* was 4.1 min in average in the groups with Low Performance (i.e., dealing with the *arrows* was the most



**Fig. 14** Selected screenshots of the absolute frequency of the students' actions as well as the average durations of the long inactive (waiting) times in the high performance groups (*up*) and the low performance groups (*down*). To simplify the process models, a confidence (threshold) of 90% was chosen for both graphs.

difficult part of the activity for Low Performance groups) (Martinez-Maldonado 2014; Martinez-Maldonado et al. 2013b).

Moreover, by contemplating on the results of Absolute Frequency of Actions in Fig. 14, we realized that adding an arrow (*Add-A*) with total frequency of 102 times, editing an arrow (*Edit-A*) with total frequency of 93 times, and editing a component (*Edit-C*) with total frequency of 93 times, were the most significant repetitive actions in the High Performance groups, respectively. Quite differently, shifting a component (*Shift-C*) with total frequency of 75 times, adding an arrow and shifting an arrow both with frequency of 51 times (*Add-A* and *Shift-A*), and editing a component (*Edit-C*) with total frequency of 45 times, were the most significant repetitive actions in the Low Performance groups, respectively (Martinez-Maldonado 2014; Martinez-Maldonado et al. 2013b).

#### Level of communication results

Using a qualitative approach (i.e., observation) to investigate the level of communication between the group members during the distance activity via *Chat Box*, we realized that the total number of words typed during the TAM concept mapping activity (7,893 words) was 3.7 times greater in the High Performance groups compared with the low performance groups (2,130 words). However, by considering the fact that the total number of High Performance groups was double compared with the low performance groups (20 groups versus 10 groups), we decide to calculate the average number of words typed per group via *Chat Box*. As shown in Table 7, the High Performance groups typed 395 words on average (i.e., divided by 20 groups) while the Low Performance groups typed 213 words on average (i.e., divided by 20 groups). Therefore, the average number of words typed per group was almost 2 times greater in the High Performance groups compared with the Low Performance groups.

Similarly, the average number of questions asked (per group) via *Chat Box* during the TAM concept mapping activity also was almost 2.5 times greater in the High Performance groups (3.63 questions asked on average) compared with the Low Performance groups (1.6 questions asked on average). In the same way, the average number of written

**Table 7 A semantic and textual analysis of the contributions typed/written and shared by students via chat rooms during the distance learning activity**

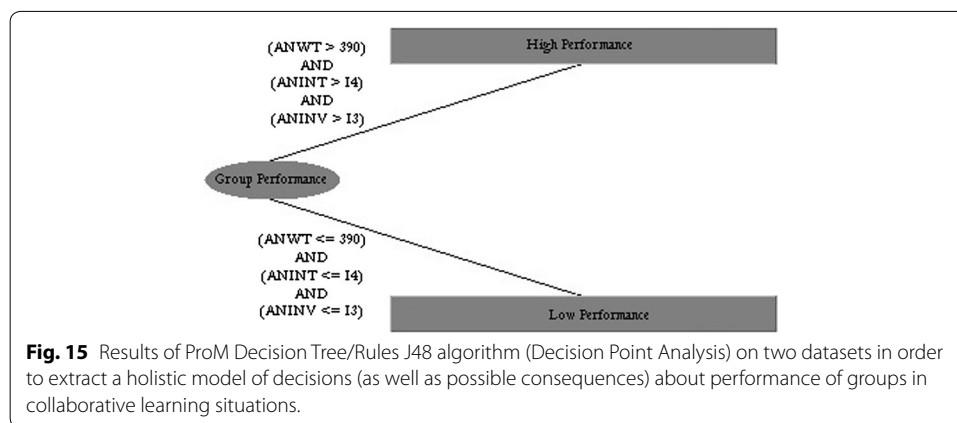
Contributions	High performance	Low performance
Total number of words typed during the activity	7,893	2,130
Average number of words typed (per group)	395	213
Total number of questions asked during the activity	73	16
Average number of questions asked during the activity (per group)	3.65	1.6
Total number of written sentences addressed to the whole group	13	4
Total number of written sentences addressed to a specific person	2	17
Total number of encouragements in the sentences	11	3
Total number of acknowledgements in the sentences	2	0
Total number of uncertainty expressions due to lack of experience in the sentences	0	4
Total number of uncertain statements (such as perhaps, maybe) in the sentences	5	37
Total number of certain statements (such as It is, I think, I'm sure) in the sentences	47	6

