Amara, S., Macedo, J., Bendella, F., & Santos, A. (2016). Group Formation in Mobile Computer Supported Collaborative Learning Contexts: A Systematic Literature Review. *Educational Technology & Society, 19* (2), 258–273.

Group Formation in Mobile Computer Supported Collaborative Learning Contexts: A Systematic Literature Review

Sofiane Amara[1*], Joaquim Macedo[2], Fatima Bendella[1] and Alexandre Santos[2]
[1]Department of Informatics, Faculty of Mathematics and Informatics, University of Sciences and Technology of Oran "Mohamed Boudiaf", USTO-MB, Oran, Algeria // [2]Centro Algoritmi, University of Minho, Braga, Portugal // sofiane.amara@univ-usto.dz // macedo@di.uminho.pt // fatima.bendella@univ-usto.dz // alex@di.uminho.pt
[*]Corresponding author

(Submitted February 13, 2015; Revised June 18, 2015; Accepted July 28, 2015)

ABSTRACT

Learners are becoming increasingly divers. They may have much personal, social, cultural, psychological, and cognitive diversity. Forming suitable learning groups represents, therefore, a hard and time-consuming task. In Mobile Computer Supported Collaborative Learning (MCSCL) environments, this task is more difficult. Instructors need to consider many more issues, such as the rapid change of mobile learners' context, their direct and naturel interaction, and the characteristics of mobile devices and networks. This paper presents a systematic literature review (SLR) that examines the relevant solutions for the problem of group formation in MCSCL environments. In the context of this SLR, an initial list of 178 papers was reviewed. After careful analysis of each paper using specific selection criteria and a quality assessment method, a final list of 12 relevant studies was filtered and used to answer the research questions. The findings revealed that: (a) there is a lack of approaches addressing the group formation problem in MCSCL environments; (b) the most proposed solutions do not allow instructors to customize the grouping process; (c) there is no useful solutions to automatically capture and evaluate many of learners' behaviours and context information; (d) the majority of approaches do not support a dynamic formation of learning groups; (e) the majority of approaches do not provide descriptions about the implemented grouping algorithms nor about the evaluation methods. Extracted and synthesized data from the selected studies is discussed in this paper, together with current research gaps and recommendations for further works.

Keywords

Learning group formation, Collaborative learning, M-learning, MCSCL, Systematic literature review

Introduction

Collaborative Learning (CL) represents an essential educational approach defined by Dillenbourg as "a situation, in which two or more people learn or attempt to learn something together" (Dillenbourg, 1999). Henri and Lundgren-Cayrol (2001) consider CL as an active process by which the learner is working on the construction of knowledge, the instructor plays the role of facilitator of learning, and the group participates as a source of information, as a motivator, as a mean of self-help and mutual support, and as a preferred place of interaction for collective construction of knowledge. Many researchers demonstrated how CL is useful for improving the cognitive, psychological, and social development of learners (Dillenbourg, 1999; Zurita et al., 2005).

The development of information and communication technologies has led to the emergence of e-learning. It is a kind of learning based principally on the use of computers for constructing and delivering knowledge. Education researchers began then to search how to benefit from this technological evolution to improve the CL pedagogies. As result, since the late 1990s, a new branch of collaborative learning called Computer Supported Collaborative Learning (CSCL) has emerged.

Furthermore, the rapid development of wireless communication and mobile technologies enabled the emergence of a new form of learning termed M-learning. It allows learners through the use of mobile devices (PDAs, tablets, Smartphones, etc.) to learn anytime and anywhere, in formal or informal places. As result, CL has become possible in mobile situations and real world environments. Hence, Mobile Computer Supported Collaborative Learning (MCSCL) represents a new paradigm of CL that is growing in use. This reality is confirmed by the big number of established MCSCL projects (Yatani et al., 2004; Zurita et al., 2005; Boticki et al., 2011; Huang et al., 2014).

On the other hand, forming effective learning groups represents one of the important factors that determine the efficiency of CL. According to (Dillenbourg, 2002), studies show that three key conditions are required for any successful CL: the task features, the communication media, and the group composition. The importance of learning group formation (LGF) process is confirmed by many researchers in the literature (Huang & Wu, 2011; Webb et al., 1998). Nevertheless, the social, cultural, psychological, and cognitive diversities of learners make the operation of forming suitable learning groups a hard and time-consuming task.

Although MCSCL represents a multidisciplinary research field (e.g., psychology, education, computer science), and although the importance of LGF process in succeeding the MCSCL activities, there is until date no effort to analyse the state of research on this topic. Thus, this paper reviews the studies addressing the topic of LGF in MCSCL using a systematic literature review (SLR) methodology. This SLR provides explicit information on what criteria are used for forming learning group, how to manually or automatically collect and use these criteria, how to support the dynamic formation of learning groups, what algorithms are used to form learning groups, and what methods are followed to evaluate the proposed approaches.

The paper is organised in four main parts. The first one shows how forming groups in MCSCL is compared to that of traditional environments. The second part describes the research methodology used in this work. Then, the main findings and gaps from this SLR, together with recommendations for further research are presented. Finally, our conclusions are provided.

Related research problem

One of the important questions raised in this research is whether the LGF approaches addressing the traditional environments are effective for MCSCL. To answer this question, it is important to know if MCSCL can be considered as only an extension of CSCL supported by mobile devices. The majority of researchers affirm that the answer is NO. M-learning context is different from that of more traditional e-learning (Parsons & Ryu, 2006). Additionally, Mobile collaborative applications do not replicate traditional learning scenarios, but they offer new learning opportunities, which cannot be reached without mobile technologies (Patten et al., 2006). Therefore, MCSCL does not mean “mobile + CSCL” (Looi et al., 2013); each paradigm has its particular environments, technologies, characteristics, practices and objectives.

MCSCL is highly dynamic in terms of users’ contexts. Mobile learners can dynamically obtain helps, recommendations, or learning content depending on their current context. For instance, when learners are near to a point of interest (POI), they receive information related to that POI, and when they move to another POI, the provided content is changed too. Contrarily, traditional environments are unable to adapt to dynamic changes of users’ context (current location, distances between learners, learning objects availability, etc.).

