Quality of Group Formation in CSCL Environments

Naseebah Maqtary[1] [a], Abdulqader Mohsen[1] [b] and Kamal Bechkoum[2]

[1] *Faculty of Computing and Information Technology, University of Science and Technology, Sana'a, Yemen*

[2] *School of Business and Technology, University of Gloucestershire, Gloucestershire, U.K.*

Keywords: Group, Quality, Collaboration, Framework, Review.

Abstract: Group Formation (GF) plays a vital role in groupwork performance, for it is the opening phase of the group development process. Many studies have been conducted to form groups in various scenarios to enhance collaborative learning. These studies used different clustering techniques, and therefore, the applied evaluation measures in each study depend on the context of the group formation process. However, there is a lack of an integrative framework to qualify the overall process of group formation. This paper proposes such a framework that is composed of layers to tackle each issue related to the GF process. The framework is called the framework of the Quality of Group Formation (QoGF). QoGF includes three different levels in which every level has its evaluation measures. These measures are group quality, group formation quality and empirical quality, which are totaled in an aggregative measure called Total Quality (TQ).

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. 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).

CSCL has emerged during the mid-1990s. Various tools have been used and employed to merge collaboration within educational activities (Stahl et al., 2006) such as emails, blogs videoconferencing systems, content management systems (CMS), and others. Focusing on collaborative learning has brought groupwork to the fore. Many studies in the CSCL environment have been carried out on administrating groupwork activities like group formation, monitoring, and evaluation (Sun and Shen, 2013).

Forming a group that collaboratively learns is one of the most challenging tasks in the CSCLs context. This topic attracted the interest of several researchers (Amara et al., 2016; Khandaker et al., 2006; Srba and Bielikova, 2015).

Group formation (GF) is the opening step of the group development life cycle. It affects the outcomes of the group in some way. The process of formation should be performed with specific issues in mind. These issues can vary from the group's special characteristics to the goal and objectives of the group's task. In educational contexts, the learners are grouped according to their characteristics, the nature of the required task, and the mechanism of the formation.

However, group formation still has shortcomings in various perspectives such as attributes, techniques, and measures, which affect the whole process. GF process with its specified attributes and techniques

[a] https://orcid.org/0000-0002-9061-019X
[b] https://orcid.org/0000-0002-3820-918X

Maqtary, N., Mohsen, A. and Bechkoum, K.
Quality of Group Formation in CSCL Environments.
DOI: 10.5220/0009391402990306
In *Proceedings of the 12th International Conference on Computer Supported Education (CSEDU 2020) - Volume 2*, pages 299-306
ISBN: 978-989-758-417-6

have been investigated throughly in (Maqtary et al., 2017), but quality measures are still not covered in a holistic manner. There is a need to survey the applied measures in contributed studies. Thus, it will help to compare these measures and ascertain their effects on GF. The main motivation behind this study is to provide a summary of quality measures of group formation to enhance collaborative learning in CSCL environments.

This paper proposes a framework for group formation quality. To achieve this, many steps are traced. Firstly, surveying the related work and answering the following questions:

- What are the applied evaluation measures?
- What are the used data, real or synthesized, and how much are their size?
- Is there a comparison with others' work?

Secondly, the gaps in the quality of GF in the literature are specified based on the given answers. Finally, an integrative framework is proposed to cover such gaps.

This paper is organized as follows: the next section introduces the process of group formation. Related work is detailed in Section 3. Section 4 is dedicated to describing findings while Section 5 shows the proposed framework. Discussing the applicability of this framework is presented in Section 6. Final remarks and conclusion is drawn in Section 7.

2 GROUP FORMATION

Group formation is the first and opening process of the group development life cycle in which efforts should be devoted to ensuring the effectiveness and efficiency of the process (Bonebright, 2010). 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 and efficiently achieving the learning outcomes (Khandaker et al., 2006). To understand the context of group formation process, some issues and details should be introduced. The next Subsection (2.1) highlights the basic concepts of GF such as the definition, the attributes, the grouping mode, and the group size while Subsection (2.2) discusses the required steps of group formation.

