



Examining the effect of a genetic algorithm-enabled grouping method on collaborative performances, processes, and perceptions

Xu Li¹ · Fan Ouyang² · WenZhi Chen¹

Accepted: 20 April 2022 / Published online: 31 May 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Group formation is a critical factor which influences collaborative processes and performances in computer-supported collaborative learning (CSCL). Automatic grouping has been widely used to generate groups with heterogeneous attributes and to maximize the diversity of students' characteristics within a group. But there are two dominant challenges that automatic grouping methods need to address, namely the barriers of uneven group size problem, and the inaccessibility of student characteristics. This research proposes an optimized, genetic algorithm-based grouping method that includes a conceptual model and an algorithm module to address these challenges. Through a quasi-experiment research, we compare collaborative groups' performance, processes, and perceptions in China's higher education. The results indicate that the experimental groups outperform the traditional grouping methods (i.e., random groups and student-formed groups) in terms of final performances, collaborative processes, and student perceptions. Based on the results, we propose implications for implementation of automatic grouping methods, and the use of collaborative analytics methods in CSCL.

Keywords Computer-supported collaborative learning · Group formation · Higher education · Genetic algorithm · Learning analytics

✉ Fan Ouyang
fanouyang@zju.edu.cn

✉ WenZhi Chen
chenwz@zju.edu.cn

¹ College of Computer Science and Technology, Zhejiang University, 38 Zheda Rd., Hangzhou 310027, People's Republic of China

² College of Education, Zhejiang University, 866 Yuhangtang Rd., Hangzhou 310058, People's Republic of China

Introduction

Grounded upon the socio-cognitive constructivism (Sawyer, 2005), computer-supported collaborative learning (CSCL) is a socio-cultural process completed by a group of students with the support of purposeful instructions and technologies and formed through emergent interactions and communications (Hernández-Sellés et al., 2019; Jeong et al., 2019; Stahl et al., 2014). To achieve a high quality of CSCL, group members need to pool together their expertise and knowledge, build mutual understandings of shared goals, and form sustained interactions and dialogues (Hernández-Sellés et al., 2019; Ludvigsen, 2016; Wang & Hwang, 2012). Therefore, the group's formations and structures are critical factors that can significantly influence the group's collaborative processes and performances (Chen & Kuo, 2019; Lin et al., 2016; Moreno et al., 2012; Reis et al., 2018). Due to the importance of the group formation for assuring the CSCL quality, varied grouping methods have been studied in the CSCL field (Chen & Kuo, 2019; Lin et al., 2016; Moreno et al., 2012; Reis et al., 2018). A widely used approach is using automatic grouping methods to generate groups with heterogeneous attributes within a group, that is to maximize the diversity of students' characteristics (Lambić et al., 2018; Lin et al., 2016; Moreno et al., 2012). But there are two major challenges the automatic grouping methods need to address, as previous studies indicated, namely the barriers of uneven student numbers within groups (i.e., the uneven group size problem) (Ahmad et al., 2021; Holenko Dlab et al., 2020), and the inaccessibility of student characteristics at the starting point (i.e., the "Cold Start" problem) (Lika et al., 2014; Pliakos et al., 2019). To address those two challenges, this research proposed an optimized, genetic algorithm-based grouping method that includes a conceptual model named *Feature Categorization Model* (FCM) to cope with the "Cold Start" problem and a GA-enabled module named *Insert Virtual Members* (IVM_{GA}) to address the group size problem. Furthermore, this research conducted a quasi-experiment in small groups' collaborative writing activities in China's higher education context and used a multi-method approach to compare the groups' collaborative processes, final performances, and student perceptions towards collaboration. Groups were generated by the proposed grouping method and two traditional grouping methods, namely, the random groups and student-formed groups. The empirical research results indicated that our proposed method outperformed traditional grouping methods. Based on the results, we proposed implications for design and implementation of grouping methods, and the use of collaborative analytics methods in CSCL.

Literature review

Existing grouping methods in CSCL

CSCL requires a group of students to pool knowledge and skills together, get to know and learn from peers, and share and construct ideas to achieve shared goals that cannot be completed by any individual alone (Chen & Kuo, 2019; Jeong

et al., 2019; Stahl et al., 2014). One of the core issues in CSCL is grouping formation, because a group's learning atmospheres, processes, and performances are determined by how well the group members work together (Ahmad et al., 2021; Chen & Kuo, 2019; Uto et al., 2020). The traditional, non-automatic grouping methods include random grouping, student-formed groups, and the instructor-assigned groups. Simple random grouping may result in merely a few members in a group to complete the collaborative task while others work as free riders, which does not allow all group members to contribute their knowledge, capacities, and skills (Chen & Kuo, 2019; Costaguta, 2015). The student-formed groups are structured by students themselves, such that they usually form groups based on their prior acquaintance, without fully considering educational factors (Krouska & Virvou, 2020; Srba & Bielikova, 2015). In addition, although the instructor-assigned groups are determined by the instructor who might mix different students' characteristics to some extent, they only focus on a few characteristics such as gender or grade (Lin et al., 2016). The primary disadvantage of those traditional grouping methods is that the diversity of students' characteristics such as communication skills, leadership capacities, and knowledge levels, are not adequately taken into account, which may result in undesired collaborative outcomes (Costaguta, 2015; Huxham & Land, 2000; Lin et al., 2016). In other words, those non-automatic grouping methods merely take into consideration some characteristics from students, but cannot find a global optimal solution (Chen & Kuo, 2019). When group members' characteristics and capacities cannot supplement each other, the group's collaborative potential could be weakened, in some cases even resulting in negative effects of collaboration (Alfonseca et al., 2006).

Considering the complexity of grouping in actual educational situations, it is difficult to generate optimal groups through traditional grouping methods by considering multiple factors, and to cope with the large size of students manually in a short period of time (Lin et al., 2010; Takači et al., 2017; Yannibelli & Amandi, 2012). To address these limitations, a series of studies have successfully used automatic grouping methods to form optimized groups in the K-12 and higher education settings. For example, Jong et al. (2006) used the dynamic-grouping or partial-grouping methods to support students find the most suitable partners based on individual student's knowledge structures, which is represented by a conceptual graph tool. Lin et al. (2010) proposed an enhanced particle swarm optimization to form well-structured collaborative learning groups. Lambić et al. (2018) used the variable neighborhood search (VNS) algorithm to automatically generate heterogeneous groups based on students' pretest scores, interpersonal relationships, prosocial behaviors, and openness characteristics. In addition, Chen and Kuo (2019) used genetic algorithms (GA) to transfer the group composition problem into a multi-objective optimized problem in order to generate optimized groups. They used a fitness function to evaluate the availability and then generated the globally optimized solution by assigning different weights to student characteristics. Overall, compared to traditional grouping, the automatic grouping methods have advantages to generate optimized groups to improve teaching and learning quality in higher education, by taking multiple student characteristics as input variables and addressing the increasing class size challenge.

Critical factors in automatic grouping methods

As stated in previous studies, individual characteristics, grouping criteria, and grouping method have critical influences on the group performance (Chen & Kuo, 2019; Moreno et al., 2012; Qiu & McDougall, 2015). There are three critical factors to consider in the automatic grouping methods, namely (1) What student characteristics to be used, (2) What grouping criteria are adopted, and (3) What algorithms to be used for generating groups. First, student characteristics include *static* characteristics and *dynamic* characteristics. The static characteristics usually do not change or at least do not change within a short period of learning, such as gender, age, prior knowledge levels, or learning styles. For instance, learning style reflects the learner's preferred way of learning, do not change in a short-term process (Alfonseca et al., 2006; Gibbs & Bernas, 2007). In contrast, the dynamic characteristics—cannot be captured at a static point—tend to continuous change during the students' learning processes, such as interactive levels or emotional status. Previous studies do not explicitly distinguish static and dynamic characteristics from students. A combination of student characteristics is usually used in previous group formation research. For example, Moreno et al. (2012) considered student's subject knowledge levels, communicative levels, and leadership levels to generate optimized groups. Chen and Kuo (2019) argued that interaction and gender were important factors for the grouping composition. Krouska and Virvou (2020) considered seventeen factors from academic, cognitive, and social aspects to form groups. However, the assumption is weakened when some dynamic characteristics are not available. For example, students' social interactions cannot be observed at the beginning of a course, which raises a "Cold Start" problem for the grouping method. Therefore, relevant studies should consider the explicit distinguish between static and dynamic characteristics from students, and carefully consider the availability of students' characteristics in the research.