CSCL grouping algorithms are implemented to be run on high performances computers. Without Internet, those CSCL’s algorithms cannot be executed on mobile devices due to their weak technical characteristics. Therefore, new lightweight algorithms specific for mobile devices should be implemented and used with other types of networks such as Delay Tolerant Networks (DTNs).

CSCL’s learners are unable to communicate naturally, to move freely together, and to interact directly with learning objects. In the contrary, MCSCL’s learners are always in motion with face-to-face interactions. They may have different movement patterns (active, passive, etc.), levels of dialog and communication (social, shy, introvert, etc.), and preferences (preferred places, partners, learning objects, etc.). Those kinds of behavioural information are not considered in traditional environments.

Unlike CSCL, MCSCL’s activities are generally exposed to many technical problems (disconnections, battery depletion, memory saturation, etc.); social problems (misunderstanding, disunion, selfishness, etc.); and natural/geographical problems (land degradation, weather changes, etc.). Those problems force the LGF approaches to be dynamic and able to (re)-form the groups in real-time. Such a mechanism of dynamically forming groups is completely ignored in traditional environments.

Taking into consideration those differences, one can affirm that using the same traditional solutions to form groups in MCSCL environments without considering their particularities could cause many problems such as, disunion of groups (e.g., when ignoring the learners’ geographical locations); demotivation, introversion, and isolation of learners (e.g., when ignoring the personal traits of mobile learners, and their different learning and social behaviours); and obstruction of collaborative activities (e.g., when ignoring the different social, technical, and geographical problems that can happen in MCSCL environments).

Methodology

In this research, a guideline for performing systematic literature reviews (SLR) proposed by Kitchenham (2007) is followed. This methodology represents an efficient way to evaluate existing works relevant to particular research questions. The disadvantage of this method is that it necessities more efforts than traditional methods of literature review. Figure 1 shows the followed steps to carry out this SLR.

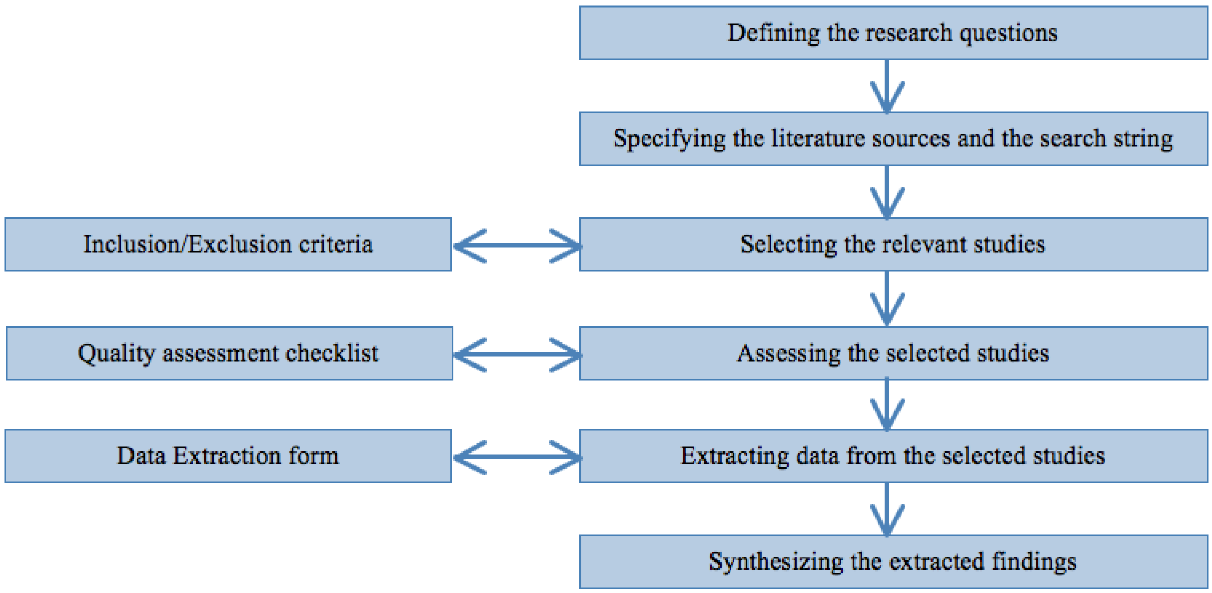

Figure 1. The followed steps for conducting the SLR

Research questions

RQ1: What are the learners' personal characteristics used as grouping criteria?

Learners' personal characteristics could be used whether in mobile environment or not. But, the objective of using this RQ is to identify which personal characteristics are more used and more appropriate for MCSCL context, and how they are combined with the other grouping criteria (learners' behaviours and contextual information).

RQ2: Which learners' behaviours are used as grouping criteria?

Mobile learners may have different behaviours (they can be confident/afraid, active/passive, social/introvert, nervous/calm, etc.). This RQ aims to identify the different considered learners' behaviours, and how the proposed MCSCL systems obtain, evaluate, and store this behavioural data.

RQ3: Which kinds of context information are used as grouping criteria?

Context is defined as any information that can be used to characterize the situation of an entity (e.g., learner, learning object, device). While MCSCL environments are known by their context awareness, the learning groups can be properly formed according to different context information. In this SLR, we search to identify which kinds of context information are more considered, how this information is gathered from the mobile devices or other technologies (sensors, smart objects, etc.), and how it is used and combined with the other criteria to form learning groups.

RQ4: How the dynamic grouping can be supported in MCSCL environments?

MCSCL activities are generally exhibited to several problems that may lead to stop the collaboration between learners at any moment. For instance, the face-to-face interactions between learners may cause some social problems (e.g., selfishness, misunderstanding, disunion). Additionally, the technical limitations of mobile devices (connection rupture, battery depletion, memory saturation, etc.), and the geographical dispersion of learners in vast learning areas may break the communication between learners and destroy the learning groups. In such cases, the dynamic composition of groups plays an important role for quickly regrouping learners based on their updated information. The dynamic grouping processes can have two forms:

- Inter-sessions grouping: allows forming groups only before starting or after ending learning process. This kind of grouping is more useful for asynchronous mobile CL.
- Intra-session grouping: allows changing group members during the learning process. This kind of grouping is more useful for synchronous mobile CL.

