CrossMark

ORIGINAL RESEARCH

Group Formation Techniques in Computer-Supported Collaborative Learning: A Systematic Literature Review

Naseebah Maqtary[1] (iD) · **Abdulqader Mohsen[1]** ·
Kamal Bechkoum[2]

Abstract Group formation is an essential process for group development lifecycle. It has been a growing concern to many researchers to be applied automatically in collaborative learning contexts. Forming a group is an atomic process that is affected by various factors. These factors differ depending on the group members characteristics, the context of the grouping process and the techniques used to form the group(s). This paper surveys the recently published work in group formation process providing a systematic literature review in which 30 relevant studies were analyzed. The findings of this review propose two taxonomies. The first one is for the attributes of group formation while the second is for the grouping techniques. Furthermore, we present the main findings and highlight the limitations of existing approaches in computer supported collaborative learning environment. We suggest some potential directions for future research with group formation process in both theoretical and practical aspects. In addition, We emphasize other improvements that may be inter-related with other computing areas such as cloud computing and mobility.

Keywords Group development · Attributes of group formation · Group formation techniques · Computer supported collaborative learning (CSCL) · Systematic literature review (SLR)

✉ Naseebah Maqtary
n.almaqtari@ust.edu

Abdulqader Mohsen
a.alabadi@ust.edu

Kamal Bechkoum
kbechkoum@glos.ac.uk

[1] University of Science and Technology, Sanaa, Yemen

[2] University of Gloucestershire, Gloucester, UK

1 Introduction

Education has improved smoothly through developing various approaches and technologies (Resta and Laferrière 2007; Stahl et al. 2006). It has been upgraded from the individual learning paradigm to collaborative learning. With this upgraded paradigm learners can gain more knowledge and skills through learning together from the same learning situation (Matazi et al. 2014; Resta and Laferrière 2007; Srba and Bielikova 2015; Stahl et al. 2006).

Collaborative learning is defined by Rowe et al. (2010) as an instructional method that is used by a group of learners to achieve a common goal. This type of learning is conceivably executed through a three dimensional model. This model includes the following axes: (1) a group of people either in pairs or more; (2) a credible material of learning, (i.e. course content, activity and lifelong work experience); and (3) a way of learning through collaborative interaction among group members (Dillenbourg 1999).

The environment of collaborative learning is either real or virtual (Dillenbourg 1999). Collaborative learning is performed through face to face conversations and meetings or online using computer tools and frameworks (Dillenbourg 1999; Resta and Laferrière 2007; Stahl et al. 2006). An example of such tools is computer-supported collaborative learning (CSCL) (Matazi et al. 2014; Rowe et al. 2010; Srba and Bielikova 2015; Stahl et al. 2006). CSCL is a pedagogical approach that uses networking technologies to aid the social and instructional interaction among learners in small groups and learning communities (Resta and Laferrière 2007; Rowe et al. 2010; Stahl et al. 2006). It employs generic tools such as e-mail, file attachments, electronic bulletin boards, chat, blogs, and digital audio and videoconferencing systems. Furthermore, it uses specific tools such as asynchronous/synchronous communication tools of Web-based Instructional Management Systems (Course Management System, CMS; Learning Management System, LMS), and virtual learning environments (Blackboard/WebCT, Moodle, Sakai, Claroline, FirstClass) (Resta and Laferrière 2007; Stahl et al. 2006).

CSCL has emerged during the mid-1990s. As shown previously, various tools have been used and employed to merge collaboration within educational activities (Stahl et al. 2006). Focusing on collaborative learning has brought groupwork to the fore. Many studies in CSCL environment have been carried out on administrating groupwork activities like group formation (GF), monitoring and evaluation (Sun and Shen 2013).

Forming a group that collaboratively learns is one of the most challenging tasks in CSCLs context. This topic attracted the interest of several researchers (Amara et al. 2016; Khandaker et al. 2006; Srba and Bielikova 2015). Many articles discussed group formation from different aspects. These aspects mostly focused on group development life cycle (Abnar et al. 2012; Sun and Shen 2013), optimized the process of group formation (Ho et al. 2009; Zheng and Pinkwart 2014), and discovered the attributes that optimally affect group formation (Graf and Bekele 2006; Yannibelli and Amandi 2011).

1.1 Research Objectives, Questions and Structure

Notwithstanding the valuable contributions thus far, there is still no rounded overview of the group formation process with its various effective components like attributes and techniques. In addition, no surveys were found on group formation process. This led us to hold a systematic literature review (SLR) about group formation. This review aims at summarizing the various previous work and reproducing these contributions in an organized manner. This has been done by using a systematic approach as discussed in the Sect.

2. Our contribution will present the techniques and attributes of group formation with different classified views. The attributes of group formation are different. They are chosen differently according to the grouping context. In addition, computerized tools perform the group formation process based on different techniques. These techniques are examined in specific grouping environments. Thus, this paper attempts to:

1. Discover the recent contributions in group formation in CSCL contexts.
2. explore the most effective attributes and techniques in grouping process from different viewpoints.
3. Summarize and represent the findings in a structural manner.
4. Draw the knowledge gaps, challenges and opportunities of group formation.

To achieve these objectives we have formulated some research questions and synthesized the relevant studies to answer them through conducting SLR. These questions are as it follows:

- *RQ1* What are the most effective attributes and techniques used in group formation process?
- *RQ2* How can the recent CSCL contributions be represented in group formation process within educational contexts?
- *RQ3* What are the knowledge gaps and limitations in group formation?
- *RQ4* What conclusions can be drawn from the existing studies?

This paper is organized around the following subsections: firstly, methodology of investigating this review is discussed in Sect. 2. Then our findings are introduced in Sect. 3. After that, Sect. 4 presents discussion and encountered limitations with future trends. Finally, conclusion is presented in Sect. 5.
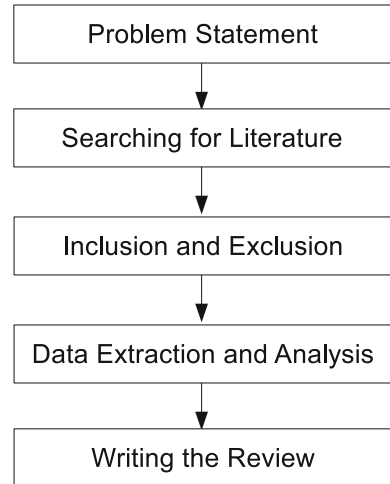
2 Research Methodology

To conduct the literature review, a systematic approach was proposed following the straightforward and simple logic described on Okoli and Schabram (2010). This approach is used to construct a systematic literature review (SLR). It consists of five sequential steps which as illustrated in Fig. 1. Results of the review are presented in Fig. 2.