Second, there are at least three types of grouping criteria, namely random grouping that does not consider specific student characteristics, homogeneous grouping where members with similar characteristics are grouped, as well as heterogeneous grouping where members with different or complementary characteristics are grouped (Chen & Kuo, 2019; Lou et al., 1996; Van der Laan & Spindle, 2007). Although previous studies reported discrepancies about the effect of different grouping criteria on collaborative learning, most studies argue that compared to randomly-selected and homogeneous groups, heterogeneous groups are beneficial to promote mutual learning and peer interactions between students with different levels of knowledge and capacities. For example, Wang et al. (2007) described the experimental results of heterogeneous groups based on students' thinking styles. Compared to random groups, the heterogeneous groups had a statistically significant capability to complete the course requirements and had a smaller inter-group variation in the final performance. Takači et al. (2017) compared the heterogeneous groups with random groups, and the result reported that heterogeneous groups had a higher average grades and pass rates than the random groups. Compared to traditional grouping methods, Chen and Kuo (2019) concluded that heterogeneous groups supported the goal of group formation in collaborative learning, where

learners learned from each other, met different learners, and shared ideas to achieve optimal learning outcomes. In summary, heterogeneous grouping is becoming a widely adopted grouping criterion as it can better cater for diverse educational scenarios to achieve desirable collaborative effects.

The third critical factor is the algorithms used to automatically generate groups (including the choice of the types of input variables). In previous research, the algorithms used to automatically generate groups include clustering, decision tree, and rank segmentation, etc. (Costaguta, 2015). These algorithms usually require a higher level of computing power to sustain computational accuracy and efficiency or external interventions from the instructors such as providing expert knowledge (Lin et al., 2010). Compared with those methods, genetic algorithm (GA) is proved to be an effective method to solve optimized grouping problem, since it transfers the problem of how to get an optimized group into a multi-objective optimization problem (Chen & Kuo, 2019; Kumar et al., 2010; Moreno et al., 2012). Specifically, GA can obtain an optimized solution for a task facing numerous solutions in a limited time. Group formation is a typical optimization problem requiring a suitable solution from tens of thousands of grouping solutions. GA-enabled grouping methods can take multiple student characteristics as input to generate optimized heterogeneous groups (Chan et al., 2010; Chen et al., 2012; Moreno et al., 2012). In addition, compared to other methods, GA is more flexible to generate groups under different criteria, and also maintains a certain level of computational efficiency (Chen & Kuo, 2019; Krouska & Virvou, 2020; Sukstrienwong, 2017). GA has been proved to be effective for forming groups with desirable learning outcomes and experiences (e.g., Chen & Kuo, 2019). In addition, when feeding student characteristics to the grouping algorithms, the input variable includes the continuous numerical type (e.g., Moreno et al., 2012) and the categorical variable type (e.g., Chen & Kuo, 2019; Krouska & Virvou, 2020). The continuous numerical type is usually used as input to a genetic algorithm, since it not only improves the computational efficiency, but also increases the robustness (Srba & Bielikova, 2015). In summary, three critical factors are considered in the automatic grouping methods in this research: a combination of student characteristics, heterogeneous grouping criteria, and the enhanced GA algorithms with continuous numerical type as input.

Challenges encountered in the automatic grouping methods

As previous studies indicated, there are two major challenges that automatic grouping methods need to address, namely the uneven group size problem (Holenko Dlab et al., 2020), and the data inaccessibility (“Cold Start”) problem (Lika et al., 2014; Pliakos et al., 2019). In CSCL, instructors usually configure a group with the size of 3 or 4 members, as prior research results have showed that a large group size would weaken the group performance (Gibbs et al., 2001). But in actual educational situations, a challenge is raised: the class size cannot be divided evenly (i.e., the uneven grouping size problem). Previous studies usually reported the student size that can be evenly assigned into groups, e.g., 135 participants of five students per group (Moreno et al., 2012), 32 participants with the group size of 4 (Sadeghi & Kardan,

2016), 48 students with 3 members per group (Krouska & Virvou, 2020). Chen and Kuo (2019) selected 83 students in which 27 students are divided into the experiment group with different group sizes (size=4 and size=5). In this study, they admitted that it was impossible to assure all groups with an even student population as well as an even distribution of roles. Although many studies have demonstrated the effect of the GA-enabled algorithms when groups can be equally assigned with student numbers (Krouska & Virvou, 2020; Moreno et al., 2012; Wang et al., 2007), it is often impossible to evenly split the number of students in the practical educational contexts. To address this practical challenge, it is necessary to assure that the grouping algorithm module can achieve a similar effect, including collaborative performance, engagement process, and student perception, when the student numbers cannot be evenly split.

Another challenge in group formation is the “Cold Start” problem, namely the difficulties of acquiring student characteristics data at the start of the collaborative learning (Lika et al., 2014; Pliakos et al., 2019). As we mentioned above, unlike static characteristics such as gender, age, learning style, some other characteristics are not available at the beginning, such as cognitive changes, social interactions, and emotions. Under this circumstance, the automatic grouping algorithms encounter the intractable challenge of “Cold Start” problem. The “Cold Start” problem is a common challenge in data-driven applications. For example, Schein et al. (2002) described that their naive Bayes classification system did not work well when the new products were not rated and fed into the algorithm. Lika et al. (2014) also reported the recommendation system did not have enough information to make recommendations when new users were entered. Similarly, in the education field, Pliakos et al. (2019) elaborated a challenge that adaptive learning systems faced, that is, they did not have enough information at the starting point about new students when they initially entered a learning environment. Since the automatic grouping method is a typical data-driven method that heavily depends on the student characteristics dataset, it is critical to obtain expected input variables of student characteristics in order to generate desirable grouping. However, due to the complexity of educational contexts, obtaining data of student characteristics is not an easy task, and it is particularly difficult to acquire the dynamic characteristics during the learning process. A solution is obtaining available static features for group formation, such that to some extent overcoming the “Cold Start” problem, encountered with data-driven algorithms. In summary, to achieve optimized solution, the uneven group size problem and the inaccessibility of student characteristics are two challenges need to be addressed in the automatic grouping methods.

Analytical methods used to examine the effect of CSCL

With the support of quantitative methods, relevant work has collected and analyzed process-oriented and performance-oriented data to reveal groups’ final performances, perceptions, and collaborative processes in order to understand the complex effect of CSCL. First, to examine the groups’ final performance, statistical analysis methods (e.g., ANOVA and t-test) are usually used to examine the significant differences of

the groups' collaborative performance (Moreno et al., 2012; Wang et al., 2007). The final performance analysis focuses on the summative evaluations of the group or individual scores, rather than the process-oriented evaluation. Second, self-report methods are usually used to examine student perceptions about the collaboration. For example, Delaney et al. (2019) adopt a survey to explore students' attitudes toward the group-based collaborative work. Chen and Kuo (2019) adopted the semi-structured interview approach to analyze the perception of group members about the collaboration in the groups. Extant studies have compared the differences of grouping methods on final student achievements and individual perceptions, but there is a lack of the comparison of groups' collaborative processes (Takači et al., 2017; Wang et al., 2007).