2.1 Preliminaries of GF

Group formation is defined in (Konert, 2014, p.16) as *"the challenge to optimize learning group formation from a given set of learners, respecting homogeneously and heterogeneously in simultaneous to match criteria and aiming for a balanced quality of the build groups"*.

As understood from the definition mentioned above, the idea of formation is to organize the learners in some clusters that need to be balanced. Balanced groups are groups that achieve their assigned tasks successfully (Zheng et al., 2018). We add to the definition *the balanced groups should match specified criteria set by the instructor or the learning situation*.

In addressing the high level of collaboration, some of the current work concentrated on the learner's attributes/characteristics and their effect on group performance, while other focused on the group's achievement and collaboration strategies.

Learner's attributes represent various aspects of learner's readiness to learn. They can be competences, personality traits, learning style, team role, or social interaction (Maqtary et al., 2017). Group formation process uses learner's attributes to decide what are the most suitable groups that meet the task goal and requirements.

Another issue related to GF is the grouping mode, which specifies the type of formed groups. Formed groups can be homogeneous, heterogeneous, or mixed depending on the members' characteristics. Homogeneous groups have objects that strongly relate to each other. Heterogeneous groups have high diversity between their objects. Mixed groups have objects that are similar in some attributes and different in others. As stated in (Graf and Bekele, 2006), homogeneous groups are better to achieve specific goals, while heterogeneous groups are preferred when innovative and creative solutions are required. Mixed-mode needs both homogeneity in some characteristics and heterogeneity in others. The decision about the best choice of formation mode is based on the task nature and the required level of collaboration between learners.

Group size is also another issue that affects group performance but still lacks attention in terms of the number of significant studies that investigate its effectiveness (Resta and Laferrière, 2007). Stahl et al. (2006) determined that group size is one of the controlling independent variables that affects collaborative learning. Kooloos et al. (2011) reported that small group size is better because it stimulates motivation, cohesiveness, development, and cognition. It is worth noting that there is no consensus in literature to determine the optimal group size in collaborative learning. However, adequate group size is argued to be five to six (Kooloos et al., 2011).

2.2 Process of GF

A detailed view of the group formation process is drawn after surveying existing work. This view may help researchers to consider the specific issues of the process and qualify it. It consists of many steps, as illustrated in Figure 1. These steps are summarized as follows:

1. ***GF Data Gathering:*** i)gathering the learner and group attributes, and ii) ordering the priorities of these attributes according to their effectiveness on collaborative learning.
2. ***GF Configuration:*** specifying the attributes and mode of the grouping process.
3. ***GF Execution:***it means executing the formation process using the proposed technique to output the groups.
4. ***GF Evaluation:*** evaluating the output groups using various metrics.
5. ***GF Final Approval:*** giving the instructor the option of manually final groups' approving.

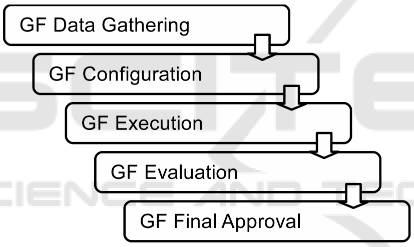

Figure 1: Steps of GF Process.

This overview of GF gives the most critical aspects to be taken into account when considering the quality of GF. In the next section, the contributions in this field are presented and summarized.