sentences addressed to the whole group was 1.5 times greater in the High Performance groups (0.6 sentences addressed to the whole group on average per group) compared with the Low Performance groups (0.4 sentences addressed to the whole group on average per group). On the contrary, the average number of written sentences addressed to a specific person was 17 times greater in the Low Performance groups (1.7 sentences addressed to a specific person per group on average) compared with the High Performance groups (0.1 sentences addressed to a specific person per group on average). This was completely consistent with our results earlier achieved about the fact that the level of interactions in the Low Performance groups is much less than the level of interactions in the High Performance groups. The average number of encouragements written in the sentences was almost two times greater in the High Performance groups (0.55 encouragements written per group on average) compared with the Low Performance groups (0.3 encouragements written per group on average). The total number of acknowledgements written in the sentences was 2 in the High Performance groups compared with 0 in the Low Performance groups. Therefore, the average number of encouragements and acknowledgements written in the sentences was more than two times greater in the High Performance groups (0.325 encouragements and acknowledgements written per group on average) compared with the Low Performance groups (0.15 encouragements and acknowledgements written per group on average). This means that students in the High Performance groups showed more positive feelings and constructive emotions in their communications with other fellow group members via *Chat Box* during the distance activity using *flowchart.com*.

Consequently, the average number of uncertainty expressions (due to lack of experience) and uncertain statements (such as perhaps, maybe) written in the sentences was 16.4 times greater in the Low Performance groups (4.1 per group on average) compared with the High Performance groups (0.25 per group on average). In other words, students in the Low Performance groups showed more negative feelings and emotions in their communications with other fellow group members via *Chat Box* during the distance activity using *flowchart.com*. And finally, the average number of certain statements (such as It is, I think, I'm sure, I'm positive, Definitely) written in the sentences during the distance activity via *Chat Box* was almost 4 times greater in the High Performance groups (2.35 certain statements per group on average) compared with the Low Performance groups (0.6 certain statements per group on average). This means that the level of certainty and self-confidence was noticeably greater in the groups with high performance compared with the groups with low performance. By defining the *level of communication* in terms of the "Total Number of Words Typed during the Activity" as well as the "Total Number of Questions Asked during the Activity", we realized that the *level of communication* was almost *double* in the High Performance groups (398.3 per group on average) compared with the Low Performance groups (214.6 per group on average).

### Decision mining results

By using Decision Tree/Rules technique and by mixing the event log collected from the Technology Acceptance Model (TAM) distance learning activity (by means of *flowchart.com*) with another event log previously collected from Theory of Reasoned Action (TRA) concept mapping activity (by means of *SMART Tables*), we were able to extract



important knowledge about the performance and behavior of groups in collaborative concept mapping situations. The rationale behind of applying a deciding mining approach was to discover a strategy that most likely is true about small groups of students—with respect to level of communication, level of interaction and level of involvement parameters—during the collaborative concept mapping activities.

Accordingly, based on the Decision Point Analysis technique (ProM 5.2) and based on the Decision Tree/Rules J48 algorithm we could identify the possible rules and consequences of collaborative actions/activities in both of the High and Low Performance groups. Figure 15 illustrates the common rules found in the High Performance groups during the two collaborative activities (i.e., TAM and TRA event logs) launched and collected in Thailand. As you consider, if *ANWT* (i.e., *Average Number of Words Typed*) in whole of the 2 collected datasets was greater than 390 words (per group) and then if *ANINT* (i.e., *Average Number of Interactions*) or absolute frequencies of *NoPossess* type of actions (per group) also was greater than 14 interactions, and then again if *ANINV* (i.e., *Average Number of Involvements*) or absolute frequencies of *Another* and *Simultaneous* types of actions (per group) also were greater than 13; then the performance of group is certainly ruled out as a group with *HP* (i.e., *High Performance*).