Therefore, this RQ serves to identify the different solutions proposed for ensuring the two forms of dynamic formation of MCSCL groups.

RQ5: What algorithms are used for forming learning groups?

In MCSCL environments, the grouping algorithms do not only serve to use some stored learners' data to form the groups, but it should capture and evaluate the learners' behaviours, and interact with different components of the system to get instantaneous context information. Hence, this RQ aims to identify and examine the proposed or re-implemented LGF algorithms specific for MCSCL environments.

RQ6: What methods are used to evaluate group formation processes?

The best way to evaluate a MCSCL's LGF approach is to test it in real world context. However doing such experiments requires human resources (learners, instructors, etc.), material resources (mobile devices, networks, sensors, etc.), and a long time period. At the contrary, simulation methods allow researcher to assess their approaches several times with several settings and grouping criteria. However, modelling the different components of MCSCL systems (especially the human behaviours) represents another difficult challenge. Therefore, the answers of this RQ will show the different methods used to assess the proposed LGF solutions.

Beside those principal six RQs, this SLR extracts other information related to LGF problem, such as the types of learning groups, and the customization of the grouping processes.

Literature sources and search terms

The list of literature sources used in this SLR includes online databases, search engines, individual journals, and proceedings of scientific events (conferences and workshops) (Table 1).

Table 1. Literature sources

Resource type	Resource name
Online databases	ACM, IEEE Explore, ScienceDirect, Scopus, SpringerLink, Web of Science,
Online search engines	CiteSeer, Google Scholar
Individual journals	Journal of Educational Technology & Society, International Journal on E-learning, Journal of Educational Data Mining, International Journal of Learning Technology
Others	Proceedings of scientific events

After defining the list of literature sources, a search string was constructed using the following method (Kitchenham et al., 2007):

- Derive the major search terms (Table 2);
- Check the keywords list of already analysed papers to find more search terms;
- Identify alternative spellings and synonyms for each major term;
- Construct the search string using Boolean ORs to join alternative spellings and synonyms, and Boolean ANDs to join major search terms.

Table 2. Search terms with their alternative spellings and synonyms

Terms	Alternative spellings and synonyms
Collaborative learning	MCSCL, CSCL, Mobil learning; Ubiquitous learning; Pervasive learning.
Group	Team; Cluster; Set; Community; Assembly.
Formation	Construction; Composition; Organization; Constitution; Creation; Building; Assigning.

The major search terms used in this SLR are “Collaborative learning” and “Group” and “Formation.”

The resulting search string is as follows: (Collaborative learning ***OR*** MCSCL ***OR*** CSCL ***OR*** Mobile learning ***OR*** Ubiquitous learning ***OR*** Pervasive learning) **AND** (Group ***OR*** Team ***OR*** Cluster ***OR*** Set ***OR*** Community ***OR*** Assembly) **AND** (Formation ***OR*** Construction ***OR*** Composition ***OR*** Organization ***OR*** Constitution ***OR*** Creation ***OR*** Building ***OR*** Assigning).

Studies selection

To ensure identifying all relevant studies, this SLR used the following search method:

- Searching online engines and online databases;
- Searching manually scientific events proceedings and individual journals;
- Scanning the references lists of all found papers in order to avoid missing any interesting study.

By following this search method, an initial list of 178 papers was identified. However, this list includes some studies that do not address exactly the described research problem, or studies that are stored in multiple databases or published in many sources. Therefore, the objective of this stage is to eliminate duplicate papers and filter the relevant ones from the set of all found studies. The selection method is based on the use of the following inclusion and exclusion criteria.

Inclusion criteria

- If a study has both conference version and journal version, only journal version is included;
- If a study has many published versions, only the newest and the most complete version is included;
- If a study is stored in multiple sources, only one copy of this paper is included.

Exclusion criteria

- Papers that do not consider MCSCL environments;
- Papers that do not address the problem of LGF.

After applying those inclusion/exclusion criteria, a list of 20 papers was filtered (Table 3).

Table 3. Distribution of found and selected studies in search sources

Source	Found studies	Selected studies
ACM	10	1
IEEE Explore	23	5
ScienceDirect	19	1

Scopus	55	3
SpringerLink	12	1
Web of Science	38	0
Educational Technology & Society	5	4
International Journal on E-learning	1	1
Journal of Educational Data Mining	1	1
International Journal of Learning Technology	1	0
Other conferences and workshops	13	3
Total	178	20

Quality assessment

For well assessing and classifying the selected studies, a quality assessment checklist was developed (Table 4). Each paper was assessed independently by two authors. The quality assessment checklist is composed of nine questions labelled from QA1 to QA9, and each question is scored as follows:

QA1 is evaluated depending on the source of the paper:

- For conferences and workshops papers, the computer science conference ranking (CORE list) is used. The possible values are: A (1.5); B (1); C (0.5); No CORE ranking (0).
- For Journals articles, the Journal Citation Reports (JCR) is used. The possible values are: Q1 (2); Q2 (1.5); Q3 (1); Q4 (0.5); No JCR (0).

QA2 to QA9 should have one of the following values: Yes (1); Partially (0.5); No (0).

Table 4. Quality assessment checklist

ID	Question	Value
QA1	Is the study published in a recognized journal or scientific event proceeding?	• CORE ranking • JCR ranking
QA2	Is there a clear statement of the aim of research?	• Yes • No • Partially
QA3	Does the study discuss any of related studies?	• Yes • No
QA4	Does the proposed approach allow a dynamic grouping?	• Yes • No
QA5	Does the study consider learners' learning behaviours?	• Yes • No
QA6	Does the study consider context information?	• Yes • No
QA7	Is the experimental procedure carefully explained?	• Yes • No • Partially
QA8	Are the findings clearly stated and presented?	• Yes • No • Partially
QA9	Was the paper cited by other researchers?	• Yes • No

The quality assessment scores of each study given by two independent authors were saved and used to calculate the average score between them. Each study that has a quality score less than 5 was removed. At the end of this stage, a final list of 12 papers was obtained. The selected studies are labelled from S1 to S12 (Table 5). The quality assessment scores of each study are presented in Table 6.