The interactive, collaborative process is a critical aspect to demonstrate groups' collaborative quality (Chen & Kuo, 2019). According to Stahl (2009a, 2009b, 2009c), the qualitative, ethnographic approaches (e.g., conversation or discourse analysis) have been used to examine micro-level turn-taking relevancies between interactional, behavioral, and cognitive activities during a short time period of collaborative learning. Damşa (2014) used in-depth qualitative analysis to investigate the nature of productive interactions, the joint efforts to co-construct knowledge and the shared epistemic agency that emerged during groups' problem-solving processes over time. With the development of learning analytics and educational data mining techniques, recent work has used quantifiable methods, such as social network analysis, or quantifiable content analysis to analyze the CSCL process. An important advantage of quantitative approaches is that they can be applied in large size of data or long discourse sequences to investigate the structures, patterns, or orderings based on standardized procedures (Cohen et al., 2013). For example, Chen and Kuo (2019) used the social network analysis to assess the collaborative processes of groups with different group methods. Moreover, since cognitive engagement is a critical factor to reflect a group's collaboration, either at the individual level or group level, we argue that it is important to examine the cognitive engagement in the collaborative learning process. Considering the multi-dimensional characteristics of the collaborative learning, merely focusing on the final performance or perceptions of collaboration may cause incomprehensive results. A comprehensive quantitative method can be used to analyze both performance-oriented and process-oriented data with an aim to provide a more holistic, multilevel, and multidimensional analysis of groups' collaborative characteristics (Janssen et al., 2013; Joksimović et al., 2018; Suthers et al., 2013). Echoing this trend, this research collects both performance and process data during CSCL and uses a multi-method approach to analyze groups' final products, student perceptions, and the processes on the social and cognitive engagement dimensions, in order to gain a deep understanding of the effect of different grouping methods on CSCL.

The proposed grouping method

There are two modules in the overarching framework of the proposed grouping method, namely a conceptual module and a GA-enabled algorithm module. The conceptual model is *Feature Categorization Module* (FCM) that captures and

divides student characteristics into static and dynamic attributes in order to address the “Cold Start” problem (see Fig. 1, part 1). To address the unbalanced group size challenge, an algorithm module is designed based on genetic algorithm, namely *GA-enabled Insert Virtual Members* (IVM_{GA}) to insert virtual members with the average level of the actual measures of student characteristics (see Fig. 1, part 2).

First, to address a “Cold Start” problem, the conceptual module *Feature Categorization Module* (FCM) is proposed. Prior studies have indicated that student characteristics related to group work include gender, major, age, leadership (Moreno et al., 2012; Yilmaz et al., 2020), communication skills (Moreno et al., 2012), collaborative learning skills (Soller et al., 1998), learning styles (Alfonseca et al., 2006; Sukstrienwong, 2017), knowledge levels (Chen & Kuo, 2019), and social interaction (Chen & Chang, 2014; Chen & Kuo, 2019). However, some characteristics are not always accessible at the beginning of the course, such as social interaction levels, which cause a “Cold Start” problem. To address this challenge, FCM classifies student characteristics into the static and dynamic categorizations: the static category indicates students’ fixed characteristics that are not changing; the dynamic category indicates students’ characteristics that cannot be captured at the beginning of the class which is changing throughout the learning process. Since static characteristics do not change in a short period and are always available, those characteristics can be used to initiate grouping when dynamic characteristics are not available. In this way, the “Cold Start” problem can be addressed to some extent.

In addition, to address the unbalanced group size challenge, an algorithm module *GA-enabled Insert Virtual Members* strategy (IVM_{GA}) is proposed (see Fig. 1, Part 2). When the number of students can be equally assigned into groups with the same group size, IVM_{GA} activates the FCM module to capture students’ static and dynamic characteristics and then automatically generate optimized groups. But in actual educational situations, students usually cannot be equally divided into groups with an equal size, which cause the unbalanced group size problem. When groups cannot be assigned with the same group size, IVM_{GA} first generates several virtual members in terms of the average level of actual students’ characteristics and then

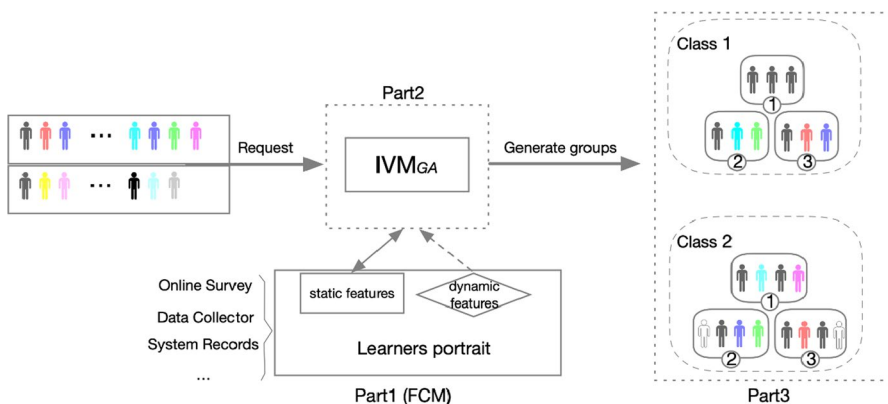


Fig. 1 The overarching framework of the proposed grouping method

generates optimized groups with the same group size by inserting virtual members (see Fig. 1, Part 3). Particularly, IVM_{GA} aims to generate virtual members to sustain a certain level of heterogeneity within the group and sustain a certain level of homogeneity among groups. There are four steps in the IVM_{GA} module. The first step is to create student characteristic matrix to describe a student with variable characteristics and then preprocess the student data as input data for the grouping algorithm (see Fig. 2). The input student data is a $N \times M$ matrix, where a row represents one student in the class and a column represents the measurements of characteristics of the student. Since different scales of the characteristic impact the final result, IVM_{GA} normalizes all characteristic data into the same scale of 0–1. $F_{n,m}$ represents a student's value of a specific characteristic after normalization.

The second step is to generate a global characteristic matrix. If virtual students are needed, the IVM module generates a characteristic matrix of those virtual students with appropriate characteristic data. The IVM module first calculates the number of virtual members that need to be generated and then generates virtual students' characteristic matrices. The STU_{IVM} represents a virtual member's characteristic, with $\overline{F_m}$ representing the m -th characteristic (see Eq. 1).

$$STU_{IVM} = \left\{ \overline{F_1}, \overline{F_2}, \dots, \overline{F_m} \right\} \quad (1)$$

Then the module inserts the virtual member STU_{IVM} into the global characteristic matrix, which includes all actual students and all virtual students generated (see Eq. 2).

$$FM = \left\{ \overline{F_1}, \overline{F_2}, \dots, \overline{F_m} \right\} \quad (2)$$

In this step, IVM_{GA} is used to generate students' global characteristic matrix (see Algorithm 1). Then, the IVM_{GA} uses the global characteristic matrix to generate a group population that includes many random group methods. Next, a fitness function will assess these group methods to find the better group genes and process crossover and mutation operation. This is an iterative process to find a better grouping scheme, and this process ends with the maximum iterations.

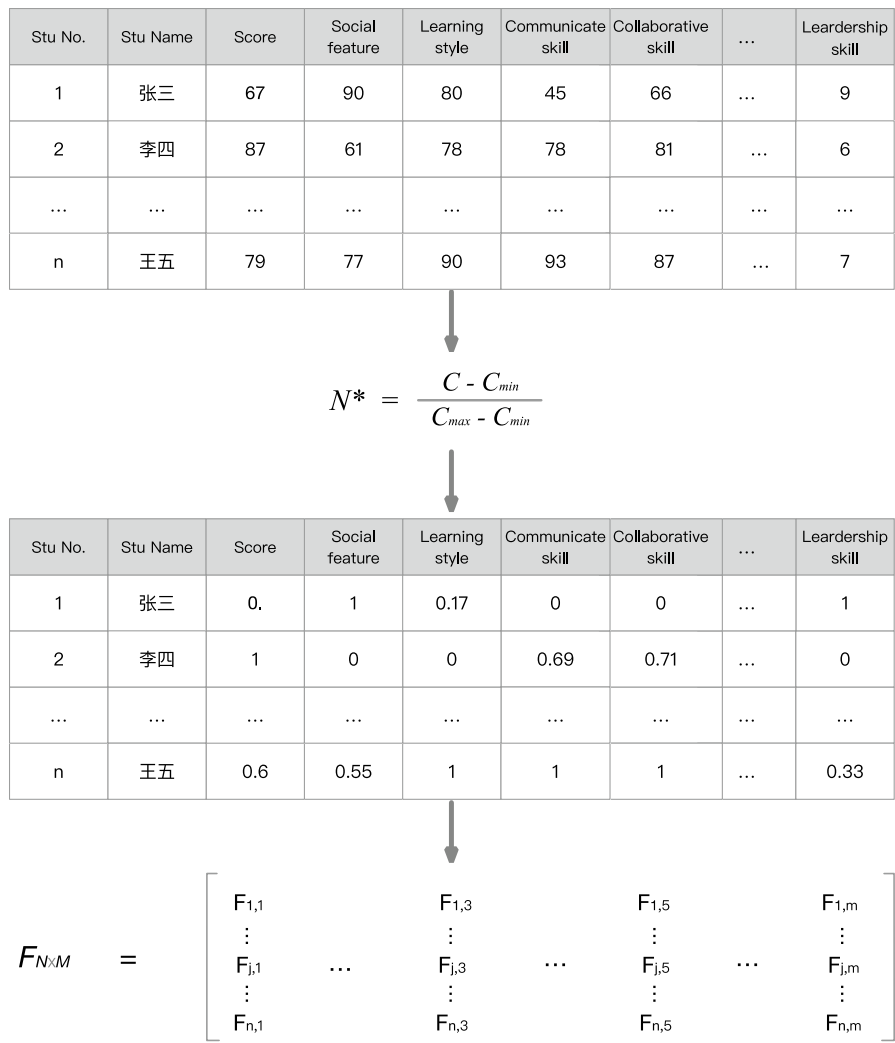


Fig. 2 The student characteristic matrix

Algorithm 1. Descriptions of core parts of IVM_{GA}

Input: The current students' characteristics matrix size, $F_{current}$; and the group size, G