In the same way, if *ANWT* (i.e., *Average Number of Words Typed*) in whole of the 2 collected datasets was equal or less than 390 words (per group) and then if *ANINT* (i.e., *Average Number of Interactions*) or absolute frequencies of *NoPossess* type of actions (per group) also was equal or less than 14 interactions, and then again if *ANINV* (i.e., *Average Number of Involvements*) or absolute frequencies of *Another* and *Simultaneous* types of actions (per group) also were equal or less than 13; then the performance of group is certainly ruled out as a group with *LP* (i.e., *Low Performance*).

## Conclusions and discussions

The main objectives of this study were firstly to identify the most important factors affecting the performance of groups in Thailand, and secondly to conduct an empirical study to analyze student's collaborative behavior (as well as textual contributions) through process mining and statistical techniques in order to provide support to lecturers by increasing their awareness toward students' collaboration process in distance learning activities. Using *flowchart.com* as a free online real-time multi-user

collaborative concept map maker service, students were asked to build/construct a *Technology Acceptance Model (TAM)* in a distance learning approach.

In order to address *Question 1* of the study (i.e., *What are the most important factors that affect the performance of groups in distance learning situations in Thailand?*); a quantitative type of research in terms of a survey—and by distributing online questionnaires to 86 Bachelor students of 3 private universities in different parts of the Bangkok—was conducted. The results showed that “*Level of Communication*”, “*Level of Interaction*”, and “*Level of Involvement*” were the top-3 most significant factors/variables affecting the performance of groups in Thailand.

Referring to *Question 2* of the study (i.e., *How can the students’ interaction data be analyzed and interpreted in order to increase the lecturer’s awareness toward the collaborative activity process in the groups of students?*); specific Alphabets and Keywords were pre-defined and applied. In order to measure the “*Level of Communication*”, the *textual and semantic contributions*—made by the students in the chat environment (*Chat Box*)—was studied based on the qualitative metrics (through observation). In order to measure the “*Level of Interaction*”, the *ownership* of the actions that students perform on the objects created by themselves (or by other fellow group members) was studied (through statistical and process mining techniques). In order to measure the “*Level of Involvement*”, the *sequence or order* of the actions performed by the students during the distance learning activity was studied (through statistical and process mining techniques). In order to investigate the students’ actions with respect to the intervals of time, the inactivity times were divided into two main categories of *short idle time* and *long idle time*. In order to analyze the students’ actions with respect to the level of influence or impact on the concept mapping assignment, the actions were divided into two main categories of *high-impact* actions and *low-impact* actions. And finally, in order to distinguish the interaction behaviors between the small groups of students, the groups were divided into two main categories of *High Performance* groups and *Low Performance* groups.

Referring to *Question 3* of the study (i.e., *After the end of the class, can the lecturer discover and compare statistically relevant relationships on the interaction data between high performance groups and low performance groups based?*); the average time spent to finish the concept mapping activity were 12 and 18.6 min in the High Performance groups and the Low Performance groups, respectively. Therefore, none of the two groups consumed whole the legitimate time (30 min) to accomplish the distance concept mapping activity. The maximum and minimum durations of time spent to finish the distance concept mapping activity were 18 min and 5 s and 5 min and 55 s in the high Performance groups, respectively. The maximum and minimum durations of time to finish the same activity were 19 min and 57 s and 16 min and 25 s in the Low Performance groups, respectively. The average number of actions (events) executed in the High Performance groups was 27.25 actions.