3 RELATED WORK

Based on state of the art, the studies concentrated on various distance measures used to evaluate each group. In some cases, the distance measure should be minimal to clarify the homogeneity within each group. On the other hand, the bigger the distance measure is, the better the result of heterogeneous formation is. Distance measures are considered as evaluation measures to indicate the acceptance level of formed groups. However, some studies added further measures that can be considered as post-performance measures for the formed groups. They tried to evaluate the quality of formed groups through relying on groups' performance indicators such as post-tests and members' responses to specific questionnaires.
Furthermore, the experiments of these studies have been carried out using different datasets. This issue complicates the comparison between these studies to ensure the effectiveness of the introduced technique. This unstandardized evaluation in the process of GF is the crucial motivation to overview the contributed factors related to quality and used datasets and conclude with sum-up review.
In (Graf and Bekele, 2006), heterogeneous groups were formed. The heterogeneity was based on the learner's score. Average of distance (AD_i) in a specified group was computed through summing the maximum and minimum values of members' attributes. Then the goodness of heterogeneity (GH) was calculated by Equation 1.

$$AD_i = \frac{maxscore(S_j) + minscore(S_j)}{2}$$
$$GH_i = \frac{maxscore(S_j) - minscore(S_j)}{1 + \Sigma_j \left| AD_i - S_{j(i)} \right|}, \quad (1)$$

where $S_{j(i)}$ means the j^{th} learner in i^{th} group.
The higher the GH, the feasible the heterogeneity in the group is. Graf and Bekele (2006) used five datasets of 100 learners and checked the quality of the grouping through the GH only. It also measured its scalability by executing the algorithm on a dataset of 512 learners. No comparison was held with others.
Ounnas et al. (2007) proposed a general framework for evaluating the quality of the GF process. The researchers proposed a metrics framework to evaluate constraint satisfaction-based group formation. The framework based on ontologies. They used formation quality (FQ) and some constraints violations (NCV) parameters, which were built on average and standard deviation (SD) to evaluate their framework. The proposed FQ dealt with goals satisfaction. As researchers stated that from the learning viewpoint, the group formation quality is a multi-dimensional concept which implies the formation efficiency besides the groups' performance indicators. As shown in Figure 2, the framework is composed of the following three parts:

1. Formation metrics
   - Constraint Satisfaction (Group, Cohort).
   - Perceived formation satisfaction (Individual, Group, Cohort).
2. Productivity metrics
3. Goal satisfaction metrics
   - Goal satisfaction quality (Group, Cohort).
   - Formation quality (Group, Cohort).

The framework was not empirically implemented and no comparison was held with others.

Procedure 1 Calculating group formation quality

Given task t, $form_{opt}(t)$, $f_{optG}(t)$
for each formation $form_n$
  for each group g_i in the cohort
    for each goal $\alpha_k \in t$
      for each constraint $c_i \in \alpha_i$
        calculate $f_{cj}(g_i, c_j)$
      calculate $f_g(g_i, \alpha_k)$
    calculate $f_{fg}(g_i, t) = f_{fg}(g_i, \alpha_1, \alpha_2 \dots \alpha_K)$
  calculate $f_{fG}(t) = f_{fG}(\alpha_1, \alpha_2 \dots \alpha_K)$
  if $f_{fG}(t) > f_{optG}(t)$
  then $form_{opt} \leftarrow form_n$
return $form_{opt}$

Figure 2: Group Formation Quality Ounnas et al. (2007).

Ho et al. (2009) used particle swarm optimization (PSO) technique to form heterogeneous groups. This study used one real dataset of 61 learners. However, no comparison was held with others. The required formation quality was obtained by maximizing the objective function, as shown in Equation 2.

$$O_{\max} = w_1 \Sigma_{a=1}^{m} \Sigma_{b=a+1}^{n} (-|CPT_a - CPT_b|) + w_2 \Sigma_{j=1}^{m} (DIF_j) + w_3 \Sigma_{j=1}^{m} (INT_j / \Sigma_{x=1}^{n} x_{ij}), \tag{2}$$

where CPT is the total competences of the j_{th} group, INT is the overall interaction among students in the j_{th} group, and DIF is the summation of the style differences among students in the j_{th} group.
Yannibelli and Amandi (2011) used evolutionary algorithm to form heterogeneous groups with balancing roles as illustrated in Equations 3, 4 and 5.