2.1 Problem Statement

This stage is concerned with two issues: 1. specifying the review aim and 2. assigning the work protocol. These issues are discussed as follows.

a. Specifying the purpose of the literature review
   The main purpose of the study is to conduct a systematic literature review on group formation to contribute to the body of knowledge in such a field.
b. Assigning the protocol of work
   Assigning the work protocol deals with formulating the research question which leads the researchers to set the research objectives. It also comes up with the plan (work protocol) which organizes the methodology of accomplishing the review. As this review aims at conducting a systematic literature review on group formation process, we have formulated the research questions. These questions were specified in Sect. 1. The plan drawn by researchers begins with specifying the review aim, sketching the

Fig. 1 The main steps of the
review

```
┌─────────────────────────────────┐
│        Problem Statement        │
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│     Searching for Literature    │
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│     Inclusion and Exclusion     │
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│   Data Extraction and Analysis  │
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│        Writing the Review       │
└─────────────────────────────────┘
```

necessary steps of conducting the review and finalizing the output of the review to be ready for publication while setting the protocol, the researchers focus on the following issues:

- Concentrating on group formation process in CSCL systems.
- Searching the recent related contributions in various journals reviews, conferences papers, theses and dissertations specifically published between 2005 and 2015.

2.2 Searching the Literature

In this review of literature, the researchers investigated more than 110 studies. They covered various available sources i.e. scholar search engines, reputable journals and

Searching the literature review
- Using various resources
- Ending with ≈ 110 studies

Practical screening
- Using relative keywords
- Ending with ≈ 30 studies

Quality appraisal
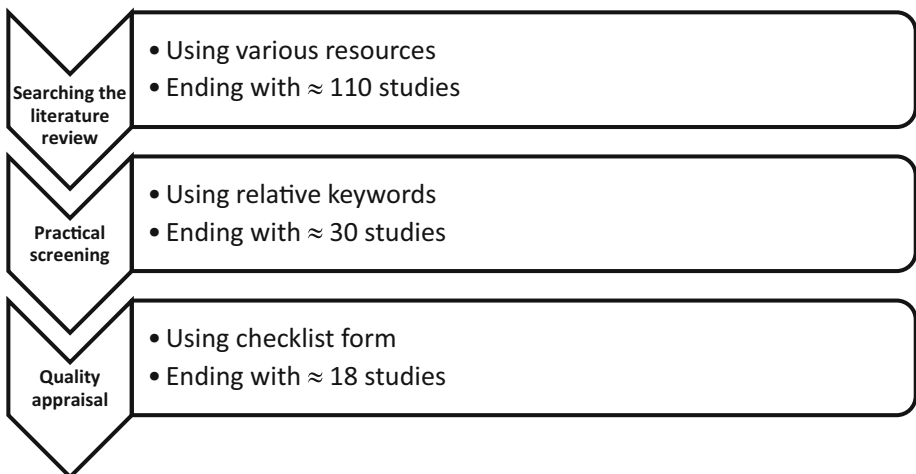- Using checklist form
- Ending with ≈ 18 studies

Fig. 2 Results of review

conference proceedings as shown in Table 1. These studies varied in type and scope. The review involved journal reviews, conference papers, theses and book chapters.

Statistically, conference papers are the most used source in this review. They are almost 58% of the total used sources while journal articles and reviews are approximately 40%. Book chapters are the least used sources which represent only 2% as illustrated in Fig. 3. The researchers used many keywords for searching such as collaborative learning, computer supported collaborative learning, group, group formation, coalition formation and group development.

2.3 Inclusion and Exclusion Processes

Collected studies were filtered using multi-level criteria. These criteria were either for inclusion or exclusion as shown in Table 2. The inclusion process is called practical screening while exclusion process is called quality appraisal. These two processes are detailed as follows.

Table 1 Main sources of the review

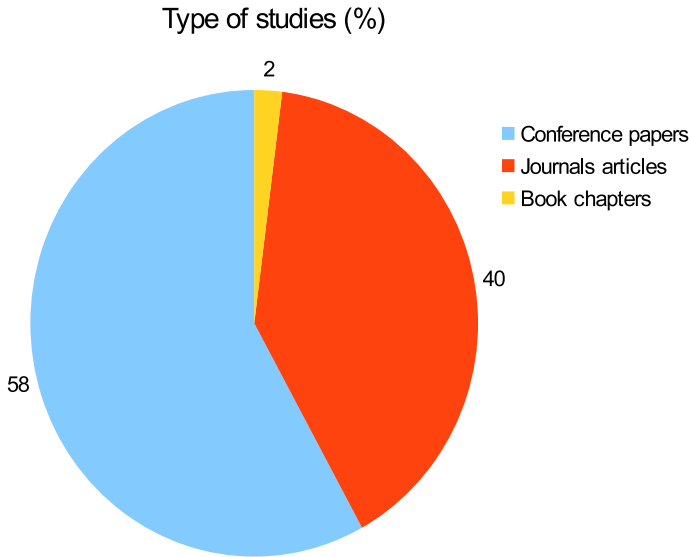| Conferences	Journals	Others
Adaptive hypermedia and collaborative web-based systems	Artificial Intelligence in Education	ACM digital library
Advanced learning technologies collaborative-learning: cognitive and computational approaches	Cambridge Handbook of the Learning Sciences Advances in Web-Based Learning Computer Science Research and Application	The Learning and Skills Research Centre
Autonomous agents and multiagent systems	Computers and Education	ScienceDirect
Computational intelligence and informatics	Creativity and Collaborative Learning	Elsiver
Computational science and engineering	Educational Psychology Review	IEEE Explore
Computer science and information engineering	Educational Technology and Society	
Forming and maintaining coalitions and teams in adaptive multiagent systems	Frontiers in Artificial Intelligence and Applications	
Information Society	Group and Organization Management	
Intelligent systems: theories and applications	Human Resource Development International	
Intelligent tutoring systems	JALT CALL Journal	
Interactive Collaborative Learning	Learning Technologies	
Recent advances in information science: European conference of computer science	Student Centered Learning	
Recent trends in information technology		
Supporting group work: ACM		
Technology enhanced education		
Tools with artificial intelligence		
User modeling		
Web intelligence and intelligent agent technology		

Type of studies (%)

Fig. 3 Types of sources used in the review

a. Practical Screening

   This step dealt with narrowing down the range of studies. This was done by reading each research abstract to ensure the availability strong relationship to the review aims and questions. Also, studies were screened through choosing more relative keywords. As a result, the involved studies in the review were only those related to group formation process in CSCL environment (i.e. 30 studies).

b. Quality Appraisal

   After specifying the related studies, another sorting level was executed to ensure the strongly correlated studies to the proposed review. This subtask used the exclusion filter on chosen studies. The task was fulfilled through answering the questions in the checklist. The checklist consists of seven questions as shown in Fig. 4. This stage ended up with 18 studies that positively answer the proposed checklist form.

2.4 Data Extraction and Analysis

The data required to build the review of group formation were identified to include type of grouping, number and type of grouping attributes, used technique and special features of grouping. This information was gathered from 18 studies in a tabular form to help in

Table 2 Inclusion and exclusion criteria

Inclusion	Exclusion
Papers published between 2005 and 2015	Nonacademic papers and gray literature such as reports, technical reports, and working papers
Using relative keywords	Papers with weak analysis and writing
Prestige and academic papers and articles	Using checklist form

No.	Question	Yes	No
1	Does the study apply the group formation in educational context?		
2	Does the study use a scientific approach in proving its contribution?		
3	Is the type of the group formation defined? (homogenous, heterogeneous or mixed)?		
4	Does the study identify specific attributes for group formation?		
5	Is the used technique clear and approved through experiments?		
6	Does the study use suitable datasets to prove the experiments' results?		
7	Does the study add valuable contribution to the state of the art of group formation?		