Output: The new characteristics matrix, F_{new}

Function calculateCharacteristicMatrix($F_{current}$, G):

initialization

$F_{current}$, read current characteristic matrix

N , the number of students, calculate the number of $F_{current}$ rows

G , current group size parameter

/* calculate the number of IVM students, $N_{virtual}$

/* % is mod operation

$N_{virtual} = (G - N \% G)$

STU_{IVM} , calculate the virtual student' matrix.

if $N_{virtual} == 0$ **then**

$F_{new} = F_{current}$

return F_{new}

else

if $N_{virtual} > G/2$ and $G \neq 3$ **then**

$N_{virtualNew} = 0$

For select G_{new} from the [3,4,5]

$N_{temp} = (G_{new} - N \% G_{new})$

if $N_{temp} \leq G_{new} / 2$

$N_{virtualNew} = N_{temp}$

break

$G_{new} += 1$

pop-up window confirms whether instructor reassign the suggested

optimal group size parameter, $N_{virtualNew}$

if selected use the suggested optimal group size **then**

STU_{IVM_all} , generate $N_{virtualNew}$ STU_{IVM}

F_{new} , concatenate $F_{current}$ and STU_{IVM_all}

return F_{new}

STU_{IVM_all} , generate $N_{virtual}$ STU_{IVM}

F_{new} , concatenate $F_{current}$ and STU_{IVM_all}

return F_{new}

end

end

The third step is to generate a matrix to represent each characteristic's average value (see Eq. 3). For one group ($1 \leq g \leq G$) under a group scheme, the group is a small matrix consisting of team's students. $T_{g,m}^i$ indicates the average of the characteristic at each dimension.

$$IM_g^i = \left\{ \overline{T_{g,1}^i}, \overline{T_{g,2}^i}, \dots, \overline{T_{g,m}^i} \right\} \quad (3)$$

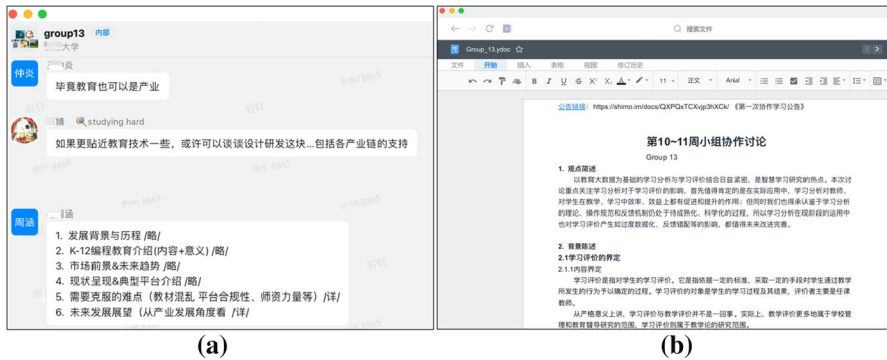


Fig. 3 Screenshots of DingTalk (a) and ZheDaYunPan (b)

The fourth step is to calculate each group schema's fitness value (see Eq. 4). The fitness value can be used to estimate the suitability of the grouping method.

$$F^i = \sum_{g=1}^G [(\overline{F_1} - \overline{T_{g,1}})^2 + (\overline{F_2} - \overline{T_{g,2}})^2 + (\overline{F_m} - \overline{T_{g,m}})^2] \quad (4)$$

Moreover, the IVM_{GA} method also can form random groups when we lower the value of “iteration times”. In our experiment, we use IVM_{GA} to generate random groups. The global parameter of the number of initialized populations is also considered. The higher the parameter we set, the larger the solution space is and the longer execute duration we get. We used the IVM_{GA} algorithm to generate both IVM_{GA} groups and GR groups; the difference was that a reduced number of iterations (a parameter configuration) was used to generate GR groups, while a larger optimization space was required (i.e., more iterations) to generate IVM_{GA} groups.

The research methodology

The research purposes and questions

The quasi-experimental design aims to examine whether there are significant differences of the collaborative performances, perceptions, and processes under different grouping methods, namely the experimental groups, and the control group of student-formed groups, and the control group of randomly-assigned groups. There are three research questions as follows:

- (1) To what extent did three grouping methods differentiate in terms of the groups' final performance?
- (2) To what extent did three grouping methods differentiate in terms of the social and cognitive engagement during the collaborative processes?

Table 1 Statistics of groups based on three grouping methods

Grouping methods	Number of groups	Number of 3-person groups	Number of 4-person groups	Number of students
IVM _{GA}	6	1	5	23
GR	8	3	5	29
GS	4	0	4	16
Total	18	4	14	68

- (3) To what extent did three grouping methods differentiate in terms of students' perceptions?

The instructional context, participants, and procedure

The research context was an undergraduate-level course titled “*Modern Educational Technologies*” offered at a Chinese research-intensive university. This course focused on learning theories, instructional design, educational technologies, emerging tools, and trending topics. This course was typically offered in the face-to-face, classroom environments; due to COVID-19, all courses moved online through an online learning management system XueZaiZheDa. This course was co-taught by three instructors, taking responsibilities to facilitate different weekly topics. 68 undergraduate sophomore students (51 females, 17 males) enrolled in this course.

Table 2 The descriptions of SNA metrics

SNA metrics	Description
APL	The average number of the shortest paths for all possible pairs of nodes
Density	The ratio of actual links between any nodes to all potential possible links), and it represents the whole network structure's cohesion
Average degree	Degree denotes the number of links pointing to or away from a node, including out-degree and in-degree; average degree indicates the mean of all group members' degree
Average closeness	Closeness is the length of paths from a student to all others in the network, defined as the inverse total length; average closeness is the mean of all group members' closeness
Average betweenness	Betweenness is the number of shortest paths that passes through a student; average betweenness is the mean of all group members' betweenness
Reciprocity	Ratio of symmetric dyads to non-null dyads in a network
GCC	GCC reflects the degree of connectedness between this node and its neighborhood, ranged between 0 and 1
ICV	The inverse of the standard deviation of student–student interaction frequency divided by its mean
Centralization	The extent to which degree centrality is concentrated within a small number of participants in a network, ranged between 0 and 1

There were two phases of this course: collaborative forum discussions during the first phase (Week 1—Week 8) and the small group collaborative writing during the second phase (Week 9—Week 16). The experiment carried out during the second phase, where students were grouped into 18 small groups to complete four collaborative writing tasks about emerging educational technology topics (2 weeks/task). Four collaborative writing tasks included “How Learning Analytics Influence the Learning Assessments”, “Artificial Intelligence Education: Opportunities and Challenges”, “STEM Education”, and “K-12 Programming Education”. These discussion topics were closely related to the course content. Students created small groups in DingTalk (a social communication tool) to discuss the topics (see Fig. 3a). Students conducted discussions within the DingTalk group in the synchronous or asynchronous ways. During the discussion, if the students responded to a partner’s points of view, they could use the @ symbol to specify the peer they responded to. The first author (as a teaching assistant) joined the groups to provide instructional and technical supports. The online collaborative writing platform was ZheDaYunPan (see Fig. 3b), where students can view, edit, and revise their peers’ writing content. During the collaborative process, the instructor and teaching assistant did not provide any instructional interventions related to the writing tasks.