$$nr(G_i, r) = \begin{cases} 1 & r \text{ is naturally played by one member of } G_i \\ -2 & r \text{ is not naturally played in } G_i \\ -p & r \text{ is naturally played by } p \text{ members in } G_i \end{cases} \tag{3}$$

$$nb(G_i) = \Sigma_{r=1}^{9} nr(G_i, r) \tag{4}$$

$$\max_{\forall G \in C} \left(b(G) = \frac{\Sigma_{i=1}^{g} nb(G_i)}{g} \right), \tag{5}$$

where $nr(G_i, r)$ calculates the balance levels of the roles in each group, as shown in Equation 3 and $nb(G_i)$ computes the average of balances in all groups as presented in Equation 4. The last Equation 5 calculates the best population that generates the balanced groups (nb should equal to nine roles based on Belbin's model). The study conducted many experiments using ten synthesized datasets. It also compared its performance by the execution time with the exhaustive method (EM).
Moreno et al. (2012) formed intra-heterogeneous and inter-homogeneous groups using genetic algorithm (GA). The study used a real, local dataset composed of 135 learners. It compared the GA with the other two algorithms; random method (RM) and (EM). It used D_i as an evaluation measure, as shown in Equation 6. The less the distance D_i is, the more the inter-homogeneous group.

$$D_i = \Sigma_{g=1}^{G} \left[(\overline{C_1} - \overline{X_{g,1}^i})^2 + (\overline{C_2} - \overline{X_{g,2}^i})^2 + \dots + (\overline{C_M} - \overline{X_{g,M}^i})^2 \right], \tag{6}$$

where $\overline{C_i}$ is the mean of each attribute in the group i, $\overline{X_g^i}$ is the mean of each attribute in all individuals.
In the same vein, Tien et al. (2013) improved GA to form heterogeneous groups. The study used the fitness function that was represented in Equation 7.

$$F_i = \omega \cdot \overline{GP} + (1 - \omega) \cdot (1 - GP^{\sigma}), \tag{7}$$

where $\overline{GP}$, GP^{σ} and ω respectively denote the mean value of all groups, the standard deviation of all groups, and the weight of $\overline{GP}$, GP^{σ}. The mean of the relative closeness to the optimal solution is indicated using GP_g, as shown in Equation 8.

$$GP_g = \frac{\sqrt{\Sigma_{q=1}^{Q} \left(\frac{\Sigma_{s=1}^{Eq} Z_{s,q}^{g}}{E_g} - V_q - \right)^2 \cdot \omega_q}}{\sqrt{\Sigma_{q=1}^{Q} \left(\frac{\Sigma_{s=1}^{Eq} Z_{s,q}^{g}}{E_g} - V_q + \right)^2 \cdot \omega_q} + \sqrt{\Sigma_{q=1}^{Q} \left(\frac{\Sigma_{s=1}^{Eq} Z_{s,q}^{g}}{E_g} - V_q - \right)^2 \cdot \omega_q}} \tag{8}$$

As stated in the study, the proposed algorithm gave slightly better solutions compared to other algorithms, which were simple GA (SGA) and (RM). It used five simulated datasets with various volumes to evaluate the proposed GA.
Amara et al. (2016) used the Average Intra-cluster Distance (AID) without mentioning the equation. The bigger the measure is, the more heterogeneity the cluster is. They used simulated dataset without specifying its volume, and no comparison was held with others
Acharya and Sinha (2018) formed mixed groups. They used the measure E, which is called the sum of square errors to ensure the homogeneity, as shown in Equation 9.

$$E = \Sigma_{i=1}^{K} \Sigma a \in C_i |a - m_i|^2, \tag{9}$$

where a is an individual, and m_i is the mean of the group C_i. Also, heterogeneity was expressed by Equation 10.

$$H(G_i) = \frac{x}{r}, \tag{10}$$

where x is the number of intervals that represent a certain attribute, and r is the number of group's members. Comparison was held with other studies including (Graf and Bekele, 2006) and (Christodoulopoulos and Papanikolaou, 2007). The study used one real dataset with 72 learners.

4 FINDINGS

As summarization showed in Table 1, group formation quality in the literature lacks a comprehensive

Table 1: Summary of GF Quality in Literature Work.