Table 5. Selected studies

Study	Reference	Research questions addressed					
S1	Huang & Wu, 2011		2	3		5	
S2	Zurita et al., 2005	1			4		6
S3	Huang et al., 2010		2				6
S4	Messeguer et al., 2010		2	3		5	
S5	El-Bishouty et al., 2010	1		3			
S6	Hsieh et al., 2010	1	2	3		5	
S7	Tan et al., 2010	1		3	4		
S8	Giemza et al., 2013	1	2				6
S9	Mujkanovic et al., 2012	1	2				6
S10	Yin et al., 2012		2	3			
S11	Yang et al., 2007		2		4	5	6
S12	Muehlenbrock, 2005	1		3		5	

Table 6. Quality assessment scores

Study	QA1	QA2	QA3	QA4	QA5	QA6	QA7	QA8	QA9	SCORE
S1	2	1	1	1	1	1	1	1	1	10
S2	2	1	1	1	0	0	1	1	1	8
S3	2	1	1	1	1	0	0.5	0.5	1	8
S4	0	1	1	1	0	1	1	1	1	7
S5	2	0	1	0	0	1	1	1	1	7
S6	0	1	1	1	0	1	1	0	1	6
S7	1	1	1	1	0	1	0	0	1	6
S8	1	0.5	1	0	0.5	1	0.5	0.5	0	5
S9	0	0.5	0.5	1	0.5	0.5	1	1	0	5
S10	0	0	1	1	1	1	0	0	1	5
S11	2	0.5	1	1	1	0	1	1	1	8.5
S12	1.5	1	1	0	0	1	1	1	1	7.5

Data extraction

To facilitate extracting the relevant data, each reviewer has used a data extraction form that lists the key information to be collected from each study (Table 7). Data extraction tables were then used to record the extracted data.

Table 7. Data extraction form

Data ID	Data
D01	Study identifier
D02	Name of author(s)
D03	Paper’s title
D04	Year of publication
D05	Paper’s type (Journal, conference/workshop proceedings)
D06	Quality assessment score
D07	Grouping type (homogeneous/heterogeneous)
D08	Whether the study supports a dynamic grouping or not
D09	Personal characteristics used as grouping criteria
D10	Learner’s behaviours used as grouping criteria
D11	Context information considered by the LGF process
D12	Used grouping algorithm(s)
D13	Methods used for evaluating the proposed solution

Data synthesis

Data synthesis phase serves to summarize and report the important results obtained from the analysis of the selected studies. To achieve this objective, the following strategy was followed.

- Answer individually the research questions by consulting the data extracted from the previous stage;
- Search additional findings besides the ones directly related to the research questions;
- Identify research gaps and provide recommendations for further research.

Results

Overview of studies

As shown in Figure 2, the majority of papers were published after 2005. This is probably due to the significant increase of mobile technologies' usage in this period. For instance, the ratio of mobile cellular telephone subscription has reached from 20% in 2004 to 95% in 2014 (Figure 3) (World-Telecommunication, 2014). Although the mobile technologies are growing in use compared to the fixed ones, we cannot affirm that mobile CL is replacing CSCL. Each one of them has its particular characteristics, methods, practices, and objectives. For instance, we cannot develop complex applications using mobile devices, and we cannot perform location-based activities in real world context using desktop computers.

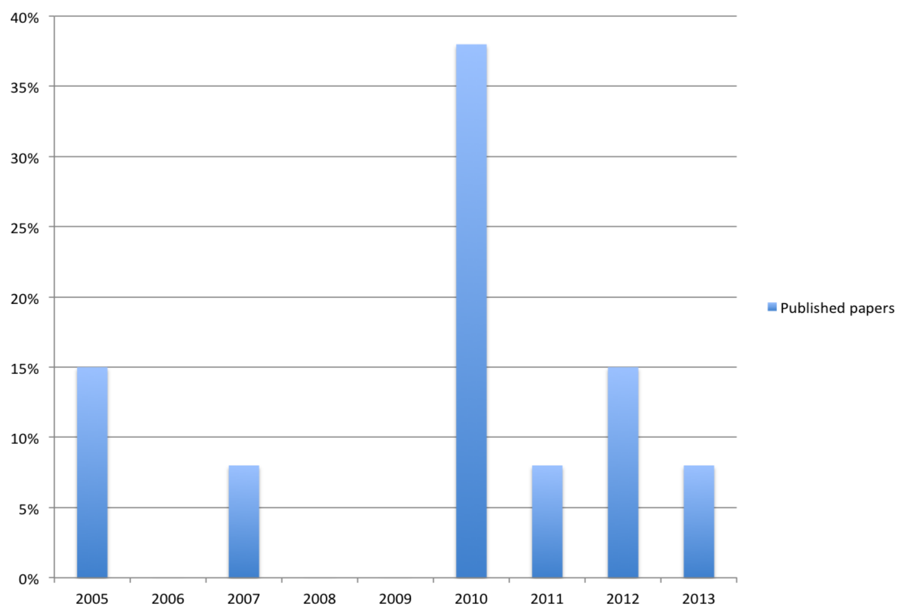

Figure 2. Publication percentage per year

As shown in Table 8, the majority of studies do not combine the three types of grouping criteria in a single grouping process. By analysing the nature of learning groups in terms of homogeneity/heterogeneity of learners, this SLR classifies the selected studies into three groups: (a) Studies aim to form heterogeneous groups by maximizing the diversity within group (S1, S2, and S6); (b) Studies aim to form homogeneous groups by minimizing the diversity within group (S3, S8, S10 and S11); (c) Studies consider principally the criteria related to learning environments, and do not pay attention to the similarities/differences of learners (S4, S5, S7, S9 and S12). Generally, researchers in literature recommend the first type (heterogeneous grouping) to be beneficial, because it helps removing barriers between learners and improving their interaction and creativity (Hübscher, 2010). Other researchers find that it is useless to apply a specific grouping kind for all types of learners. Therefore, it is useful to leave the choice to instructors for selecting the nature of the groups according to different learning's objectives, learners' needs, activities' types, etc.