Fig. 4 Checklist form used to analyze the relevant studies in SLR on group formation

exploring group formation process as presented in Table 3. Then, gathered data were synthesized using qualitative analysis, quantitative analysis or both. In the review which was based on both quantitatively and qualitatively analysis of extracted data, taxonomic maps were used to reorder the ideas and extracted data from the group formation studies. This stage ended up with two taxonomies which were discussed in details in Sect. 3.

2.5 Writing the Review

The final stage in conducting systematic literature review was writing the review, the adopted methodology, results and explanations. The findings of this paper are also reported based on a systematic approach. Next sections discuss the contributions found in group formation studies in CSCLs. These sections concentrate on the techniques of grouping and the effective attributes.

3 Findings

After reviewing more than 110 studies using a scientific approach called systematic literature review, the final analysis only involved 18 studies. In this approach, many steps were followed gradually filter the studies and determine the most related ones to the research objectives. The outcomes of SLR phases are presented in Table 4. This section presents the findings of group formation studies in CSCL contexts. These findings are organized in a structural manner. Thus, group formation process is introduced and discussed with its effective attributes. Also, the relevant literature is presented in a historical order so that most remarkable features and contributions are highlighted. After that, two taxonomies are illustrated and discussed to represent group formation attributes and techniques, respectively.

3.1 Group Formation

As mentioned above, various research studies were conducted to explore new provisions in group formation. These studies attempted to ensure that all group members are smoothly

Table 3 Overview of extracted data from relevant studies in group formation

Author and publication year	GF attributes					No. of attributes	Techniques	Contributions
	Knowledge	Learning style	Social interaction	Team role	Personality traits			
Graf and Bekele (2006)	✓				✓	2	Ant colony algorithm	Heterogeneous groups Introducing the measure called Goodness of Heterogeneity
Christodoulopoulos and Papanikolaou (2007)	✓	✓				2	Fuzzy C-means Random selection algorithm	Homogeneous and heterogeneous groups Manual instructor interference – Equality is provided in group size
Soh and Khandaker (2007)	Various					–	Multi-agents	Introduces I-MINDS system which is infrastructure of collaborative learning Models the group formation problem as rules and policies
Ounnas et al. (2008a)			✓	✓	✓	3	Semantic Web Ontologies	Introduces a framework of group formation based on teacher preferences Solves the problem of orphan student
Ho et al. (2009)	✓	✓	✓			3	Particle swarm optimization	Using student feature model to model attributes
Rubens et al. (2009)	Informal (non-structured) data					–	Mash-up technique	Proposing a data-driven automatic group formation model and corresponding methods for mash-up, mining and inference utilizing data originating from multiple data sources
Hubscher (2010)	Various					–	Tabu search	It forms groups after instructor assignment of the desired attributes
Yannibelli and Amandi (2011)				✓		1	Knowledge base evolutionary algorithm	Introducing different models of team roles, especially Belbins team role First research dealing with team role
Abnar et al. (2012)		✓				1	Genetic algorithm Greedy algorithm	Homogeneous and heterogeneous grouping Mixed groups are formed with equal size Iterative process for satisfying the tutor about grouping

Table 3 continued

Author and publication year	GF attributes					No. of attributes	Techniques	Contributions
	Knowledge	Learning style	Social interaction	Team role	Personality traits			
Mujkanovic et al. (2012)	✓				✓	1	Regression analysis optimization	Grouping is based on idea of no prior rules or data about members because they are from different environments
Using updating rules to adapt the approach of forming groups
Brauer and Schmidt (2012)	✓	✓	✓			3	Genetic algorithm	It concentrated on forming group after gathering attributes from OSN
Moreno et al. (2012)	Various					–	Genetic algorithm	Formulating group problem into multi-objective optimization problem
Inter-homogeneous and intra-heterogeneous groups
It compares between 3 algorithms (random, exhaustive and proposed method)
Tien et al. (2013)	Various					–	Genetic algorithm with TOPSIS technique	Inter-homogeneous and intra-heterogeneous groups
It compares between 3 algorithms (random, genetic and proposed method)
Jozan and Taghiyareh (2013)	Various					–	Genetic algorithm	They used inversion concept to represent the priority of attributes of group members
Srba and Bielikova (2015)	✓					1	Clustering with applying the theory of group technology	Dynamic group with short term and iterative formation
Applying group technology
Zheng and Pinkwart (2014)					✓	1	Discrete particle swarm optimization	Heterogeneous groups
It compares between 3 algorithms (exhaustive method, random selection and DPSO based on time complexity and stability)
Amara et al. (2016)	Various					–	K-means	Held the whole process of groupwork
Dynamic and customized group formation through MCSCL

Table 4 Outcomes of SLR phases

Phase	Outcomes
Reviewing literature	110
Practical screening	30
Quality appraisal	18

and easily achieving the learning outcomes. (Khandaker et al. 2006). Group formation is the first process of the group development life cycle in which efforts should be devoted to ensure effectiveness and efficiency of the process (Bonebright 2010). The group development life cycle is divided into many phases that translate the process of forming and monitoring the performance of the group. There are various models and theories about group development. The most known model that expresses group development life cycle is Tuckmans model, which has been frequently reviewed and extended by researchers. One example of these extensions is the Tuckman and Jensens model. This model divides group development life cycle into five stages: forming, storming, norming, performing and adjourning (Bonebright 2010; Srba and Bielikova 2015; Tuckman and Jensen 1977). The first stage is forming the group. It is concerned with introducing the backgrounds, preferences and experience of each group member to each other to form the first impression. In this, the group leader should be aware about group goals, roles and responsibilities to clarify them for the members (Bonebright 2010). The next stage, storming, is dealing with setting the rules of group management to minimize, or avoid, conflict. The third stage, norming, is the phase during which agreement is reached about how members work together to maximize group performance and achievement (Bonebright 2010). The performing stage is about the group functioning towards the stated goals, which are finally evaluated in the adjourning stage (Bonebright 2010).

As stated in the methodology, 18 studies were selected to explore the contributions in group formation process. These contributions are briefly discussed in Sect. 3.2.

3.2 Related Literature

This section reviews the state of the art in group formation process. It is written in a historical order to summarize the contributions of each study and its technique. It gives a brief discussion on how this technique was executed within the context of group formation. It also shows the attributes used as identified in Table 3.

Graf and Bekele used Ant colony optimization algorithm (ACO) to build heterogeneous groups. These groups were built according to members performance and personality traits. Researchers introduced a quality parameter called goodness of heterogeneity (GH) to measure the level of heterogeneity of the group members. They proved the scalability of the strategy employed by iterative experiments with different group sizes (Graf and Bekele 2006).

In Christodoulopoulos and Papanikolaou (2007), the authors implemented a web-based group formation tool. This tool has the ability of grouping members homogenously and heterogeneously. Researchers used a fuzzy c-Means algorithm to homogeneously assign a member to the most appropriate group. The appropriate group was selected by looking at the different probabilities of members belonging to different groups. In addition, the proposed strategy used the random selection algorithm in order to heterogeneously create groups. Knowledge level and learning style of the members were used as criteria for the grouping process.