Author and publication year	Evaluation Equations			Dataset			Comparison with Others
	Homo	Hete	Mixed	Local	Real	Synth.	
Graf and Bekele (2006)	-	Equ. 1	-	5	✓	-	-
Ounnas et al. (2007)	-	*	-	1	✓	✓	-
Ho et al. (2009)	-	Equ. 2	-	1	✓	-	-
Yannibelli and Amandi (2011)	-	Equ. 3 & 4 & 5	-	10	-	✓	EM
Moreno et al. (2012)	-	Equ. 6	-	1	✓	-	EM & RM
Tien et al. (2013)	-	Equ. 7 & 8	-	5	-	✓	SGA & RM
Amara et al. (2016)	*	*	-	1	-	✓	-
Acharya and Sinha (2018)	-	-	Equ. 9 & 10	1	✓	-	*

* Not available.
\- Not applied.

quality framework that considers all expected contexts. Each study used its proposed measure and, in the best cases, ran some comparisons with others to evaluate their proposed metrics and techniques. The only introduced quality framework was detailed in (Ounnas et al., 2007) and presented in Figure 2. However, it is still a proposal framework and not implemented or evaluated yet. Moreover, it is not suitable for all group formation models since it was formulated based on ontologies. Its measures and levels concentrate on empirical performance.
According to the above discussion, inadequate contributions are apparent in the field of quality metrics that measure the quality of the group formation process from different viewpoints.
Therefore, it is essential to implement a comprehensive quality framework for the group formation process. The framework should give directive indicators about the groups' formation process for the instructors. There is no need to wait until the end of the groups' works to evaluate their formation. Instead, some measures need to be proposed to assess the formed groups before beginning the actual work. The concentration should be on how well the groups were formed rather than how well they will perform. These measures will indicate the success of the formation process. In this context, the following questions may arise: How can the group formation process be evaluated? what are the most important factors that indicate the success of the formation process?

5 QUALITY OF GROUP FORMATION FRAMEWORK (QoGF)

Group formation in CSCL is essential, especially in large cohorts and short-term groups. To qualify the formation process, there is a need for quality metrics. Here, a comprehensive framework quality of group formation (QoGF) is proposed. Based on the surveyed literature, there is a lack of comprehensive quality metrics that deal with all specifications of the formation process. QoGF is composed of three levels to reflect various aspects of group formation. Intentionally, the order of the levels reveals that the quality should be considered while forming groups. Group quality, formation process quality, and empirical quality are the levels of QoGF. The following subsections define these levels and present their applied evaluation measures.

5.1 Levels of QoGF

As Figure 3 shows, QoGF has three nested levels. The most inner level is the group quality (GQ), while the most outer one is the empirical quality (EQ). Formation process quality (FPQ) is in the middle of the proposed framework. Below sections define each level and discuss its role in the GF process.

5.1.1 Group Quality (GQ)

GQ is the inner level, which ensures two perspectives: each group's quality and satisfaction of assigned constraints. The first perspective is the quality of each group, which is called *intra-class quality*. Various distance measures can be used to examine the intra-class quality, such as Davis-Bouldin Index (DBI), Dunn, etc. GQ measures evaluate the homogeneity and compactness inside each group according to the clusters' validity indicators. Some of them also measure the separability between groups. Choosing the measure is based on the nature of the data and the context of the formation process. The second perspective is the *constraints satisfaction*, which verifies if the formation process fulfills the predefined constraints such as the task's nature, the instructor's constraints, and the learner's preferences. These constraints are selectively applied depending on the context of the GF process. They cover the various required configurations as discussed below.

1. Task's nature deals with different aspects that rely on the context of the required task. These aspects are various, including grouping mode (Homogeneous, heterogeneous and/or mixed), size (variable or fixed) and/or priorities of attributes.
2. The instructor's constraints give opportunities for the instructor to set up his constraints, such as excluding somebody from a specific group or solving the problem of unassigned learners (orphans) to any group at the end of the process.
3. The learner's preference takes in to account the preferences of learners as some tasks take care of

learners' satisfaction (e.g., preserving the friendship between the learners).