On the other hand, the maximum and minimum numbers of students’ total actions (or events) to finish the distance concept mapping activity in the High Performance groups were 42 and 11 actions, respectively. By the same token, the maximum and minimum numbers of students’ total actions during the distance concept mapping activity in the Low Performance groups were 77 actions and 15 actions, respectively. The average

number of actions (events) executed in the low performance groups was 41.2 actions. This means that the students in the low performance groups performed more actions (on average) and created more events (on average).

Moreover, the distribution diagram of the low performance groups significantly demonstrates a very high ratio of actions performed per second a moment just before the end of the distance concept mapping activity, while the distribution diagram of the high performance groups shows an almost high ratio of actions performed per second during the middle of the distance concept mapping activity.

In addition, the occurrence of the keyword *NoPossess* was four times greater in the High Performance groups. Consequently, the level of students' interaction was almost four times greater in the High Performance groups compared with the Low Performance groups. Similarly, the occurrence of the keywords *Simultaneous* and *Another* was almost three times greater in the groups with high performance. Therefore, the level of students' involvement was almost three times greater in the high performance groups compared with the low performance groups.

Referring to *Question 4* of the study (i.e., *After the end of the class, can the lecturer discover and compare distinguished patterns (graphs/models) of interaction based on the students' interaction data between high and low performance groups?*); by visually comparing the resulting fuzzy mining graphs/models we realized that both of the High and Low Performance groups share identical core blocks of activity. However, the block "*HighOnly-NoPossess*" was the most significant trace of the interaction in the High Performance groups. Quite the opposite, the block "*HighOnly-Possess*" was the most significant trace of the interaction in the Low Performance groups. In other words, the Low Performance groups showed more tendencies to perform High-Impact actions on the objects *created and possessed by themselves* while students in the High Performance groups showed more tendencies to perform High-Impact actions on the objects *previously created and possessed by their other fellow group members*. In the same way, the Low Performance groups (on average) performed more *Low-Impact* and *No-Impact* types of actions compared with the High Performance groups.

The most significant blocks of activity in the High Performance groups (with regard to the fuzzy mining graph/model) were as follows: (1) *HighOnly-NoPossess*, (2) *HighOnly-NoPossess*, (3) *NoImpact-NoPossess*, (4) *NoImpact-Possess*, (5) *IdleLong*, (6) *IdleShort*, and (7) *LowOnly-NoPossess*. However, the most significant blocks of activity in the Low Performance groups (with regard to the fuzzy mining graph/model) were as follows: (1) *HighOnly-Possess*, (2) *NoImpact-NoPossess*, (3) *LowOnly-Possess*, (4) *NoImpact-Possess*, (5) *IdleShort*, and (6) *IdleLong*. Therefore, except the fourth significant factor (i.e., *NoImpact-Possess*), the rest of the blocks of activities were not consistent and compatible with each other in a comparison between the High and Low Performance groups. Additionally, the High Performance groups were more involved in production of more collaborative processes with higher level of interaction while expressing more complex handover of tasks from one student to another student. The High Performance groups also were more involved in production of more collaborative processes with higher level of involvement and participation compared with the Low Performance groups.

By using *Fuzzy Mining graphs* and based on the *Mean Duration* analysis of time intervals in Disco Fluxicon, we realized that the high performance groups showed long



waiting times a moment just after creating the first component (during the online concept mapping activity) which means the students were mostly brainstorming together. On the other hand, the low performance groups (quite opposite) did not show any long waiting times after creating the first component. Instead, the low performance groups spent long waiting times only after editing the arrows which means dealing with the *arrows* was the most difficult part of the activity for low performance groups.

Furthermore, by using Decision Point Analysis technique and based on the Decision Tree/Rules J48 algorithm we concluded that if “*Average Number of Words Typed*” was greater than 390 words (per group) and if “*Average Number of Interactions*” also was greater than 14 interactions, and then again if “*Average Number of Involvements*” also was greater than 13; then the performance of group is certainly ruled out as a group with “*High Performance*”.