Soh and khandaker presented a multi agent framework for group formation of students. They implemented VALCAM, an algorithm that groups students based on the idea of iterative auctions. This framework was applied in a computer supported collaborative learning environment called I-MINDS (Soh and Khandaker 2007). I-MINDS is a distributed computing infrastructure that uses intelligent multiagent information system for education (Soh 2004; Soh et al. 2008). It reforms the problem of group formation through using rules and policies. These rules and policies are realized by agents while the process of group formation is executed (Soh 2004; Soh and Khandaker 2007).

Another approach of group formation was presented by Ounnas et al. (2007b) who modeled a semantic framework to represent the interaction data of learners through using FOAF ontology. Moreover, researchers introduced the term orphan student which means the left or unassigned student to a group (Ounnas et al. 2008b). They used web semantic technologies and logic programming (Ounnas et al. 2008a).

Particle swarm optimization algorithm was the technique used in Ho et al. (2009). Researchers chose social interaction, competences and learning style attributes to form the groups. They involved the time complexity of the problem that considers more than one attribute while forming groups.

Some contexts of collaborative learning in informal environments lack prior information about learners. This issue attracted Rubens and his research team to propose a data-driven model for extracting information of learners from various data sources such as blogs, wikis, and forums. The extracted information was built in a mash-up way that led to automatic group formation through grouping the learners with sharable knowledge (Rubens et al. 2009).

Yannibelli and Amandi applied the evolutionary algorithm to form groups based on the attributes of teams roles. To form a well-balanced group with various team roles, the group formation process should be heterogeneous. Heterogeneity raised the time complexity of the problem, which was solved by using this algorithm. Researchers evaluated the result of formation process by ensuring that each outcome group has highly diverse team roles (Yannibelli and Amandi 2011).

Brauer and Schmidt proposed another approach to capture data about members attributes. They developed a graph model for modeling members data from online social networks (OSN). The data included attributes such as knowledge, learning style and social interaction. They used various graph traversal algorithms to capture the candidate members of groups. These candidate members were grouped through using a genetic algorithm which handles the group formation process execution and scalability (Brauer and Schmidt 2012).

Genetic algorithms were also used to form heterogeneous groups in Sukstrienwong (2012). The researcher modeled a fitness function with fairness and equity in terms of members performance to ensure the fair formation, which means that each group has various knowledge levels of the members.

Moreno with his group suggested using genetic algorithms to form groups with multiple attributes. They formulated the grouping problem into multi-objective optimization problem under combinatorial scenario. To validate the result, comparative study was done in contrast with exhaustive and random algorithms. They stated that their proposed strategy has contributed better results in both computational and pedagogical directions (Moreno et al. 2012).

A different approach was applied by Abnar et al. (2012) where groups were formed iteratively to reach the continuously tuned fitness threshold of the genetic algorithm. This approach was featured by the flexible facility of using different attributes with ranking

(prioritizing) them according to the group task. Same feature was handled in Hubscher (2010) so that group formation was iteratively adaptable to the context of the used strategy in teaching and learning. A new criterion called evenly skilled, which depends on reciprocal teaching method, was proposed to assign students to groups. Researcher employed the tabu search algorithm to form groups because of its stable and systematic use of memory.

The situation of forming groups that have members who are geographically distributed and no prior rules are known about them were investigated in Mujkanovic et al. (2012). The research approach used self-learning algorithm, called regression analysis optimization that adapted rules of forming groups over time yielding a progressively improving group performance.

Hui-Wen Tien and his research group examined the effectiveness of group formation process through proposing new strategy that adopted genetic algorithm with TOPSIS technique. They formulated the fitness value to achieve the goal of obtaining inter-homogeneous and intra-heterogeneous groups. The authors compared between the random algorithm, genetic algorithms and the proposed strategy, they concluded that the proposed method recorded better grouping results than random and genetic algorithms against different number of characteristics (Tien et al. 2013).

A similar approach was employed in Jozan and Taghiyareh (2013) where a genetic algorithm was reapplied to group formation process. Researchers also applied the idea of priority and weights of the members attributes with the concept of inversion. They evaluated the quality of the formed groups through inter-group fitness and intra-group fitness measures. The concluded remarks stated that the proposed strategy returned better results in group formation process with inter-group fitness criteria while it showed weak performance with intra-group fitness criteria.

Discrete particle swarm optimization algorithm (DPSO) was used to group heterogeneous learners according to their personality traits and gender in Zheng and Pinkwart (2014). It was also evaluated in contrast with exhaustive and random algorithms. The researchers argued that DPSO algorithm gave better performance and stability results in a reasonable time than other evaluated algorithms for group formation.

In Srba and Bielikova (2015), a group was formed through applying group technology (GT). It is a concept in manufacturing and engineering management sector, whereby its most applied task is cell manufacturing. Researchers proposed a novel method by using clustering algorithms to involve the collaboration feedback of the members dynamically and iteratively each time the group was formed. They simplified the group lifecycle to suit the short-term groups in virtual domains.

Recently, a research was conducted to form homogeneous groups in mobile collaborative learning environment (MCSCL) (Amara et al. 2016). It held all the activities of groupwork. It also added the dynamism of the group formation at any level of groupwork and customized selection of forming attributes. The research enabled the instructors to determine type, number and weight of grouping criteria. The technique used for group formation was K-means algorithm.

From the above discussion and literature, it was noticed that many of the issues raised in group formation process and still need further investigations. For example, the formation process was conducted under various contexts to cover all collaborative learning aspects. The group formation can be characterized by different parameters such as the group size, duration of the groupwork, the ideal method of grouping, the authority of grouping and/or type of the formed groups. In addition, attention is needed to the educational and psychological characteristics of members involved in the grouping. These characteristics vary

from one group to another if the group goal and tasks are different. In some of the previous work, groups were formed using member characteristics such as knowledge, skills and competences while other work grouping were based on learning styles, personality traits and other characteristics. These characteristics are discussed in details and classified within a proposed taxonomy in Sect. 3.3.

3.3 Taxonomy of Group Formation Attributes

The group formation process can be achieved either manually or automatically (Srba and Bielikova 2015). Manual formation of the group is either self-selection or instructor assignment (Resta and Laferrière 2007; Srba and Bielikova 2015; Ounnas et al. 2007a). In the self-selection approach the member has the right to choose the most suitable group for him/her. This approach does not guarantee a balanced grouping and thus violates the ideal group formation (Abnar et al. 2012; Zheng and Pinkwart 2014). The second approach is managed by the instructor decisions about which member will form part of which group (Srba and Bielikova 2015). This kind of selection guarantees better results, in terms of a balanced grouping, but it is a fairly complex process when large numbers of members are grouped manually (Mujkanovic et al. 2012; Srba and Bielikova 2015). In order to assign members to groups automatically, there exist many CSCLs environments that offer the option of creating groups automatically with or without human intervention (Abnar et al. 2012). Random selection is one way of achieving group formation automatically (Srba and Bielikova 2015). Other approaches are used to form groups according to the context.

During the formation of groups, various attributes should be taken into consideration to ensure that groups will achieve their goals (Coffield et al. 2004). Thus, the attributes can be categorized into two classes: member attributes and group attributes. Member attributes are the attributes that describe the person who will be included in groups while the group attributes describe the group characteristics as a whole.

3.3.1 Member Attributes

Examples of member attributes are knowledge, skills, learning styles and personality traits. These attributes are used to decide the most suitable group for each member (Abnar et al. 2012; Graf and Bekele 2006). Other member attributes such as social interaction and team roles are also considered in some situations of group formation (Yannibelli and Amandi 2011).