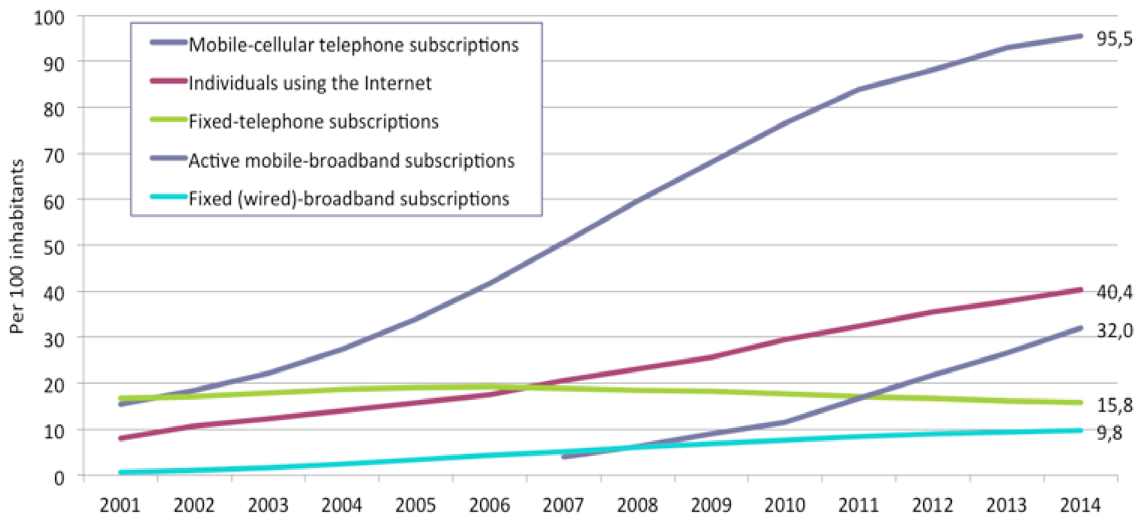

Figure 3. Evolution of telecommunication technology usage

Table 8. General features of the reviewed studies

Study	Grouping type		LGF criteria			LGF characteristics	
	Heterogeneous	Homogeneous	Personal characteristics	Learning behaviours	Context information	Customized	Dynamic
S1	✓	✗	✗	✓	✓	✓	✗
S2	✓	✗	✓	✗	✗	✓	✓
S3	✗	✓	✗	✓	✗	✗	✗
S4	✗	✗	✗	✓	✓	✗	✗
S5	✗	✗	✓	✗	✓	✗	✗
S6	✓	✗	✓	✓	✓	✗	✗
S7	✗	✗	✓	✗	✓	✗	✓
S8	✗	✓	✓	✓	✗	✗	✗
S9	✗	✗	✓	✗	✗	✗	✗
S10	✗	✓	✗	✓	✓	✗	✗
S11	✗	✓	✓	✓	✗	✗	✓
S12	✗	✗	✓	✗	✓	✓	✓

This SLR classifies also the reviewed studies according to their ability to customize the grouping process. We found that the majority of studies are not customizable. Instructors are generally unable to introduce their own choices for forming groups. For instance, they cannot select the nature of learning groups, the type and the number of grouping criteria, the number of learners in each group, etc. Instead, Studies S1, S2, and S7 provide some opportunities for customizing the groping process. S1 allows choosing the number of groups and introducing a threshold of difference between learners. S2 enables (re)composing the learning groups by selecting the criteria that adapt the learning objectives. S7 Allows learners to freely select one of two provided grouping options to be used by the algorithm.

Findings on the research questions

The following subsections present the extracted answers from the reviewed studies.

RQ 1: Learners' characteristics used as grouping criteria

Since the majority of selected studies address MCSCL environments, the researchers pay more attention to criteria related to these environments and ignore the personal characteristics of learners. The most used learners' personal characteristics are: gender (S6, S9), age (S7, S9), preferences (S2), interests (S3, S7, S8, S10), and learning backgrounds and experiences (e.g., languages, learning scores, learning capacities) (S5, S6, S7, S9, S11, S12).

RQ2: Learners' behaviours used as grouping criteria

S1 proposes the use of an U-learning portfolio tool to collect and evaluate certain behaviours (see Table 9). The authors developed a systematic grouping scheme based on three successive processes: transformation of data from U-portfolio to a Portfolio Grid, building a learner similarity matrix, and using two clustering algorithms to distinguish learners into appropriate heterogeneous groups.

S3 presents a process of partners recommendation based on the analysis of learners' reading interests. The proposed approach collects behavioural data from the social platform Del.icio.us, and creates a behaviour profile for each learner. Collaborative services use then these profiles to recommend partners with similar interests and form appropriate learning communities.

S4 proposes a LGF approach based on a supervised machine-learning intelligent system. Depending on available contextual information (time, place, neighbourhood, etc.) captured from the learners' devices, the system can automatically estimate the composition of learning groups.

S6 uses learners' social interactions as the essential criterion to form learning groups. Highly interactive learners are identified as "hub's learners," and low interactive learners are identified as "island's learners." The two types of learners should be assigned evenly in each group.

S8 presents a mobile application for supporting the learners to form informal learning groups. The grouping mechanism uses learners' profiles containing data about the course of studies, the number of semesters and the attended lectures. Based on this data the system provides recommendations to each learner.

S10 shows an approach for recommending partners according to the helping history between learners. The system could evaluate automatically the level of personal relationship between each pair of learners according to the frequency of the helping each other. Based on this information, recommendations of partners are provided to each learner.

RQ3: Context information used as grouping criteria

S1 proposes a location profile to record learners' movements during their learning activities using Radio Frequency Identification (RFID) technology.

S4 considers some kinds of context information, such as: time and day of week, place, and list of neighbours (using Bluetooth' MAC addresses). This information is combined with learners' behaviours to train and test the machine-learning model.

S5 uses RFID technology to detect the surrounding objects. Information about detected objects in addition to the learners' location allow the system to provide social knowledge awareness map for peer helpers.