Referring to *Question 5* of the study [i.e., *After the end of the class, can the lecturer analyze and investigate the contributions made by the students (i.e., semantically and textually) while working with distance learning activity in high and low performance groups?*]; the average number of words typed per group was almost 2 times greater in the High Performance groups. Similarly, the average number of questions asked (per group) via *Chat Box* during the online concept mapping activity also was almost 2.5 times greater in the High Performance groups compared with the Low Performance groups. In the same way, the average number of written sentences addressed to the whole group was 1.5 times greater in the High Performance groups while the average number of written sentences addressed to a specific person was 17 times greater in the Low Performance groups. The average number of both encouragements and acknowledgements written in the sentences was more than two times greater in the High Performance groups compared with the Low Performance groups. In other words, students in the High Performance groups showed more positive feelings and constructive emotions in their communications with other fellow group members during the distance activity.

Consequently, the average number of uncertainty expressions and uncertain statements written in the sentences was 16.4 times greater in the Low Performance groups compared with the High Performance groups. This means that students in the Low Performance groups showed more negative feelings and emotions in their communications with other fellow group members during the distance activity. Consequently, by defining the *level of communication* based on the “Total number of words typed during the activity” and “Total number of questions asked during the activity”, we realized that the *level of communication* was almost two times greater in the High Performance groups compared with the Low Performance groups.

One of the main limitations of the research was the fact that, in this study we considered textual and semantic words in order to analyze the extent of communication between small groups of students during the activity. However, the extent of communication definitely depends on many other variables and factors (such as body language, gestures, verbal and audio interactions, face-to-face communications, and etc.) as well. In the future, we aim to investigate the level of speech and verbal communication of students (by using microphones at their personal computers) during the distance learning activity as well. Another limitation of the study was the fact that different process mining algorithm (such as alpha mining, heuristic mining, genetic mining, and etc.) may

lead to different models/graphs with different structures and sequences. Similarly, different threshold (i.e., conformance or cutoff level) in each method results in different maps and models with dissimilar layouts and arrangement. Another limitation of the study is the fact that by using an online concept mapping activity we cannot merely analyze and investigate distinguished behavior of students during distance learning activities. In the future, we aim to apply more sophisticated techniques and methods of decision analytics on more number of students within e-Learning, Learning Management Systems (LMS) and Massive Open Online Course (MOOC) situations.

## Endnotes

<sup>a</sup>For further information about the flowchart.com please check the website: <http://flowchart.com/>.

<sup>b</sup>For further information about the Technology Acceptance Mode (TAM) please check the website: [https://en.wikipedia.org/wiki/Technology\\_acceptance\\_model](https://en.wikipedia.org/wiki/Technology_acceptance_model).

<sup>c</sup>For detailed information about the Process Mining and ProM please check the website: <http://www.processmining.org/>.

<sup>d</sup>For more information about Disco Fluxicon please check the website: <https://fluxicon.com/disco/>.

<sup>e</sup>For more detailed information about Decision Point Analysis supported in ProM framework please check the link: [http://www.processmining.org/\\_media/publications/rozinat2006.pdf](http://www.processmining.org/_media/publications/rozinat2006.pdf).

## Authors' contributions

WP supervised the research; made substantial contributions to conception and design of data analysis and interpretation of data; has been involved in revising the manuscript critically for important intellectual content. PP made substantial contributions to conception and design of data analysis and interpretation of data, collecting and acquisition of data; has been involved in drafting the manuscript. Both authors read and approved the final manuscript.

## Acknowledgements

The authors would like to thank Emeritus Professor Dr. James G. Williams from School of Information Sciences (University of Pittsburgh) for reviewing this paper and providing helpful insights.

## Compliance with ethical guidelines

### Competing interests

The authors declare that they have no competing interests.