Different studies are conducted to tackle grouping students for a specific task or assignment based on their knowledge and skills. The attributes are measured through Likert scale stated by the instructor himself/herself according to his/her knowledge about students (Graf and Bekele 2006; Ho et al. 2009). Attributes can also be collected from different learning management systems which store student information and their academic progress (Brauer and Schmidt 2012).

Learning styles or personality traits are arbitrarily used in different works to check the effectiveness of group formation (Abnar et al. 2012; Brauer and Schmidt 2012; Christo-doulopoulos and Papanikolaou 2007; Ho et al. 2009; Martin and Paredes 2004). There is a strong intuitive appeal to consider learning styles as an indicator of the speed, manner and confidence of picking information and data (Coffield et al. 2004; Martin and Paredes 2004). There are various models of learning styles such as Kolbs learning style Inventory (LSI),Herrmann whole brain model (HDBI) and Myers–Briggs Type Indicator (MBTI) (Coffield et al. 2004). To capture the learning style of a group member, a questionnaire is

run and then an index is used to specify the learning style of the member after classifying his responses to the questionnaire (Coffield et al. 2004; Martin and Paredes 2004).

Some researchers used the personal information/traits of the group members, such as age, gender, IQ and race, as an attribute that affects the group formation process (Graf and Bekele 2006; Mujkanovic et al. 2012; Ounnas et al. 2008b; Zheng and Pinkwart 2014).

Social interaction is another attribute that attracts attentions of psychological and educational researchers. Social interaction and negotiation contribute to the way in which people learn how to develop shared understanding about certain concepts or tasks (Kreijns et al. 2002). The attribute of social interaction consists of different social skills that learners should have while working collaboratively within the group. These social skills are participation, social grounding, active learning conversation skills, performance analysis and group processing and promotive interaction (Soller 2001).

Members within a group should play different roles according to the mission of the group and their behavior. A role is the way a person is expected to behave, contribute and interrelate with others throughout collaborative work. Several team role models were proposed and investigated in the literature about group formation. These models are based on the concept of well-balanced groups, which should be formed with members having heterogeneous roles. The most known team role model is Belbins which is applied in training activities by different organizations, consulting firms and executive education programs (Yannibelli and Amandi 2011).

3.3.2 Group Attributes

The group attributes are related to the context of the group goal and task. For example, the nature of the task determines the homogeneity of the group members. Accomplishing some tasks need homogeneous characteristics of the group members while other tasks need diverse characteristics of group members in order to complete the tasks (Christodoulopoulos and Papanikolaou 2007; Srba and Bielikova 2015). In addition, the duration of completing a group task is another attribute which affects the group formation process. Thus, there are short term and long term groups (Huang et al. 2009; Srba and Bielikova 2015). Moreover, the process of assigning members can be static for the duration of the task. it can benefit from previous information about group members and their abilities in accomplishing group tasks. The latter type of formation is called dynamic or adaptive formation (Mujkanovic et al. 2012; Srba and Bielikova 2015).

All the aforementioned review of the group formation attributes were reorganized and classified within a proposed taxonomy that reflects the above presentation. This taxanomy is depicted in Fig. 5, which has a multi-level categorization. Attributes are divided into two groups: (1) member attributes and (2) group attributes. Member attributes are categorized into five different attributes. Group attributes are also classified into four subcategories which are: (1) assignment method, (2) homogeneity of the group members' characteristics, (3) group duration and (4) adaptability of the group. Also a tabular mapping is used to map each study with its specified grouping attributes to facilitate grasping information about studies contributions. This mapping is presented in Table 3.

3.4 Taxonomy of Group Formation Techniques

The literature on group formation is quite rich. The publications range from gathering data of group members to applying various models on group formation process with different
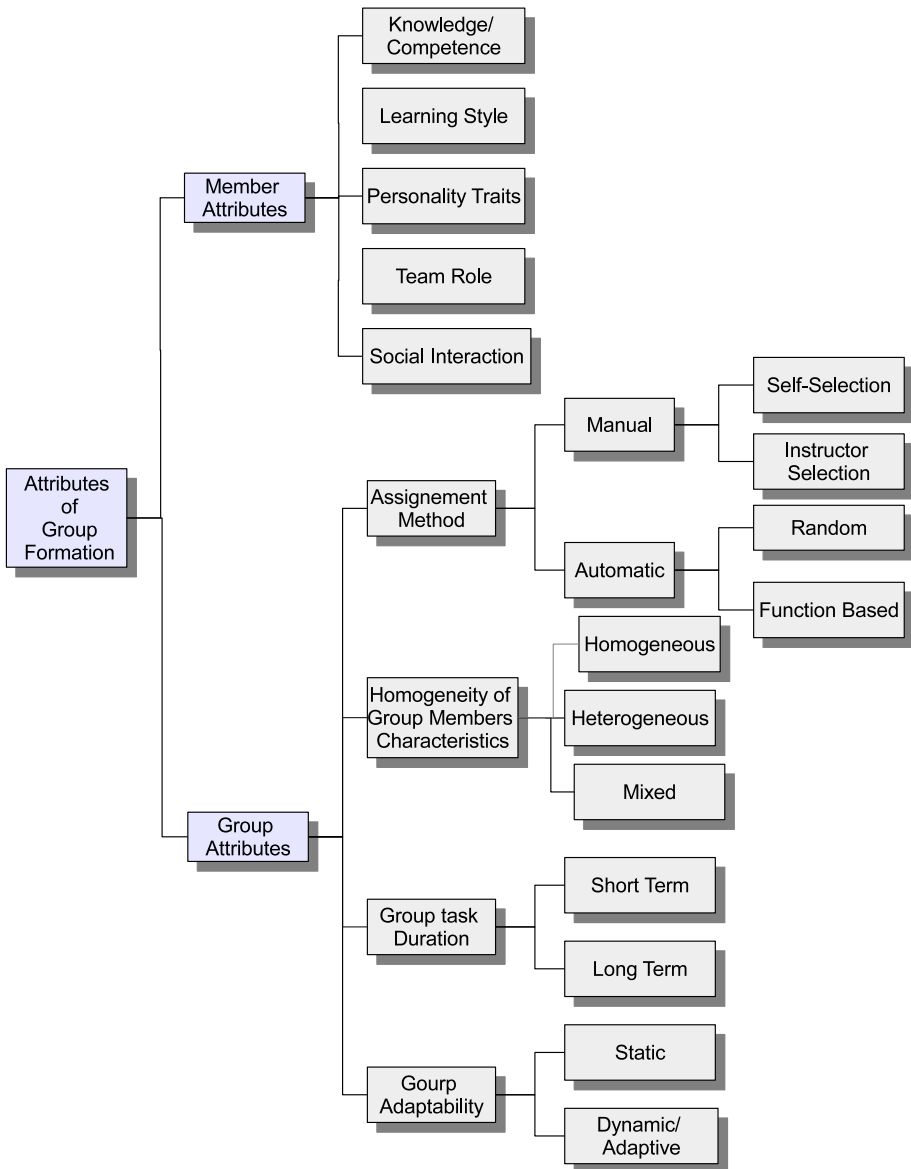
Fig. 5 Taxonomy of group formation attributes

perspectives in various contexts. This section highlights the distinction between reviewed studies under different points of view as illustrated in the taxonomy shown in Fig. 6.