The experimental design and procedures

The participants were divided into an experimental group (IVM_{GA}), control group of student-formed groups (GS), and control group of randomly-assigned groups (GR). Before grouping, an online questionnaire was delivered to students in order to collect students’ static characteristics, including gender, age, major, communication skill level, collaborative capacity level, and leadership level (see Appendix 1). A dynamic characteristic, namely students’ interactive levels, was measured by the social network analysis method, based on student degree (in-degree and out-degree) in the large group forum discussion (Week 1—Week 8). The instructor informed 68 students about the three types of grouping. IVM_{GA} created heterogeneous groups taking students’ static and dynamic characteristics as input; GR used a random grouping method to form groups by changing a parameter in the IVM_{GA} ; GS groups asked students to form their own groups autonomously. There were 6 IVM_{GA} groups, 8 GR

Table 3 The coding framework of discussion content (Garrison et al., 1999; Malmberg et al., 2017)

Code	Description
Cognitive engagement (CE)	Students’ sharing of information, as well as idea explanations and elaborations
Task regulation (TR)	Students’ task-related discourses, including planning, monitoring, management, and reflection
Social chatting (SC)	Students’ social chatting irrelevant to the collaborative topics, or students’ social encouragements to peers

Table 4 The comparison of performance among three grouping methods

Grouping method	Mean	Std	F	<i>p</i>	Multiple comparison
IVM _{GA}	81.45	5.48	4.16	0.02*	IVM _{GA} > GR
GR	70.38	23.71			GS > GR
GS	81.69	4.39			

* $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$

groups, and 4 GS groups. Most groups were configured with 4 members, but some groups had only 3 members. Specifically, there were four groups, namely group 9 (in IVM_{GA}) and groups 5, 10, and 12 (in GR), consisted of only three students (see Table 1).

Analysis procedures and strategies

We collected data from three sources, the collaborative writing documents from ZheDaYunPan, the collaborative discussion data from DingTalk, and the students' questionnaire data from an online questionnaire platform WJX (www.wjx.cn). First, to evaluate the collaborative writing documents, three external experts were invited to evaluate groups' final collaborative write-ups (four write-up documents for each group). An evaluation standard was provided to three experts, including two dimensions of content quality (85 points) and formats (15 points) (see Appendix 2). The result of Cohen's Kappa interrater reliability of three experts was $\alpha = 0.98$, which indicated a high reliability. The average score of four write-up files was used as an indicator to measure the groups' final performance.

Table 5 The comparison of performance among three grouping methods in four tasks

Tasks	Grouping method	Mean (Std.)	F	<i>p</i>	Multiple comparison
Task 1	IVM _{GA}	78.74 (11.69)	1.54	0.22	
	GR	75.91 (11.08)			
	GS	82.77 (10.23)			
Task 2	IVM _{GA}	83.83 (7.27)	3.31	0.04*	IVM _{GA} > GR
	GR	69.01 (27.94)			
	GS	80.88 (8.31)			
Task 3	IVM _{GA}	81.47 (7.95)	2.21	0.12	
	GR	75.96 (10.65)			
	GS	81.39 (9.21)			
Task 4	IVM _{GA}	81.77 (8.48)	4.72	0.01**	IVM _{GA} > GR GS > GR
	GR	60.64 (36.20)			
	GS	81.73 (8.19)			

* $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$

Second, to examine the collaborative process, this research focused on the social interaction and discussion content dimensions. This research used social network analysis (SNA) and quantitative content analysis (QCA) to analyze those two dimensions. On the social interaction dimension (see Table 2), this research used the group-level SNA metrics to uncover groups' social attributes, including *average path length (APL)*, *density*, *average degree*, *average closeness*, *average betweenness*, *reciprocity*, *global clustering coefficients (GCC)*, *the inverse coefficient of variation (ICV) of student interaction*, and *centralization* (see Ouyang, Chen, et al., 2021; Ouyang, Ling, et al., 2021; Ouyang & Scharber, 2017). Since students sometimes did not refer to a specific peer during the online discussions, it was difficult to identify the receiver; in this situation, the receiver was denoted as "all" when the network data was processed. For SNA metrics of *reciprocity* and *ICV* that needed to consider interactions between two identified members, we excluded the data of "all"

Table 6 The comparison of the social interaction of three grouping methods

SNA metrics	Grouping method	Mean value	F	p
APL	IVM _{GA}	1.10	1.34	0.27
	GR	1.02		
	GS	1.15		
Density	IVM _{GA}	5.49	1.03	0.36
	GR	6.19		
	GS	3.55		
Average degree	IVM _{GA}	22.97	0.58	0.56
	GR	22.68		
	GS	16.94		
Average closeness	IVM _{GA}	0.10	2.16	0.12
	GR	0.12		
	GS	0.07		
Average betweenness	IVM _{GA}	0.26	0.79	0.46
	GR	0.31		
	GS	0.44		
Reciprocity	IVM _{GA}	1.00	1.98	0.15
	GR	0.91		
	GS	1.00		
GCC	IVM _{GA}	0.57	3.39	0.04*
	GR	0.57		
	GS	0.32		
ICV	IVM _{GA}	1.27	1.84	0.17
	GR	1.06		
	GS	0.69		
Centralization	IVM _{GA}	0.18	0.89	0.42
	GR	0.22		
	GS	0.16		

* $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$

in the SNA measure processes. R packages *sna*, *igraph* and *tnet* were used to measure those SNA metrics.

On the discussion content dimension, a coding framework was used to analyze students' discussion content in DingTalk, including cognitive engagement (CE), task regulation (TR), and social chatting (SC) (see Table 3). The first author and another research assistant coded the discussion content based on the framework. They coded the content separately and reached a Cohen's Kappa's interrater reliability of 0.96.

Third, we designed an online questionnaire (with 5-Likert Scale), including six items about students' attitudes, emotions, and feedback (see Appendix 3) to collect student perceptions about the group's collaboration.

Finally, this research compared the effects of three grouping methods (i.e., IVM_{GA}, GR, GS) on the final performances, the collaborative processes, and students' perceptions through using the statistical analysis approaches (ANOVA, T-test). In addition, we also compared the IVM_{GA} and GR with two group sizes (i.e., size = 3 or size = 4); we excluded the GS groups since it had the same group size of 4. Overall, a multi-method approach (including statistic methods, social network analysis, quantitative content analysis) was used in this research.

Table 7 The comparison of the social interaction of IVM_{GA} and GR with two group sizes

SNA metrics	Size	IVM _{GA}		GR	
		Mean value	<i>p</i>	Mean value	<i>p</i>
APL	3	1.03	0.16	0.99	0.70
	4	1.12		1.04	
Density	3	7.45	0.30	9.94	0.04*
	4	5.10		3.94	
Average Degree	3	23.83	0.89	31.15	0.12
	4	22.80		17.60	
Average closeness	3	0.17	0.02*	0.14	0.27
	4	0.09		0.11	
Average betweenness	3	0.06	0.18	0.14	0.04*
	4	0.30		0.42	
Reciprocity	3	1.00	0.67	0.92	0.88
	4	1.00		0.90	
GCC	3	0.64	0.62	0.52	0.53
	4	0.55		0.60	
ICV	3	0.85	0.35	1.05	0.99
	4	1.35		1.06	
Centralization	3	0.08	0.19	0.22	0.96
	4	0.20		0.23	

* $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$

Table 8 The comparison of content analysis codes among three grouping methods

Content codes	Grouping method	Mean	Std	F	<i>p</i>
CE	IVM _{GA}	51.33	37.09	0.31	0.73
	GR	57.20	54.16		
	GS	46.63	32.12		
TR	IVM _{GA}	36.75	20.17	0.24	0.78
	GR	33.07	26.22		
	GS	27.50	37.80		
SC	IVM _{GA}	57.29	37.30	0.54	0.58
	GR	49.57	46.42		
	GS	47.80	63.16		

* $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$ **Table 9** The comparison of content analysis codes of IVM_{GA} and GR with two group sizes

Content codes	Size	IVM _{GA}		GR	
		mean	<i>p</i>	mean	<i>p</i>
CE	3	21.25	0.00***	66.33	0.51
	4	57.35		51.11	
TR	3	31.75	0.65	31.00	0.75
	4	37.75		34.44	
SC	3	64.25	0.77	62.58	0.29
	4	55.90		40.89	

* $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$

Results

To what extent did three grouping methods differentiate in terms of the groups' final performance?