Received: 5 March 2015 Accepted: 22 July 2015

Published online: 04 August 2015

## References

- Aalst, W. M. P. (2009). *Process mining: discovery, conformance and enhancement of business processes*. Berlin: Springer Verlag. ISBN 978-3-642-19344-6.
- Agarwal, S., Pandey, G. N., & Tiwari, M. D. (2012). Data Mining in Education: Data Classification and and Decision Tree Approach. *International Journal of e-Education, e-Business, e-Management and e-Learning*, 2(2), 140–144.
- Al-Barrak, M. A., & Al-Razgan, M. (2015). Predicting students final GPA using decision trees: a case study. *International Journal of Information and Education Technology*, 6(7), 528–533.
- Berea, A., Tsvetov, M., Daun-Barnett, N., Greenwald, M., & Cox, E. (2015). A new multi-dimensional conceptualization of individual achievement in college. *Decision Analytics*, 2, Art ID 3.
- Coleman, D. (2011). What is Web 3.0, and Why Do You Care. Retrieved 23 June 2015, from cmswire. <http://www.cmswire.com/cms/social-business/what-is-web-30-and-why-do-you-care-013072.php>.
- Concept map (2015). n.d. [https://en.wikipedia.org/wiki/Concept\\_map](https://en.wikipedia.org/wiki/Concept_map). Accessed 8 June 2015.
- Concept Mapping Fuels (2008). n.d. [http://www.energyeducation.tx.gov/pdf/223\\_inv.pdf](http://www.energyeducation.tx.gov/pdf/223_inv.pdf). Accessed 12 Feb 2015.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340. doi:10.2307/249008.
- Disco (2012). <https://fluxicon.com/disco/>. Accessed 8 Jan 2015.