Research studies made by Brauer and Schmidt (2012), Rubens et al. (2009) employed data-driven models to collect data about learners from different environments such as online social networks (OSN). In almost all the work reviewed, the group formation process is discussed in various circumstances.

Some research studies depended on either single attribute to form a collaborative group e.g., (Abnar et al. 2012; Mujkanovic et al. 2012; Srba and Bielikova 2015; Sukstrienwong 2012; Yannibelli and Amandi 2011; Zheng and Pinkwart 2014) or multiple attributes e.g., (Brauer and Schmidt 2012; Christodoulopoulos and Papanikolaou 2007; Graf and Bekele 2006; Ho et al. 2009; Ounnas et al. 2008a). It is obvious that more attributes of group members, make the process of group formation more complex.

In addition, various techniques were used to prove the experiments of forming groups. these techniques include evolutionary approach (Abnar et al. 2012; Brauer and Schmidt 2012; Jozan and Taghiyareh 2013; Moreno et al. 2012; Sukstrienwong 2012; Tien et al. 2013; Yannibelli and Amandi 2011), swarm techniques (Graf and Bekele 2006; Ho et al. 2009; Zheng and Pinkwart 2014), clustering algorithms (Amara et al. 2016; Christodoulopoulos and Papanikolaou 2007; Srba and Bielikova 2015), semantic ontologies (Ounnas et al. 2008b) and multi-agent (Soh 2004).

Moreover, a wide view of recent research in group formation shows that different approaches were applied to form effective groups. These types of research can be categorized into three categories based on the research goal:

- Finding out suitable attributes that affect group formation and achievement (Yannibelli and Amandi 2011).
- Trying to optimize the process of group formation through using optimization techniques so that better and faster formation will be achieved (Abnar et al. 2012; Graf and Bekele 2006; Jozan and Taghiyareh 2013; Zheng and Pinkwart 2014).
- Applying new technologies from different scientific area, for example, using the manufacturing theory called group technology, GT (Srba and Bielikova 2015).
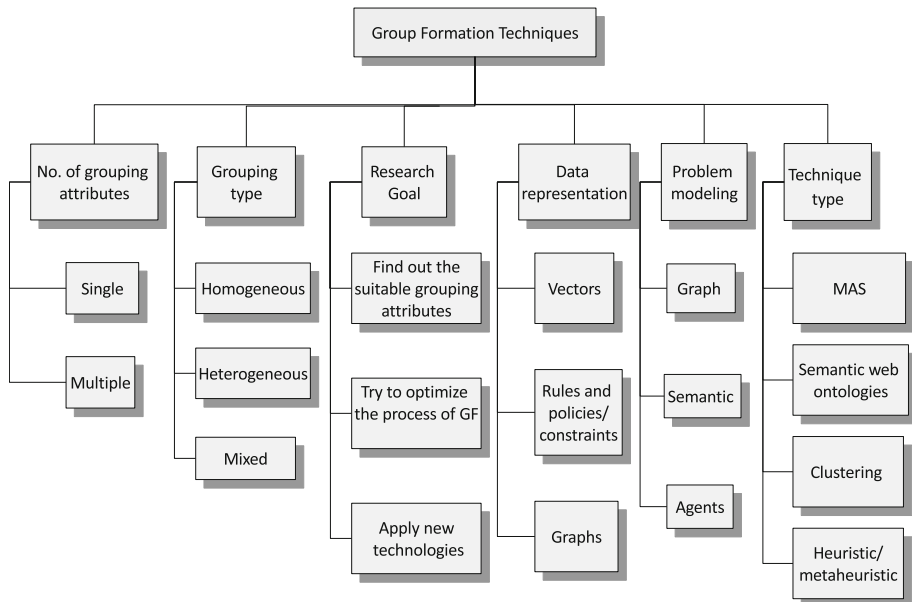

Fig. 6 Taxonomy of surveyed GF techniques

Many researchers merged more than one goal. For instance, a research may conclude with applying suitable attributes with iterative experiments to obtain the optimal solution (Abnar et al. 2012; Tien et al. 2013; Zheng and Pinkwart 2014).

Looking more deeply into previous work on group formation, it becomes apparent that there are two approaches of formation which compare the similarity of the group members characteristics: either clustering the homogeneous characteristics or constrained optimization for heterogeneous and mixed characteristics (Christodoulopoulos and Papanikolaou 2007; Hubscher 2010).

A proposed taxonomy on classification of the techniques, which are reviewed and discussed previously, is presented in Fig. 6. This taxonomy demonstrates various criteria to classify techniques from different perspectives. As shown in the figure, the applied techniques can be classified based on the final formation of the group. The grouping techniques can be homogeneous, heterogeneous or mixed. Also these techniques differ in the point of data representation according to the context of the problem. The problem of group formation process was formulated by different models in order to solve it optimally. Surveyed research modeled the GF problem through using agents, semantic networks and graphs. Clearly, these works may be classified through examining the type of the technique used. The type of technique varies from one context to another. Thus, the adoption of one particular technique could be based on heuristic/metaheuristic, multi-agents, clustering or semantic ontologies.

4 Discussion and Future Trends

This section discusses the obtained results from the SLR. 18 studies have been reviewed in group formation area to achieve the research aims. The aims were: discovering the recent contributions in group formation in CSCL contexts, exploring the effective attributes and techniques on grouping process, summarizing and representing findings in a structural manner revealing knowledge gaps, challenges and opportunities.

It is obvious that group formation process has been investigated from two important perspectives. The first one is the attributes that affect the grouping process and the second is about techniques used in CSCL contexts. These perspectives are discussed in details in Sects. 4.1 and 4.2.

The study analysis has revealed results that have led to the classification of the contributions according to various viewpoints. The classification comes up with two novel taxonomic maps to represent both perspectives. These maps were presented in the Sect. 3 and are briefly discussed within group formation perspectives.

Finally, challenges and opportunities are summarized based on the knowledge gaps found in the related literature.

4.1 Attributes of Group Formation

As illustrated in Table 3, group formation process depends on the chosen attributes in each study. It is clear that studies differed in their chosen attributes and their number. Five of them used only single attribute to form groups i.e. (Abnar et al. 2012; Mujkanovic et al. 2012; Srba and Bielikova 2015; Yannibelli and Amandi 2011; Zheng and Pinkwart 2014) while two studies used two attributes in their group formation (Christodoulopoulos and Papanikolaou 2007; Graf and Bekele 2006). On the other hand, there were 3 studies out of

the 18 used 3 attributes (Brauer and Schmidt 2012; Ho et al. 2009; Ounnas et al. 2008a). The aims of these studies were to explore the capability of forming groups with multiple criteria in more complex situations. The rest of studies used other attributes depending on their available and data sets (Amara et al. 2016; Hubscher 2010; Jozan and Taghiyareh 2013; Moreno et al. 2012).

The most common used attribute was knowledge. It was used in 6 studies i.e. (Brauer and Schmidt 2012; Christodoulopoulos and Papanikolaou 2007; Graf and Bekele 2006; Ho et al. 2009; Srba and Bielikova 2015; Sukstrienwong 2012). This seems to be the most suitable and important attribute to form educational groups because of its effects on the group outcomes.