5.1.2 Formation Process Quality (FPQ)

The purpose of FPQ is evaluating the goodness and balance of the whole process of formation. It is divided into two parts: inter-class quality and algorithm quality.

1. *Inter-class quality* evaluates whether the whole formed groups are balanced. It means that groups are similar in satisfying the formation constraints and are in the same level of quality. To measure such issue, any balance factor can be applied.
2. *Algorithm quality* evaluates the time and space complexity of the used algorithm.

5.1.3 Empirical Quality (EQ)

EQ is the outer level of the QoGF framework. It evaluates the success of the groups' formation (e.g., the groups' performance) according to the achieved goals. Abnar et al. (2012) reported that most of the GF studies did not evaluate the real outcomes after groups' task completion. The only thing researchers measured is comparing the value of quality factors they defined for the groups formed by their algorithm. This assumption leads us to consider this type of quality. It is measured through reporting and analyzing the responses of all groups members to the predesignated questionnaires that investigate many factors such as satisfaction, performance, goal achievement, and advancement in collaborative learning such as discussed in (Abnar et al., 2012).

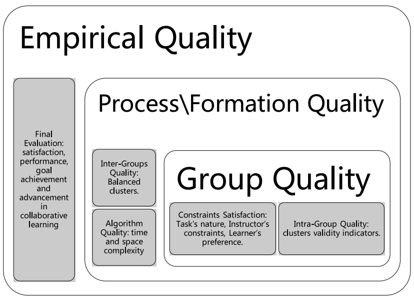

Figure 3: Proposed Quality Framework for GF (QoGF).

5.2 Evaluation Measures

As presented in the QoGF framework, many evaluation measures are available to be used among the different levels of the framework. These levels are applicable to any chosen grouping mode, which are: homogeneous, heterogeneous, and mixed. The following list clarifies each mode with its possible evaluation measure.

1. **Homogeneous mode** in a specific attribute can be achieved through minimizing the difference between the values of that attribute in the same group. Forming homogeneous groups means that similarity/homogeneity should be high intra-group and dissimilarity/heterogeneity should be also high inter-groups. Many existed validation measures are used to qualify this type of groups such as homogeneity and separation scores, silhouette width, redundant score and WADP (Chen et al., 2002).
2. **Heterogeneous mode** can be evaluated through maximizing the number of various values of a specific attribute that should be heterogeneous in the same group. Heterogeneous formed groups are those groups with high dissimilar/heterogeneous intra-group and high similar/homogeneous inter-groups. One of the most attractive measures that are used to evaluate these groups is *K-complementarity measure*. K-complementarity measure verifies the fulfillment of all roles in the each group. It is borrowed from forming groups of tutors in the educational context, as presented in (Lafifi et al., 2014).
3. **Mixed mode** means that formed groups are composed of objects with some attributes that are homogeneous, and other attributes are heterogeneous. In this mode, evaluation measures are combined from measures of both modes; homogeneous and heterogeneous.

Below subsections introduce the measures GQM, $FPQM$, and EQM of the GQ, FPQ, and EQ levels, respectively, with consideration of all modes of group formation: homogeneous, heterogeneous and mixed. Besides, another measure called total quality measure (TQM) is introduced to finalize the GF quality process. TQM is calculated based on the outcomes of the GQ, FPQ, and FPQ measures.