ANOVA results showed that there were significant differences among three grouping methods (see Table 4). The result showed that the IVM_{GA} and GS groups achieved a significant higher score compared with the GR groups ($p = 0.02$). However, there was no statistical significance between IVM_{GA} and GS groups, where GS groups, in average, achieved a slightly higher score than the IVM_{GA} groups. The standard deviation in the GR groups was the largest among three grouping methods (Std. = 23.71), indicating that the final performance of the GR groups was highly polarized, with some groups performed extremely well and others performed poor.

Table 5 showed the result of comparing the average performances among three different grouping methods under the same writing task. The result showed that the GS groups had the highest average score in Task 1, while the IVM_{GA} groups had the highest average score in Task 2, Task 3, and Task 4. In Task 2 and Task 4, significant differences were detected among three grouping methods.

Table 10 Comparison of student perceptions of three grouping methods

Item	Grouping method	Mean value	Std	F	<i>p</i>
Q1	IVM _{GA}	3.76	0.70	0.05	0.95
	GR	3.73	0.60		
	GS	3.80	0.68		
Q2	IVM _{GA}	3.10	0.30	0.56	0.57
	GR	3.08	0.39		
	GS	3.20	0.41		
Q3	IVM _{GA}	4.33	0.80	3.58	0.03*
	GR	3.73	1.04		
	GS	4.33	0.62		
Q4	IVM _{GA}	4.10	1.09	1.18	0.32
	GR	3.65	0.98		
	GS	3.93	0.88		
Q5	IVM _{GA}	3.33	0.80	0.56	0.57
	GR	3.35	0.89		
	GS	3.60	0.74		
Q6	IVM _{GA}	3.43	0.75	1.23	0.30
	GR	3.42	1.14		
	GS	3.87	0.83		

* $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$

Moreover, this research also examined whether there were differences among groups with two different sizes (i.e., group size = 3 or group size = 4). No statistical differences were identified in the GR and IVM_{GA} groups. Analyses were also carried out to assess whether there were significances in GR and IVM_{GA} groups with two different group sizes on each task. There was no significant difference in the IVM_{GA} groups in the four tasks, while the GR groups showed a significant difference of performance in Task 2.

To what extent did three grouping methods differentiate in terms of the social and cognitive engagement during the collaborative processes?

The social interaction dimension

ANOVA was used to assess whether there were differences of the SNA measurements among three grouping methods. Table 6 showed that there were no significant differences in most SNA measurements, except *GCC* ($p = 0.04$). The IVM_{GA} and GR groups had the same value of *GCC* (*value* = 0.57), followed by GS groups (*GCC value* = 0.32). Although there were no statistical significances in other SNA measurements, three grouping methods had different SNA results. Specifically, the GR groups had the shortest APL (*value* = 1.02), the highest density (*value* = 6.19) and closeness (*value* = 0.12); the GS groups had the lowest centralization (*value* = 0.16) and closeness (*value* = 0.07), and the highest betweenness

(*value* = 0.44) and reciprocity (*value* = 1); the IVM_{GA} groups had the highest degree (*value* = 22.97), reciprocity (*value* = 1), and *ICV* (*value* = 1.27).

Furthermore, this research used T-tests to examine whether there were statistical differences in the SNA measurements between two group sizes under the GR and IVM_{GA} , respectively. Table 7 showed that in the IVM_{GA} groups, there was a statistical significance of closeness ($p = 0.02$) between the two group sizes. The groups with three members had higher closeness centrality than the groups with the size of four members. In the GR groups, two measurements, namely density ($p = 0.04$) and betweenness ($p = 0.04$), had significant differences between two different group sizes. The groups with three members had a higher density (*value* = 9.94), and the groups with four members had a higher betweenness centrality (*value* = 0.42). Overall, only one metric revealed statistical significance in the IVM_{GA} groups, whereas two metrics had statistical significance in the GR groups. Therefore, the IVM_{GA} grouping method had a competitive advantage in solving the unbalanced group size problem based on the comparison of SNA measurements. The IVM_{GA} method still maintained the approximate capacity among groups in the collaborative process with different group sizes.

The discussion content dimension

The results indicated that there was no significant difference among three grouping methods on CE, TR and SC (see Table 8). Among three grouping methods, the IVM_{GA} groups had a medium level of cognitive engagement (CE = 51.33), the highest level of task regulation (TR = 36.75) and social chatting (SC = 57.29). In addition, for TR and SC codes, the standard deviation values of IVM_{GA} groups were the lowest, indicating a balanced attribute within IVM_{GA} groups on those two codes.

Table 9 indicated there was a significant difference in IVM_{GA} groups with different group sizes on CE. The mean value of CE within groups of three members was significantly lower than groups of four members.

To what extent did three grouping methods differentiate in terms of students' perceptions?

The reliability test of questionnaire responses was the Coefficient value = 0.80, indicating a high reliability. Table 10 showed the average score and standard deviation of each questionnaire question and then revealed the one-way ANOVA analysis results. There was only one item (Q3), regarding the satisfaction of the grouping result, had a significant difference ($p = 0.03$). The result indicated that students were satisfied with the IVM_{GA} and GS grouping methods. Although there was no significant difference in Q4 regarding fluency, the IVM_{GA} groups reported the highest value of fluency, which indicated that conflicts can be resolved and tasks can be coordinated more efficiently within the groups. Furthermore, different group sizes did not lead to differences in student perceptions of the collaborative experiences, either under the IVM_{GA} or GR groups.

Discussions

Addressing research questions

To address the two issues of uneven group sizes and the inaccessibility of student characteristics, this research proposed an optimized, genetic algorithm-based grouping method and conducted a quasi-experiment research. Results showed that the experimental groups outperformed the traditional grouping methods (i.e., random groups and student-formed groups) in terms of the groups' final performances, collaborative processes, and student perceptions. Regarding the first research question, the result showed that in the final collaborative writing performance, the IVM_{GA} groups outperformed the GR groups to a statistically significant extent, which was consistent with other research using the GA-enabled grouping methods (e.g., Chen & Kuo, 2019; Moreno et al., 2012; Takači et al., 2017). In addition, the GR groups' standard deviation values were highly polarized, indicating the unbalanced performance results in GR groups. Further analysis revealed two groups in GR had extremely poor performances, namely G11 (4-person group) and G12 (3-person group). The poor performance of these two groups in task 2 and task 4 caused the unbalanced performance issue. The results justified that the GR method was more likely to generate poor performing groups, compared to other two grouping methods. Moreover, although the IVM_{GA} groups had different group sizes, this difference did not impair their group performances. When we compared the 3-person groups in IVM_{GA} and GR, there were no significant differences; when both the 3-person and 4-person groups were taken into account, a significant difference emerged between the grouping methods of IVM_{GA} and GR. The reason was that the IVM_{GA} strategy minimized the differences of groups with different sizes, while the GR grouping method could not achieve similar effects (Holenko Dlab et al., 2020). Overall, IVM_{GA} was conducive to improve the groups' collaborative performances while did not weaken the group performances due to differences of the group sizes.

Regarding the second research question, the comparison of the groups' collaborative processes provided additional evidence to support the advantage of our proposed grouping method. On the social interaction dimension, the results showed that the IVM_{GA} and GR groups had the highest *GCC* value, indicating the high levels of connectedness. A possible reason was that unfamiliar students in the IVM_{GA} and GR groups tended to create connections with each other, while familiar students in GS groups were likely to interact with their acquaintance but tended not to sustain interactions with other members in the group (Gazelle et al., 2005). The highest metrics of *degree* and *ICV* also indicated that the IVM_{GA} groups had the most active, balanced, evenly-distributed interaction because our grouping method mixed the different levels of students to avoid the formation of the extremely unbalanced groups. The *ICV* results indicated that the IVM_{GA} method achieved a competitive advantage in keeping a balanced group when the group sizes were different. Therefore, the IVM_{GA} was beneficial to cultivate collaborative interactions and minimize the groups' differences, particularly when groups had different sizes. On the discussion content dimension, no significant differences were found on cognitive engagement

(CE), task regulation (TR), and social chatting (SC) among three grouping methods. Compared to other two grouping methods, IVM_{GA} groups achieved relatively low scores of standard deviations, which indicated IVM_{GA} groups kept a certain level of balance within groups. But the balance did not maintain in the IVM_{GA} groups when we compared between the different group sizes. Overall, from a collaborative process perspective, the IVM_{GA} grouping method did facilitate groups to achieve better social interaction and cognitive engagement during the collaborative processes.