In addition, the attributes of learning styles and personality traits were discussed similarly. Each one was used in 4 studies (Abnar et al. 2012; Christodoulopoulos and Papanikolaou 2007; Mujkanovic et al. 2012; Ounnas et al. 2008b; Zheng and Pinkwart 2014). Educationally, learning style is an effective attribute that plays a vital role in students learning and subsequently affects the grouping process. Besides, personality traits are the simplest gathered attributes for experiments and studies.

The recent approach has added new attributes. It focused on the learners relationships and their roles within teams. Thus, there were three studies that used social interactions as an attribute to form groups (Brauer and Schmidt 2012; Ho et al. 2009; Ounnas et al. 2008a) while other two studies used the team role attribute (Ounnas et al. 2008a; Yannibelli and Amandi 2011).

The analyzed attributes directed this research to represent them using a novel taxonomic map according to their classification and relationship to group formation process. The classification of attributes findings of the study are shoun in Fig. 5.

According to aforementioned discussion, the more flexible system that offers many choices for the instructor to form his/her groups, the grouping process can better assist in.

4.2 Techniques Used in Group Formation

It is obvious that evolutionary algorithms were used as seen in Table 3 in group formation process (Abnar et al. 2012; Brauer and Schmidt 2012; Graf and Bekele 2006; Ho et al. 2009; Jozan and Taghiyareh 2013; Moreno et al. 2012; Tien et al. 2013; Yannibelli and Amandi 2011). These types of algorithms were approximately 60% of the reviewed studies. For example, genetic algorithm was the dominant used technique. It was used in more than 5 studies (Brauer and Schmidt 2012; Jozan and Taghiyareh 2013; Moreno et al. 2012; Sukstrienwong 2012; Tien et al. 2013).

Clustering algorithms were also frequently used, especially for homogeneous grouping. Approximately, four studies used these algorithms (Amara et al. 2016; Christodoulopoulos and Papanikolaou 2007; Mujkanovic et al. 2012; Srba and Bielikova 2015). While other studies varied in their techniques because of their aims (e.g. trying new approach for grouping). Soh and Khandaker (2007), Ounnas et al. (2008b) and Srba and Bielikova (2015) used multiagents, semantic ontologies and group technology, respectively.

4.3 Opportunities and Challenges of Group Formation

Based on the related literature, trends are observed in the area of group formation and it is obvious that the reviewed work covered the automated group formation process from the viewpoint of education and collaborative learning. It is worthwhile to try reapplying this

process in other contexts, where people need to working in groups. Such contexts include training, business and psychology.

However, there are still many issues that are not sufficiently discussed and solved in group formation. These issues are gaps and shortcomings in the reviewed literature. They are used as a basis for defining the directions for future investigations, as follows:

1. Choosing a specific technique to form a group is oriented by the context of the group formation problem. However, in some situations there are many suitable techniques to be applied in the group formation process. This concept leads the interest of the researchers to ask about the reason behind applying the chosen technique, which is not clear in some of the previous work.
2. Using local datasets for evaluating the proposed strategies may be a shortcoming in the situation of comprehensive comparison among the computational and pedagogical results of the applied group formation techniques.
3. A comprehensive paradigm that expresses all the details of group formation process in different situations should be developed in conjunction with other disciplines such as education, training and psychology.
4. Incomplete solutions were introduced to solve the group formation in various contexts. Thus, there is a need to develop an autonomous system that holds main grouping operations and learning preferences in the field of group formation.
5. Poor contributions are obvious in the field of quality metrics that measure the quality of group formation process from different viewpoints. Therefore, quality of service (QoS) as an evaluation framework for group formation is a fertile area to deal with comprehensively for identifying the success of group formation process.
6. Mobile and cloudy environments are the new trends for developed systems. Thus, incorporating group formation process within these environments would offer facilities to various disciplines.

4.4 Implications

Based on the reviewed literature, various trends are observed in the area of group formation. It is obvious that the reviewed work covered the automated group formation process from the viewpoint of education and collaborative learning. It is worthwhile to attempt reapplying this process in other contexts, where people need to work in groups. Such contexts include training, business and psychology.

However, there are still many issues that are not sufficiently discussed and solved in group formation. These issues include gaps and shortcomings in the reviewed literature. They are used to indicate the implications revealed from reviewing group formation in different studies. These implications are concerned with the perspectives of research: implications for future research and implications for practice.

4.4.1 Implications for Future Research

As this review focused on contributions of recent research studies on group formation, almost implications are for future research. These implications are discussed below.

1. Choosing a specific technique to form a group is oriented by the context of the group formation problem. However, in some situations there are many suitable techniques to be applied in the group formation process. This concept arouses the interest of the

   researchers to ask about the reason behind applying the chosen technique, which is not
   clear in some of the previous work.
2. Using local datasets for evaluating the proposed strategies may be a shortcoming in the
   situation of comprehensive comparison among the computational and pedagogical
   results of the applied group formation techniques.
3. A comprehensive paradigm that expresses all the details of group formation process in
   different situations should be developed in conjunction with other disciplines such as
   education, training and psychology.
4. Poor contributions are obvious in the field of quality metrics that measure the quality
   of group formation process from different viewpoints. Therefore, quality of service
   (QoS) used as an evaluation framework for group formation is a fertile area for
   identifying the success of group formation process.

4.4.2 Implications for Practice

Group formation still needs applicable systems. These systems should consider the
following:

1. Incomplete solutions were introduced to solve the group formation in various contexts.
   Thus, there is a need to develop an autonomous system that holds main grouping
   operations and learning preferences in the field of group formation.
2. Mobile and cloudy environments are the new trends for developed systems. Thus,
   incorporating group formation process within these environments would facilitate
   various disciplines.

5 Conclusion

Automated group formation process become an important issue in terms of talking about
collaborative learning. The state of the art of group formation has shown the various and
widespread works from different viewpoints. In this study, a systematic literature review is
introduced on group formation process in CSCL contexts. It highlights the attributes
affecting the process of group formation through presenting taxonomy of these attributes.
Such attributes were categorized according to different criteria. The techniques of group
formation were reorganized in another proposed taxonomy which tackles the way of
grouping, the problem definition, the data representation and the type of the applied
technique. These taxonomies were constructed to answer the research questions: What are
the most effectively used attributes and techniques in group formation process; how can the
recent CSCL contributions be represented in group formation process within educational
context; what are the knowledge gaps and limitations in group formation?
   Through this survey, it is obvious that group formation process still needs more
improvements to be an ideal process in CSCL environment. Some of these improvements
are directly related to the process itself. For instance, obtaining complete solutions with
optimal performance is the most critical issue. On the other hand, other improvements may
be inter-related with other computing areas such as cloud computing and mobility.

References

Abnar, S., Orooji, F., & Taghiyareh, F. (2012). An evolutionary algorithm for forming mixed groups of learners in web based collaborative learning environments. In *2012 IEEE international conference on technology enhanced education (ICTEE)* (pp. 1–6). IEEE.

Amara, S., Macedo, J., Bendella, F., & Santos, A. (2016). Group formation in mobile computer supported collaborative learning contexts: A systematic literature review. *Journal of Educational Technology & Society*, 19(2), 258–273.

Bonebright, D. A. (2010). 40 years of storming: A historical review of tuckman's model of small group development. *Human Resource Development International*, 13(1), 111–120.