5.2.1 Group Quality Measure

GQM evaluates the fulfillment of GQ criteria; intra-class and constraints satisfaction. It also varies based on the chosen grouping mode. In homogeneous mode, any intra-class quality measure can be used. Also, the percentage of satisfied constraints selected by the instructor and learners should not be violated. These two criteria are formulated by $groups_{intra-class}$ and $constraints$, respectively. They should be maximized in homogeneous mode to fulfill the quality of the GQ level. On the other hand, the heterogeneous mode can use the same criteria of homogeneity with the inverse of $groups_{intra-class}$. This inverse is required to reflect the difference inside the formed groups. Finally, in the mixed-mode, the same measures of both homogeneity GQM_{homo} and heterogeneity GQM_{hete} are applied as a preferred setting that is specified by the instructor. The group quality level is totally measured by GQM, as shown in the Equation 11. The more the value of GQM is, the better the results are obtained.

$$GQM = \begin{cases} constraints + groups_{intra-class} & \text{Homogeneous groups} \\ constraints + \frac{1}{groups_{intra-class}} & \text{Heterogeneous groups} \\ GQM_{homo} + GQM_{hete} & \text{Mixed groups} \end{cases} \tag{11}$$

5.2.2 Formation Process Quality Measure

In FPQ, the QoGF middle level, different criteria are chosen to evaluate the inter-class quality and algorithm complexity. Inter-class quality should measure the balance and equilibrium among all formed groups by applying any assigned balance measure, which may be affected by the formation context and instructor preference. It is expressed by $proc_{inter-class}$, and it should be maximized. Algorithm complexity can be expressed by $algo_{complexity}$. The algorithm complexity should be minimized to gain better performance. The $FPQM$ value should be maximized to achieve better results and calculated using Equation 12.

$$FPQM = \frac{1}{algo_{complexity}} + proc_{inter-class} \tag{12}$$

5.2.3 Empirical Quality Measure

The outer level of QoGF is to model the empirical performance of formed groups. As discussed previously, it is measured by the learners' improved performance and questionnaires about their satisfaction of achievement. All these indicators should be quantified using suitable measures ($chosen - measures$) and should be maximized to reflect better results. EQM is expressed by Equation 13

$$EQM = max(chosen - measures) \tag{13}$$

5.2.4 Total Quality Measure

The total quality measure TQM expresses the quality of the whole GF process as shown in Equation 14. It is the summation of the qualities of the QoGF levels with preferred weights set by the instructor. The value of this measure should be maximized to record the best results.

$$TQM = GQ * w_{GQ} + FPQ * w_{FPQ} + EQ * w_{EQ} \tag{14}$$

6 DISCUSSION

As presented in Section 5, this paper introduces an integrative framework, QoGF, for the process of group formation in different cases and constraints. In comparison with the literature, previously conducted studies applied the GF process on specially chosen contexts, while QoGF gives an umbrella for applying the process by enabling the instructors to set the process configuration according to the required collaborative context. The GF process configuration includes the chosen grouping mode, used intra-group measure, balance among formed groups, algorithm feasibility, outcomes of the formed groups, and other assigned preferences and constraints. Further, QoGF offers the capability of evaluating the achievement of such configurations through different levels and assigned measures.

The contribution of QoGF is its ability to finalize the GF process using a quantitative measure that indicates the overall process quality. In addition, it can be considered as a basis that allows comparison among different scenarios and settings to form groups using various grouping techniques.

7 CONCLUSIONS AND FUTURE WORK

The group formation process is an important step that affects the other steps of the group development life cycle in CSCL environments. Several studies were conducted to automatically form groups in different contexts. They used various quality measures to evaluate the process. This paper reviews the existing studies and summarizes their contributions according to quality metrics and used datasets. After analysis of these studies, shortcomings in GF quality were recorded. Therefore, an integrative quality framework was proposed to alleviate the lack of quality standards in GF.

This framework includes three layers to reflect the nature of the process. It is composed of group quality (GQ), formation process quality (FPQ), and empirical quality (EQ). Besides that, different evaluation measures are formulated to permit the applicability of qualifying the GF process through various levels and contexts.

For future work, this framework needs to be implemented and evaluated to ensure its suitability to most contexts of formation in CSCL environments. According to the applied datasets, it is preferable to have a benchmark dataset that can be used in the context of group formation.

REFERENCES

Abnar, S., Orooji, F., and Taghiyareh, F. (2012). An evolutionary algorithm for forming mixed groups of learners in web based collaborative learning environments. In *Technology Enhanced Education (ICTEE), 2012 IEEE International Conference on*, pages 1–6. IEEE.