Regarding the third research question, the questionnaire results provided positive feedback about student perceptions. IVM_{GA} and GS groups significantly exceeded GR on students' satisfaction level, which was consistent with previous studies (e.g., Lin et al., 2016). Students' satisfaction had critical influence on their willingness to collaborate in a group, since if group members engaged in collaborative tasks with a negative attitude, it was almost impossible to acquire preferred outcomes (Lavy, 2017; Reis et al., 2018). Moreover, fluency was commonly used for evaluating the group performance in creative tasks (Clark & Mirels, 1970). IVM_{GA} groups reported the highest score of fluency, which indicated that conflicts could be resolved and tasks could be coordinated more efficiently within the groups. In addition, different group sizes did not lead to differences of student perceptions of the collaborative experiences. Therefore, the IVM_{GA} grouping method was beneficial to improving students' satisfaction and collaborative fluency. Overall, the proposed grouping method indicated competitive advantages to improve the group collaboration performances, increase collaborative engagement, and improve student satisfaction about the group work.

Implications for the implementation of grouping methods

In this research, two common challenges namely the uneven group sizes and inaccessibility of student characteristics were resolved to enhance the adaptability of the grouping methods. In order to further extend the study of GA-enabled grouping methods, three implications were proposed on the implementation of automatic grouping methods. First, regarding the input variables of student characteristics, the essential of heterogeneous grouping is the mixture of participants' characteristics at multiple dimensions. The selection of group characteristics is a complicated issue, since we cannot ensure that all characteristics that would impact the group performance were considered. A combination of static and dynamic characteristics from students not only achieve better collaborative effect than merely using one type of student characteristics (Krouska & Virvou, 2020; Moreno et al., 2012), but also facilitates the solution of "Cold Start" problems. Our research used social interaction metrics as the dynamic student characteristics; future research can consider other dynamic characteristics such as the changing of cognitive or emotional levels. In addition, a continuous numerical variable would show a certain level of flexibility in a complicated educational environment and decrease the unbalanced characteristics distribution. Second, the heuristic search algorithm aims to search for an optimized solution from a larger solution space in a limited period of time (Cruz & Isotani, 2014). The IVM_{GA} grouping method proposed in this research can efficiently generate the optimized groups, compared to traditional grouping methods. IVM_{GA} still has some disadvantages, particularly as the numbers

of groups and the student characteristics increase, the solution set space of the algorithm becomes very large and it is impossible to iterate the whole solution set space in the limited time (Lambić et al., 2018; Sukstrienwong, 2017; Takači et al., 2017). Future work can consider combine some other algorithms (e.g., genetic programming) to reduce the dimensionality of the student features and improve the computational efficiency. Third, this research requires the researchers to process the student characteristics data and manually used the grouping algorithm to generate groups. Future implementations should integrate the grouping function in online learning platforms to support automatic grouping in online or blended learning (Alfonseca et al., 2006). Without the technical support, it would be extremely time-consuming and labor-intensive for the instructor to conduct the grouping work (Sukstrienwong, 2017; Takači et al., 2017). With the support of automatic groups in online platforms, the instructors can further modify groups with other factors such as task types, difficulty levels, and course content (Wang et al., 2007). Moreover, online platforms can help instructors to automatically acquire student data via drop-in quiz and online questionnaire tools and convert discussion content information into dynamic social interactive characteristic data (e.g., Chen et al., 2018; Ouyang, Chen, et al., 2021; Ouyang, Ling, et al., 2021). In summary, future design and implementation of grouping methods should consider three critical factors, namely input variables of student characteristics, choice of other algorithms and integration of grouping function in online platforms.

Implications for collaborative analytics

Most previous studies focus on the evaluation of groups' summative performance rather than the process-oriented analytics (Lambić et al., 2018; Lin et al., 2016; Moreno et al., 2012; Takači et al., 2017), which may result in a lack of a comprehensive, holistic understanding of the CSCL process. There is an analytical trend in the learning science and collaborative learning fields to examine both performance and process data and to reveal the multi-dimensional characteristics of CSCL (Janssen et al., 2013; Schneider et al., 2021). The multi-method approaches have been used to conduct the multi-dimensional analyses of collaborative learning processes, such as using the statistical analysis, sequential analysis, and social network analysis approaches to investigate the sequences of students' knowledge contributions (Chen et al., 2018) and social interaction structures (Ouyang, 2021; Ouyang & Chang, 2019; Ouyang & Scharber, 2017). Considering the multi-dimensional characteristics of the collaborative learning, merely focusing on the final performance of collaboration may cause inconclusive and incomprehensive understandings. Therefore, studies have used a multi-method approach to complement each other in order to provide a more holistic, multilevel, multidimensional analysis of the collaborative process. For example, Chen and Kuo (2019) and Joksimović et al. (2018) used a multi-method approach to reveal the depth of the learning process and the experiment details to enhance the validation of the conclusion. Taking a step forward, this research took cognitive engagement analysis as an additional supplement to further explore the collaborative process. Furthermore, Schneider et al. (2021) stated that collaborative learning analytics is the intersection of collaborative learning and learning analytics, which expands the learning analytics method to focus on analytics

of groups of students. Overall, using multiple learning analytics methods, collaborative analytics can provide feedback on similarities and differences of CSCL results at the group level to improve the understanding of groups' collaborative work.

Conclusions, limitations, and future directions

One of the core issues in CSCL is grouping formation, because a group's learning atmospheres, processes, and performances can be determined by how well the group members work together (Chen & Kuo, 2019; Reis et al., 2018; Uto et al., 2020). GA-enabled grouping algorithms have been used to cope with the complexity and diversity of educational contexts (Chen & Kuo, 2019; Krouska & Virvou, 2020; Moreno et al., 2012), such as enhancing the grouping optimization effect, improving the computational efficiency, and meeting the diversity of educational needs. However, two common challenges, including the uneven group size problem and unavailability of student characteristics, weaken the effects of the grouping methods in practices. To overcome two challenges, this research proposes a feature categorization module and an enhanced genetic algorithm module, and further applied this proposed method in an educational context where group sizes could not be equally divided. The experiment result shows that the experimental groups outperformed the random or self-selected grouping methods. And the experimental groups maintained a relatively balanced capacity to complete collaborative tasks, kept an active interaction among peers, and perceived a satisfied collaborative perception.

Future directions of this line of inquiry include algorithm optimizations, integration in online platforms, and further empirical validation in different scenarios using collaborative analytics methods. First, future research can focus on improving grouping algorithm's computing efficiency by using advanced algorithms to perform dimensionality reduction and extract key characteristics as input in the algorithm model (Jolliffe & Cadima, 2016). Second, this research merely applied the grouping methods in an undergraduate course; further research should expand the educational activities to broader instructional scenarios to further test the effect of this grouping method on collaborative learning. Although we used multiple rounds of tasks to enhance its representativeness, future studies should increase the number of sample sizes and expand the educational activities to broader instructional scenarios in order to test and verify the effectiveness of grouping methods. Third, in addition to evaluations of final performance, future work should use multiple learning analytics to conduct the collaborative process-oriented analytics in order to enhance the interpretation of group work (Dindar et al., 2020; Schneider et al., 2021).

Appendix 1

See Table 11.