Brauer, S., & Schmidt, T. C. (2012). Group formation in elearning-enabled online social networks. In *2012 15th international conference on interactive collaborative learning (ICL)* (pp. 1–8). IEEE.

Christodoulopoulos, C. E., & Papanikolaou, K. A. (2007). A group formation tool in an e-learning context. In *19th IEEE international conference on tools with artificial intelligence (ICTAI 2007)*, vol. 2, (pp. 117–123). IEEE.

Coffield, F., Moseley, D., Hall, E., & Ecclestone, K. (2004). Learning styles and pedagogy in post 16 learning: A systematic and critical review. *The Learning and Skills Research Centre*.

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

Graf, S., & Bekele, R. (2006). Forming heterogeneous groups for intelligent collaborative learning systems with ant colony optimization. In *International conference on intelligent tutoring systems* (pp. 217–226). Springer.

Ho, T. F., Shyu, S. J., Wang, F. H., & Li, C. T. J. (2009). Composing high-heterogeneous and high-interaction groups in collaborative learning with particle swarm optimization. In *2009 WRI World congress on computer science and information engineering* vol. 4, (pp. 607–611). IEEE.

Huang, Y., Zhu, M., Wang, J., Pathak, N., Shen, C., Keegan, B., Williams, D., & Contractor, N. (2009). The formation of task-oriented groups: Exploring combat activities in online games. In *2009 CSE'09. International conference on computational science and engineering* vol. 4, (pp. 122–127). IEEE.

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

Jozan, M. M. B., & Taghiyareh, F. (2013). An evolutionary algorithm for homogeneous grouping to enhance web-based collaborative learning. *International Journal of Computer Science Research and Application*, 3(1), 74–85.

Khandaker, N., Soh, L. K., & Jiang, H. (2006). Student learning and team formation in a structured CSCL environment. *Frontiers in Artificial Intelligence and Applications*, 151, 185.

Kreijns, K., Kirschner, P. A., & Jochems, W. (2002). The sociability of computer-supported collaborative learning environments. *Educational Technology & Society*, 5(1), 8–22.

Martin, E., & Paredes, P. (2004). Using learning styles for dynamic group formation in adaptive collaborative hypermedia systems. In *Proceedings of the First International Workshop on Adaptive Hypermedia and Collaborative Web-based Systems (AHCW 2004)* (pp. 188–198). http://www.ii.uam.es/rcarro/AHCW04/MartinParedes.pdf.

Matazi, I., Messoussi, R., & Bennane, A. (2014). The design of an intelligent multi-agent system for supporting collaborative learning. In *2014 9th International conference on intelligent systems: Theories and applications (SITA-14)* (pp. 1–8). IEEE.

Moreno, J., Ovalle, D. A., & Vicari, R. M. (2012). A genetic algorithm approach for group formation in collaborative learning considering multiple student characteristics. *Computers & Education*, 58(1), 560–569.

Mujkanovic, A., Lowe, D., Willey, K., & Guetl, C. (2012). Unsupervised learning algorithm for adaptive group formation: Collaborative learning support in remotely accessible laboratories. In *2012 International conference on information society (i-Society)* (pp. 50–57). IEEE.

Okoli, C., & Schabram, K. (2010). A guide to conducting a systematic literature review of information systems research. *Sprouts: Working Papers on Information Systems*, 10, 26.

Ounnas, A. (2008). Semantic web-based group formation for e-learning. In: *ESWC 2008 Ph. D. Symposium* (p. 51).

Ounnas, A., Davis, H. C., Millard, D. E. (2007a). Semantic modeling for group formation. In *11th International Conference on User Modeling, Corfu, Greece* (p. 71).

Ounnas, A., Davis, H. C., & Millard, D. E. (2007b). Towards semantic group formation. In *Seventh IEEE International Conference on Advanced Learning Technologies, 2007. ICALT 2007*. IEEE (pp. 825–827).

Ounnas, A., Davis, H., & Millard, D. (2008a). A framework for semantic group formation. In *2008 Eighth IEEE international conference on advanced learning technologies* (pp. 34–38). IEEE.

Ounnas, A., Davis, H., & Millard, D. (2008b). Semantic web-based group formation for e-learning.

Resta, P., & Laferrière, T. (2007). Technology in support of collaborative learning. *Educational Psychology Review*, *19*(1), 65–83.

Rowe, J. P., Shores, L. R., Mott, B. W., & Lester, J. C. (2010). Integrating learning and engagement in narrative-centered learning environments. In *International Conference on Intelligent Tutoring Systems* (pp. 166–177). Springer.

Rubens, N., Vilenius, M., & Okamoto, T. (2009). Automatic group formation for informal collaborative learning. In *Proceedings of the 2009 IEEE/WIC/ACM international joint conference on web intelligence and intelligent agent technology-volume 03* (pp. 231–234). IEEE Computer Society.

Soh, L. K. (2004). On cooperative learning teams for multiagent team formation. In *of the AAAIs 2004 Workshop on forming and maintaining coalitions and teams in adaptive multiagent systems*.

Soh, L. K., Khandaker, N. (2007). Forming and scaffolding human coalitions with a multi-agent framework. In *Proceedings of the 6th international joint conference on autonomous agents and multiagent systems* (p. 62). ACM.

Soh, L. K., Khandaker, N., & Jiang, H. (2008). I-minds: A multiagent system for intelligent computer-supported collaborative learning and classroom management. *International Journal of Artificial Intelligence in Education*, *18*(2), 119–151.

Soller, A. (2001). Supporting social interaction in an intelligent collaborative learning system. *International Journal of Artificial Intelligence in Education (IJAIED)*, *12*, 40–62.

Srba, I., & Bielikova, M. (2015). Dynamic group formation as an approach to collaborative learning support. *IEEE Transactions on Learning Technologies*, *8*(2), 173–186.

Stahl, G., Koschmann, T., & Suthers, D. (2006). Computer-supported collaborative learning: An historical perspective. *Cambridge Handbook of the Learning Sciences*, *2006*, 409–426.

Sukstrienwong, A. (2012). Genetic algorithm for forming student groups based on heterogeneous grouping. In *Recent advances in information science: Proceedings of the 3rd European conference of computer science* (pp. 92–97).

Sun, G., & Shen, J. (2013). Teamwork as a service: a cloud-based system for enhancing teamwork performance in mobile learning. In *2013 IEEE 13th International Conference on Advanced Learning Technologies (ICALT)* (pp. 376–378), IEEE.

Tien, H. W., Lin, Y. S., Chang, Y. C., & Chu, C. P. (2013). A genetic algorithm-based multiple characteristics grouping strategy for collaborative learning. In *International conference on web-based learning* (pp. 11–22). Springer.

Tuckman, B. W., & Jensen, M. A. C. (1977). Stages of small-group development revisited. *Group & Organization Management*, *2*(4), 419–427.

Yannibelli, V., & Amandi, A. (2011). Forming well-balanced collaborative learning teams according to the roles of their members: An evolutionary approach. In *2011 IEEE 12th international symposium on computational intelligence and informatics (CINTI)* (pp. 265–270). IEEE.

Zheng, Z., & Pinkwart, N. (2014). A discrete particle swarm optimization approach to compose heterogeneous learning groups. In *2014 IEEE 14th international conference on advanced learning technologies* (pp. 49–51). IEEE.