S6 proposes a grouping solution based on the learners' locations and their interactions. To detect the location of participants, each learner carries an ubi-coin (wireless detector), and each ubi-coin represents a node. If two nodes are within a specific distance for a moment, the system considers that they are in interaction.

S7 uses learners' geographical locations as the principal criterion to form learning groups. The system uses certain technologies offered by the mobile devices (e.g., Wi-Fi, and Bluetooth, GPS) to obtain the learners' locations.

S10 helps learners to construct a social learning network based on their location information collected through GPS sensors. This context information allows each learner to know the current location of other learners who have the similar interests. The system recommends then CL activities between them.

S12 presents an important solution allowing an ad-hoc creation of learning groups, using some context information (learner's location, ambient sounds from PDA devices, and learner's availability).

Table 9 summarizes the grouping criteria considered by the reviewed studies.

Table 9. Used grouping criteria

Study	Personal characteristics	Learning behaviours	Context information
S1		Observing/Answering quiz/Interacting/Moving/Losing/ Answering questions/Referencing/ Completing tasks/Taking note	Location
S2	Preferences/Academic performance/Sociability		
S3		Past activities (read books)	
S4		Preferences (places, partners, subjects)/Time spent for learning	Time/Location/ Neighbourhood
S5	Knowledge and experiences/Interests	Interaction	Availability of educational materials/Location
S6	Gender/Personal background	Interaction	Location
S7	Age/Learning subjects/Learning results/Learning styles/Learning interests		Location
S8	Interests	Course of studies/Number of semesters/Attended lectures	
S9	Gender/Age/Motivation /Previous knowledge		
S10		Helping history	Location
S11	Knowledge	Browse online course materials/Submit questions/Send messages to others/Perform exercises/Take online tests	
S12	Knowledge (Capacities/difficulties)		Location/Ambient sound/Learner availability

RQ4: Proposed solutions to support dynamic group formation during the learning process

S2 affirms that re-composing group members at any moment leads to reach many educational and social objectives. However, the proposed system enables only destroying the entire group and form new one with new members, and no solution was proposed to partially change group members.

The system proposed in S7 allows each learner to join or leave his/her group at any moment as his/her wishes. For instance, when a learner changes his/her geographical location, the grouping algorithm provides him/her dynamically with information about the existing groups in the proximity, and she/he decides to join or not each group.

RQ5: Proposed LGF algorithm(s)

S1 is the only study that clearly describes the used algorithms for grouping learners. These algorithms are classified into two types: heterogeneous grouping algorithm with given threshold of difference, and heterogeneous grouping algorithm with given number of groups. Based on the instructor’s choice, the chosen algorithm transmits data from an U-portfolio to a similarity matrix (M) in order to calculate distances between learners’ behaviours and then create groups based on these distances. We present in Figure 4 the successive steps used by one of the two proposed algorithms to form appropriate groups.

Step 1. Assign the difference threshold T according to the empirical rule or the learning strategy design.

Step 2. Establish a triangular matrix M' as the difference matrix of M.

Step 3. Assign every learner to be an independent group by default.

Step 4. Select the element with the maximum value (the greatest difference value) from all of the elements in the matrix M'.

Step 5. If selected value (the maximum) is greater than the threshold T, find the two learners (row and column of matrix M') corresponding to the element value.

Step 6. If each of the found learners belongs to a different group, merge them into one group.

Step 7. Delete the maximum value element just selected. Repeat Steps 4-6 until a difference value greater than T cannot be found in the matrix M'.

Step 8. Based on the obtained heterogeneous grouping results, the number of people in each group may differ. Therefore, teachers can decide whether or not to set the difference threshold N_T of the number of people. If the difference between any two groups of people is greater than N_T, then teacher can execute a balance() function to conduct minor adjustments to balance the number of people.

Figure 4. Heterogeneous grouping algorithm with given difference threshold

S4 uses data traces collected from the study of learners' behaviours to train and test an intelligent system. The behavioural data should be transformed into input and output vectors to train IBL (Instance-based learning) and BayesNet (Bayesian Network) algorithms. These algorithms store the training data and use it later to predict the outputs at the arrival of a new contextual event.

S6 proposes an algorithm for recommending learning partners based on explored social networks. Through the analysis of collected data from learners' activities, the interactive type of each learner (hub or island type) is identified. The proposed algorithm uses the learners' interactive type in addition to some individual learning characteristics to assign learners to their appropriate groups.

S7 utilizes a grouping algorithm based principally on the geographical locations of learners. As grouping options, the algorithm could use two options: learning profiles together with learning styles of learners, or learners' learning interests.

RQ 6: Methods used for evaluating the learners grouping processes

S1 used both simulation and experimental evaluation methods. Simulation was used to evaluate the proposed approach by comparing its average intra-cluster diversity (AID) with the AID of groups created randomly or according to academic achievement. A higher value of AID implies greater heterogeneity, and greater heterogeneity implies better results of heterogeneous grouping. Experimental evaluation method was used to evaluate and compare the learners' behaviours. The learners are assigned to different groups using four grouping methods: random grouping; school achievement based grouping; grouping using the developed algorithm with given difference threshold; and grouping using the developed algorithm with given number of groups.

S2 uses an experimental evaluation method that compares pre-test and post-test results of a control group (formed using a random grouping) and three experimental groups (formed using three grouping criteria: preference, achievement, and sociability). The main objective of this experiment is to study the impact of different group reconfigurations on CL activities.

S4 proposes an evaluation method based on the use of same training datasets to create the learning model and to test the system. Second evaluation method uses only some parts of the training dataset to train the system, and other parts are used to test the system.

Discussion and recommendations for further research

Although mobile technologies facilitate gathering important information related to learning contexts (location, time, learners availability, learning objects availability, etc.), many considered studies (S2, S3, S8, S9, and S11) ignore completely the use of such context information. The learners' geographical location is almost the only context information used as LGF criterion.

Through this SLR, we remark the lack of approaches that combine the three kinds of criteria (learners' characteristics, learning behaviours, and context information) in a single process of LGF. Such a combination can make the grouping process more generic and adaptable to different learning contexts. Additionally, the grouping process should be customizable so that instructors can freely define the type, the number, and the weight of grouping criteria, together with different settings such as the number, the size and the type of learning groups. In such a way, instructor can customize the LGF process according to learning scenarios, type of learners, type of activities, needs, place, time, etc.