Acharya, A. and Sinha, D. (2018). A "mixed" strategy for collaborative group formation and its learning outcomes. *Journal of Educational Technology Systems*, 46(4):440–462.

Amara, S., Macedo, J., Bendella, F., and 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.

Chen, G., Jaradat, S. A., Banerjee, N., Tanaka, T. S., Ko, M. S., and Zhang, M. Q. (2002). Evaluation and comparison of clustering algorithms in analyzing es cell gene expression data. *Statistica Sinica*, pages 241–262.

Christodoulopoulos, C. E. and 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)*, volume 2, pages 117–123. IEEE.

Dillenbourg, P. (1999). What do you mean by collaborative learning. *Collaborative-learning: Cognitive and computational approaches*, 1:1–15.

Graf, S. and Bekele, R. (2006). Forming heterogeneous groups for intelligent collaborative learning systems with ant colony optimization. In *International Conference on Intelligent Tutoring Systems*, pages 217–226. Springer.

Ho, T.-F., Shyu, S. J., Wang, F.-H., and Li, C. T.-J. (2009). Composing high-heterogeneous and high-interaction groups in collaborative learning with particle swarm optimization. In *Computer Science and Information Engineering, 2009 WRI World Congress on*, volume 4, pages 607–611. IEEE.

Khandaker, N., Soh, L.-K., and Jiang, H. (2006). Student learning and team formation in a structured cscl environment. *FRONTIERS IN ARTIFICIAL INTELLIGENCE AND APPLICATIONS*, 151:185.

Konert, J. (2014). *Interactive Multimedia Learning: Using Social Media for Peer Education in Single-Player Educational Games*. Springer.

Kooloos, J. G., Klaassen, T., Vereijken, M., Van Kuppeveld, S., Bolhuis, S., and Vorstenbosch, M. (2011). Collaborative group work: Effects of group size and assignment structure on learning gain, student satisfaction and perceived participation. *Medical Teacher*, 33(12):983–988.

Lafifi, Y., Bendjebar, S., and Zedadra, A. (2014). A k-complementarity technique for forming groups of tutors in intelligent learning environments. *Journal of computing and information technology*, 22(2):115–130.

Maqtary, N., Mohsen, A., and Bechkoum, K. (2017). Group formation techniques in computer-supported collaborative learning: A systematic literature review. *Technology, Knowledge and Learning*, pages 1–22.

Matazi, I., Messoussi, R., and Bennane, A. (2014). The design of an intelligent multi-agent system for supporting collaborative learning. In *Intelligent Systems: Theories and Applications (SITA-14), 2014 9th International Conference on*, pages 1–8. IEEE.

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

Ounnas, A., Millard, D. E., and Davis, H. C. (2007). A metrics framework for evaluating group formation. In *Proceedings of the 2007 international ACM conference on Supporting group work*, pages 221–224. ACM.

Resta, P. and 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., and Lester, J. C. (2010). Integrating learning and engagement in narrative-centered learning environments. In *International Conference on Intelligent Tutoring Systems*, pages 166–177. Springer.

Srba, I. and 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. D., and Suthers, D. D. (2006). *Computer-supported collaborative learning*.

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

Tien, H.-W., Lin, Y.-S., Chang, Y.-C., and Chu, C.-P. (2013). A genetic algorithm-based multiple characteristics grouping strategy for collaborative learning. In *International Conference on Web-Based Learning*, pages 11–22. Springer.

Yannibelli, V. and Amandi, A. (2011). Forming well-balanced collaborative learning teams according to the roles of their members: An evolutionary approach. In *Computational Intelligence and Informatics (CINTI), 2011 IEEE 12th International Symposium on*, pages 265–270. IEEE.

Zheng, Y., Li, C., Liu, S., and Lu, W. (2018). An improved genetic approach for composing optimal collaborative learning groups. *Knowledge-Based Systems*, 139:214–225.