- Elias, T. (2011). Learning analytics. Retrieved 23 June 2015, from Learning Analytics: Definitions, Processes and Potential. <http://learninganalytics.net/LearningAnalyticsDefinitionsProcessesPotential.pdf>.
- Evelien, O., Ronald, R. (2002). Social network analysis: a powerful strategy, also for the information sciences. *Journal of Information Science*. SAGE Publications. doi:10.1177/016555150202800601. Retrieved 3 March 2015, from [http://www.researchgate.net/profile/Ronald\\_Rousseau/publication/242401176\\_Social\\_network\\_analysis\\_a\\_powerful\\_strategy\\_also\\_for\\_the\\_information\\_sciences/links/0c960537a4bc5c6efa000000.pdf](http://www.researchgate.net/profile/Ronald_Rousseau/publication/242401176_Social_network_analysis_a_powerful_strategy_also_for_the_information_sciences/links/0c960537a4bc5c6efa000000.pdf).
- Flowchart.com, beta 2.4-rlive. (2014). (sunny Florida). Retrieved 23 June 2015, from Features. <http://flowchart.com/>. Accessed 23 June 2015.
- Fuzzy Miner, I. (2009). <http://www.processmining.org/online/fuzzyminer>. Accessed 8 Jan 2015.
- Günther, C., & Aalst, W. M. P. (2007). Fuzzy mining: adaptive process simplification based on multi-perspective metrics. In G. Alonso, P. Dadam, & M. Rosemann (Eds.), *International Conference on Business Process Management (BPM 2007)* (Vol. 4714, pp. 328–343)., Lecture Notes in Computer Science Springer: Berlin.
- Jishan, S. T., Rashu, R. I., Haque, N., & Rahman, R. M. (2015). Improving accuracy of students' final grade prediction model using optimal equal width binning and synthetic minority over-sampling technique. *Decision Analytics*, 2(1), 1.
- Martinez-Maldonado, R. (2014). *Analysing, visualising and supporting collaborative learning using interactive tabletops*. (Doctoral dissertation). Thesis for University of Sydney, Australia.
- Martinez-Maldonado, R., Dimitriadis, Y., Kay, J., Yacef, K., Edbauer, M.T. (2013a). MTClassroom and MTDashboard: supporting analysis of teacher attention in an orchestrated multi-tabletop classroom. In: *Proceedings of the International Conference on Computer Supported Collaborative Learning (CSCL2013)*, (pp. 119–128).
- Martinez-Maldonado, R., Kay, J., Yacef, K. (2012). Analysing knowledge generation and acquisition from individual and face-to-face collaborative concept mapping. In A.J. Canas, J.D. Novak, J. Vanhear, editors, *Concept Maps: Theory, Methodology, Technology Proc. of the Fifth Int. Conference on Concept Mapping* (pp. 17–24).
- Martinez-Maldonado, R., Yacef, K., Kay, J. (2013b). Data Mining in the Classroom: Discovering Groups' Strategies at a Multi-tabletop Environment. In *Proceedings of the International Conference on Educational Data Mining 2013 (EDM 2013)* (pp. 121–128).
- McGrath, J. E. (1984). *Groups: interaction and performance* (1st ed.). New Jersey: Prentice-Hall Inc.
- McGrath, J. E. (1991). Time, interaction, and performance (TIP): a theory of groups. *Small Group Research*, 22(2), 147–174.
- Novak, J. (1990). Concept maps and Vee diagrams: two metacognitive tools to facilitate meaningful learning. *Instructional Science*, 19(1), 29–52. doi:10.1007/BF00377984.
- Östlund, B. (2008). Interaction and Collaborative Learning—If, Why and How? Retrieved 2 July 2015, from Umeå University. [http://www.eurodl.org/materials/contrib/2008/Berit\\_Ostlund.htm](http://www.eurodl.org/materials/contrib/2008/Berit_Ostlund.htm).
- Premchaiswadi, W., Porouhan, P., Premchaiswadi, N. (2012). Online Robotics Course: Factors Affecting Students' Satisfaction toward. *International Journal for e-Learning Security (IJeLS)*, 2(3/4), 181–191.
- Porouhan, P., & Premchaiswadi, W. (2011). Factors affecting the passengers' intention toward "airline electronic ticketing" in Thailand. In *ICT and Knowledge Engineering (ICT & Knowledge Engineering)* (pp. 177–186). Bangkok: IEEE Xplore.
- Process Mining Group (2009). Decision Miner. <http://www.processmining.org/online/decisionmining>. Accessed 7 June 2015.
- Process mining (2011). n.d. [https://en.wikipedia.org/wiki/Process\\_mining](https://en.wikipedia.org/wiki/Process_mining). Accessed 8 June 2015.
- ProM. (2011). <http://www.processmining.org/prom/start>. Accessed 11 Jan 2015).
- ProM Tips- Which Mining Algorithm Should You Use. (2010). <https://fluxicon.com/blog/2010/10/prom-tips-mining-algorithm/>. Accessed 11 Jan 2015.
- Rozinat, A., & Aalst, W. M. P. (2006). Decision mining in ProM. *Business Process Management*, 4102, 420–425.
- Seven Things you should know about analytics (2010). <https://net.educause.edu/ir/library/pdf/ELI7059.pdf>. Accessed 8 Jan 2015.
- Social Network Miner. (2012). <http://www.processmining.org/online/snminer>. Accessed 8 Jan 2015.
- Spivack, Nova. Web 3.0: The Third Generation Web is Coming. (2015). <http://lifeboat.com/ex/web.3.0>. Accessed 8 July 2015.
- Technology Acceptance Model (2003). Retrieved 23 June 2015, from Wikipedia. [https://en.wikipedia.org/wiki/Technology\\_acceptance\\_model#CITEREFDavis1989](https://en.wikipedia.org/wiki/Technology_acceptance_model#CITEREFDavis1989). Accessed 23 June 2015, from Wikipedia.
- Wang, S. L., Hsu, H. Y., Lin, S. S., & Hwang, G. J. (2014). The Role of Group Interaction in Collective Efficacy and CSCL. *Educational Technology & Society*, 17(4), 242–254.
- Web 3.0 Software Service. (2014). <http://flowchart.com/>. Accessed 18 June 2015.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)

---