Many kinds of personal characteristics, learning behaviours, and context information have never been proposed as grouping criteria by the reviewed studies. Therefore, the following criteria are recommended to be considered in further approaches: Learner's health status (healthy or disabled); Level of communication with instructors; Level of interaction with learning objects; Level of interaction with the system; List of disliked partners; Preferred times of learning; Availability of learners; Movement patterns of learners; Learning progression rate of learners; weather status; and learners' agenda information.

Considering the greatest possible number of grouping criteria is a good solution to make the grouping process useful for any learning situation. However, selecting manually the proper grouping criteria could represent a difficult task. Therefore, developing new solutions that help instructors to know which criteria are appropriate to each situation is much requested. We propose, in this context, the use of machine learning algorithms that analyse the relationship between the past groups' outcomes and the used grouping criteria, to form the best possible groups.

The intra-session dynamic grouping represents an essential requirement in MCSCL. However, only two papers (S2 and S7) propose some solutions for supporting the dynamic grouping. Such a dynamic groping allows instructors to relocate at anytime the members of groups, or to form new groups depending on the task that is being performed in a given time and place. Additionally, it helps newly arrived learners, or learners who want to change their groups to quickly find new appropriate groups.

Another gap from this SLR is the lack of sharing the source codes of the implemented algorithms (except S1). This would prevent the MCSCL community to well compare, evaluate, and enhance the LGF tools and algorithms.

Through this SLR, we remark that many of reviewed studies did not evaluate their approaches. Some studies compared their automatic processes of LGF with manual or random methods (S1, S2). Some others used only questionnaires to evaluate the users' satisfactions and obtain their feedbacks (S3, S5 and S8). Other studies used simulation methods to assess their approaches (S1, S4, and S9).

Among the gaps extracted through this SLR, is the lack of studies that consider the group leadership. Defining a leader for each group does not only help instructor for well controlling and communicating with the groups, but also facilitate the management of mobile CL activities within the groups. The group's leader can define the role and location of each member; define the learning/working needs, objectives, and strategies; and find solutions to internal problems. Nevertheless, selecting the leader of each group represents another difficult task. This selection can be based on the analysis of the past learning outcomes and behaviours of learners.
Through this SLR, we remark that the majority of studies do not show how to obtain and measure the grouping criteria. We propose, therefore, a new mechanism for gathering, evaluating, and storing the values of grouping criteria (see Figure 5). Those criteria are classified into three groups:

- Learners' personal traits, which could be introduced by the learners through their mobile interfaces, and stored in a specific database.
- Context information, which could be obtained in real time from the learners' devices using some specific tools (e.g., GPS, RFID).
- Learners' behaviours (Communication with partners, interaction with the system, etc.). To continuously evaluate those behaviours, we propose installing a set of event log files on the mobile device of each learner. Those files are considered as temporary databases that store different data related to the learners' behaviours. The system can then examine those log files and extract the relevant behavioural information through a specific data extraction service. This extracted information should be stored in another active database (e.g., Oracle, MySql).

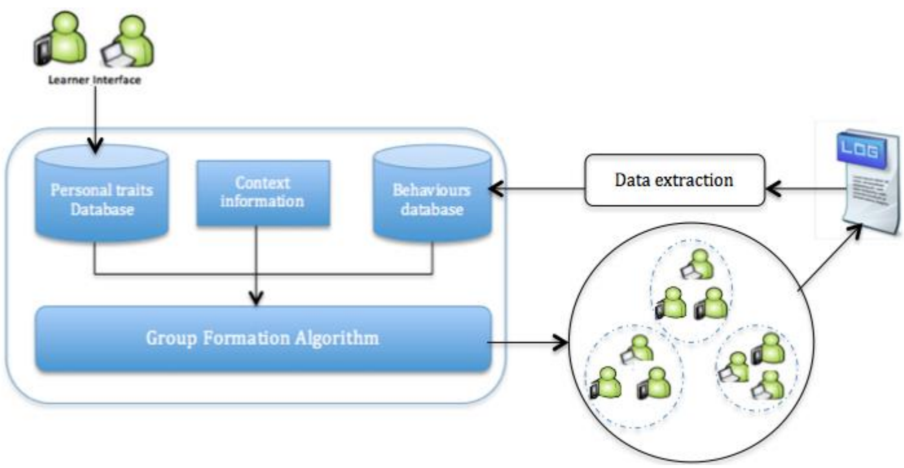

Figure 5. Proposed mechanism for evaluating and storing the group formation criteria

Limitations

The conducted SLR examines clearly the current state of research on the topic of LGF in MCSCL. However, some limitations have to be considered:

- This SLR addresses only the problem of LGF. Considering other related search problems, such as adaptive and personalised learning systems may provide other solutions for the LGF problem.
- This SLR focuses on the LGF problem in educational fields only. Considering other none-educational settings (e.g., formation of: business teams, sport teams, army troops) can provide more useful ideas for the topic of LGF.

Conclusions

This paper presents a Systematic Literature Review (SLR) on the problem of learning group formation (LGF) in Mobile Computer Supported Collaborative Learning (MCSCL) contexts. We believe that, through this SLR, we provided the MCSCL community a clear overview of the research on the LGF problem. In light of the extracted findings and research gaps, this SLR recommends the MCSCL community to: Develop new solutions for supporting intra-session dynamic grouping; Search new ways for making the grouping mechanism as useful as possible by considering the greatest possible number of grouping criteria; Propose new solutions to automatically select or recommend the relevant criteria that adapt to different learning aspects. Finally, and given the lack of considerable number of relevant approaches analysing the presented research problem, the MCSCL community are invited to pay more attention towards this issue by developing new contributions.

References

Boticki, I., Looi, C. K., & Wong, L. H. (2011). Supporting mobile collaborative activities through scaffolded flexible grouping. *Educational Technology & Society, 14*(3), 190-202.

Dillenbourg, P. (1999). What do you mean by collaborative learning? In P. Dillenbourg (Ed.), *Collaborative-learning: Cognitive and Computational Approaches* (pp.1-19). New York, NY: Elsevier.

Dillenbourg, P. (2002). Over-scripting CSCL: The Risks of blending, collaborative learning with instructional design. In P. Kirshner (Ed.), *Three Worlds of CSCL: Can we support CSCL?* (pp.61-91). Heerlen, The Netherlands: Open University of the Netherlands.

El-Bishouty, M. M., Ogata, H., Rahman, S., & Yano, Y. (2010). Social knowledge awareness map for computer supported ubiquitous learning environment. *Educational Technology & Society, 13*(4), 27-37.

Giemza, A., Manske, S., & Hoppe, H. U. (2013). Supporting the formation of informal learning groups in a heterogeneous information environment. In L. H. Wong et al. (Eds.), *Proceedings of the 21st International Conference on Computers in Education* (pp. 367-375). Bali, Indonesia: Asia-Pacific Society for Computers in Education.

Henri, F., & Lundgren-Cayrol, K. (2001). *Apprentissage collaboratif à distance: pour comprendre et concevoir les environnements d'apprentissage virtuels* [Remote Collaborative Learning: Understanding and designing virtual learning environments]. Sainte-Foy, Canada: Presse de l'Université du Québec (PUQ).

Hsieh, J. C., Chen, C. M., & Lin, H. F. (2010). Social interaction mining based on wireless sensor networks for promoting cooperative learning performance in classroom learning environment. In U. Hoppe et al. (Eds.), *The 6th IEEE international conference on wireless, mobile, and ubiquitous technologies in education* (pp. 219-221). Los Alamitos, CA: IEEE Computer Society.

Huang, J. J., Yang, S. J., Huang, Y. M., & Hsiao, I. Y. T. (2010). Social learning networks: Build mobile learning networks based on collaborative services. *Educational Technology & Society, 13*(3), 78-92.

Huang, Y. M., & Wu, T. T. (2011). A Systematic approach for learner group composition utilizing U-learning portfolio. *Educational Technology & Society, 14*(3), 102-117.

Huang, Y. M., Liao, Y. W., Huang, S. H., & Chen, H. C. (2014). A Jigsaw-based cooperative learning approach to improve learning outcomes for mobile situated learning. *Educational Technology & Society, 17*(1), 128-140.

Hübscher, R. (2010). Assigning students to groups using general and context-specific criteria. *IEEE Transactions on Learning Technologies, 3*(3), 178-189.

Kitchenham, B. A. (2007). *Guidelines for performing systematic literature reviews in software engineering* (Technical Report EBSE-2007-01). Keele, UK: Keele University and University of Durham.

Kitchenham, B. A., Mendes, E., & Travassos, G. H. (2007). Cross versus within-company cost estimation studies: A Systematic review. *IEEE Transactions on Software Engineering, 33*(5), 316-329.

Looi, C. K., Wong, L. H., & Song, Y. (2013). Mobile computer supported collaborative learning. In C. Hmelo-Silver et al. (Eds.), *The International Handbook of Collaborative* Learning (pp. 420-436). New York, NY: Routledge.

Messeguer, R., Medina, E., Royo, D., Navarro, L., & Juárez, J. P. (2010). Group prediction in collaborative learning. In Callaghan, V., & Kameas, A. (Eds.), *The 6th IEEE International Conference on Intelligent Environments* (pp. 350-355). Los Alamitos, CA: IEEE Computer Society.

Muehlenbrock, M. (2005). Formation of learning groups by using learner profiles and context information. In C. K., Looi et al. (Eds.), *Proceedings of the 2005 conference on Artificial Intelligence in Education: Supporting Learning through Intelligent and Socially Informed Technology* (pp. 507-514). Amsterdam, The Netherlands: IOS Press.

Mujkanovic, A., Lowe, D., & Willey, K. (2012). Adaptive group formation to promote desired behaviours. In M., Llewellyn, & S., Daniel (Eds.), *Profession of Engineering Education: Advancing Teaching, Research and Careers: 23rd Annual Conference of the Australasian Association for Engineering Education* (pp. 850). Melbourne, Australia: Engineers Australia.

Parsons, D., & Ryu, H. (2006). A Framework for assessing the quality of mobile learning. In R. Dawson et al. (Eds.), *Learning and Teaching Issues in Software Quality, Proceedings of the 11th International Conference for Process Improvement, Research and Education* (pp.17-27), Southampton, UK: Southampton Solent University.

Patten, B., Sánchez, I. A., & Tangney, B. (2006). Designing collaborative, constructionist and contextual applications for handheld devices. *Computers & Education, 46*(3), 294-308.

Tan, Q., Kinshuk, Jeng, Y. L., & Huang, Y. M. (2010). A Collaborative mobile virtual campus system based on location-based dynamic grouping. In J. Spector et al. (Eds.), *Proceedings of the 10th IEEE International Conference on Advanced Learning Technologies* (pp. 16-18). Los Alamitos, CA: IEEE Computer Society.

Webb, N. M., Nemer, K. M., Chizhik, A. W., & Sugrue, B. (1998). Equity issues in collaborative group assessment: Group composition and performance. *American Educational Research Journal, 35*(4), 607-651.

World-Telecommunication. (2014). Retrieved from http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx

Yang, F., Wang, M., Shen, R., & Han, P. (2007). Community-organizing agent: An Artificial intelligent system for building learning communities among large numbers of learners. *Computers & Education, 49*(2), 131-147.

Yatani, K., Onuma, M., Sugimoto, M., & Kusunoki, F. (2004). Musex: A System for supporting children's collaborative learning in a museum with PDAs. *Systems and Computers in Japan, 35*(14), 54-63.

Yin, C., Dong, Y., Tabata, Y., & Ogata, H. (2012). Recommendation of helpers based on personal connections in mobile learning. In Kinshuk et al. (Eds.), *Seventh IEEE International Conference on Wireless, Mobile and Ubiquitous Technology in Education* (pp. 137-141). Los Alamitos, CA: IEEE computer society.

Zurita, G., Nussbaum, M., & Salinas, R. (2005). Dynamic grouping in collaborative learning supported by wireless handhelds. *Educational Technology & Society, 8*(3), 149-161.