Table 11 The pre-course questionnaire of student information

Questionnaire

Basic information

1. Name 2. Gender 3. Age 4. Major

Communicative skill

5. When my thoughts are different from others, I still express my points of view
 6. I am open with other people's points of view about my thoughts
 7. I often try to figure out what other people is going to say before they finish the conversation
 8. When I don't understand what other people is saying, I will take the initiative to ask questions
 9. In the conversation, I pay attention to the facts and details as well as the emotional tone of the speaker
 10. When communicating with other people, I tend to think from the perspectives of others
 11. I am willing to listen to or accept constructive criticism from others
 12. I will avoid saying things that I think will make others unhappy
 13. When there is a disagreement, I can still communicate objectively and rationally with others
 14. When I know that I am wrong, I am willing to admit my mistakes

Collaborative skill

15. I like to collaborate with my peers, and I am willing to actively initiate collaboration with my peers
 16. I think teamwork allows me to better organize my ideas (such as refining my opinions)
 17. I feel that teamwork promotes my understanding of others and my respects of others
 18. I think teamwork should encourage all members to actively participate in the activity
 19. I think people are equal individuals when exploring information, constructing knowledge, or completing tasks together in a group
 20. When evaluating other members' work in a group, I usually express my opinions with reasons, rather than merely expressing my intuitions
 21. When completing tasks with my peers, I am willing to listen to my peers' opinions, even if I disagree with these opinions
 22. I think that the completion of any work is inseparable from the help and collaboration of others
 23. When discussing group work, I am willing to actively put forward my own views, ideas or plans
 24. Although teamwork is sometimes time-consuming, it can help produce the high-quality results in the long run

Leadership skill

25. I can usually understand the reaction of others to an opinion in advance
 26. I am very good at solving problems
 27. It is very easy for me to make plans and deal with details
 28. Management is my strengths
 29. I enjoy responding to other people's requests and concerns
 30. I like to make plans for the team
 31. I think the key to successfully handling conflicts is to respect each other
 32. I am good at obtaining resources to support the collaborative projects
 33. I try my best to find consensus in different situations
 34. I can flexibly adjust myself when the teamwork needs some changes
-

Appendix 2

See Table 12.

Table 12 The assessment standard for the collaborative writing documents

Content Quality Score (85 points)	Non-Content Quality Score (15 points)
Clear logic in writing content (20 points)	Document layout and structure (5 points)
Breadth and depth of topic expansions (20 points)	Formats of tables and Figs. (5 points)
Argumentations and viewpoints (15 points)	References (5 points)
Sufficient evidence and examples (25 points)	
Grammar (5 points)	

Appendix 3

See Table 13.

Table 13 The after-course questionnaire

No	Question (1 indicated “extremely unsatisfied”, 2 indicated “unsatisfied”, 3 indicated “neutral”, 4 indicated “satisfied”, 5 indicated “extremely satisfied”)
Q1	Are you satisfied with your level of mastery of the writing content after collaboration?
Q2	Are you satisfied with your group size?
Q3	Are you satisfied with the results of grouping?
Q4	Are you satisfied with the fluency of communication and collaboration in your team?
Q5	Are you satisfied with the quality of your group’s collaborative work?
Q6	Are you satisfied with the overall collaborative experiences during four tasks in your group?

Acknowledgements The research is financially supported by the National Natural Science Foundation of China (62177041), 2021 Key Research and Development Plan of Zhejiang Province (2021C03140), and Zhejiang Province Educational Science Planning and Research Project (2022SCG256).

Declarations

Conflict of interest We have no Conflict of Interest to declare.

References

- Ahmad, A., Zeeshan, F., Marriam, R., Samreen, A., & Ahmed, S. (2021). Does one size fit all? Investigating the effect of group size and gamification on learners’ behaviors in higher education. *Journal of Computing in Higher Education*, 33(2), 296–327. <https://doi.org/10.1007/s12528-020-09266-8>

- Alfonseca, E., Carro, R. M., Martín, E., Ortigosa, A., & Paredes, P. (2006). The impact of learning styles on student grouping for collaborative learning: A case study. *User Modeling and User-Adapted Interaction*, 16(3), 377–401. <https://doi.org/10.1007/s11257-006-9012-7>
- Chan, T., Chen, C.-M., Wu, Y.-L., Jong, B.-S., Hsia, Y.-T., & Lin, T.-W. (2010). Applying the genetic encoded conceptual graph to grouping learning. *Expert Systems with Applications*, 37(6), 4103–4118. <https://doi.org/10.1016/j.eswa.2009.11.014>
- Chen, B., Chang, Y. H., Ouyang, F., & Zhou, W. (2018). Fostering student engagement in online discussion through social learning analytics. *The Internet and Higher Education*, 37, 21–30. <https://doi.org/10.1016/j.iheduc.2017.12.002>
- Chen, C.-M., & Chang, C.-C. (2014). Mining learning social networks for cooperative learning with appropriate learning partners in a problem-based learning environment. *Interactive Learning Environments*, 22(1), 97–124. <https://doi.org/10.1080/10494820.2011.641677>
- Chen, C.-M., & Kuo, C.-H. (2019). An optimized group formation scheme to promote collaborative problem-based learning. *Computers & Education*, 133, 94–115. <https://doi.org/10.1016/j.compedu.2019.01.011>
- Chen, R.-C., Chen, S.-Y., Fan, J.-Y., & Chen, Y.-T. (2012). Grouping partners for cooperative learning using genetic algorithm and social network analysis. *Procedia Engineering*, 29, 3888–3893.
- Clark, P. M., & Mirels, H. L. (1970). Fluency as a pervasive element in the measurement of creativity. *Journal of Educational Measurement*, 7(2), 83–86.
- Cohen, L., Manion, L., & Morrison, K. (2013). *Research methods in education*. Routledge.
- Costaguta, R. (2015). Algorithms and machine learning techniques in collaborative group formation. *Advances in Artificial Intelligence and Its Applications*. https://doi.org/10.1007/978-3-319-27101-9_18
- Cruz, W. M., & Isotani, S. (2014). Group formation algorithms in collaborative learning contexts: A systematic mapping of the literature. *Collaboration and Technology*. https://doi.org/10.1007/978-3-319-10166-8_18
- Damşa, C. I. (2014). The multi-layered nature of small-group learning: Productive interactions in object-oriented collaboration. *International Journal of Computer-Supported Collaborative Learning*, 9(3), 247–281.
- Delaney, D., Kummer, T.-F., & Singh, K. (2019). Evaluating the impact of online discussion boards on student engagement with group work. *British Journal of Educational Technology*, 50(2), 902–920.
- Dindar, M., Järvelä, S., & Haataja, E. (2020). What does physiological synchrony reveal about meta-cognitive experiences and group performance? *British Journal of Educational Technology*, 51(5), 1577–1596.
- Garrison, D. R., Anderson, T., & Archer, W. (1999). Critical inquiry in a text-based environment: Computer conferencing in Higher Education. *The Internet and Higher Education*, 2(2), 87–105. [https://doi.org/10.1016/S1096-7516\(00\)00016-6](https://doi.org/10.1016/S1096-7516(00)00016-6)
- Gazelle, H., Putallaz, M., Li, Y., Grimes, C. L., Kupersmidt, J. B., & Coie, J. D. (2005). Anxious solitude across contexts: Girls' interactions with familiar and unfamiliar peers. *Child Development*, 76(1), 227–246.
- Gibbs, W. J., & Bernas, R. S. (2007). Computer-mediated-communications, learning style, and visualizing online educational conversations. *Journal of Computing in Higher Education*, 18(2), 25–50. <https://doi.org/10.1007/BF03033412>
- Gibbs, W. J., Chen, C., & Bernas, R. S. (2001). Group instruction and web-based instructional approaches for training student employees. *Journal of Computing in Higher Education*, 13(1), 71–90. <https://doi.org/10.1007/BF02940945>
- Hernández-Sellés, N., Muñoz-Carril, P.-C., & González-Sanmamed, M. (2019). Computer-supported collaborative learning: An analysis of the relationship between interaction, emotional support and online collaborative tools. *Computers & Education*, 138, 1–12.
- Holenko Dlab, M., Boticki, I., Hoic-Bozic, N., & Looi, C. K. (2020). Exploring group interactions in synchronous mobile computer-supported learning activities. *Computers & Education*, 146, 103735.
- Huxham, M., & Land, R. (2000). Assigning students in group work projects. Can we do better than random? *Innovations in Education and Training International*, 37(1), 17–22.
- Janssen, J., Cress, U., Erkens, G., & Kirschner, P. A. (2013). Chapter 6 Multilevel analysis for the analysis of collaborative learning. In C. E. Hmelo-Silver, C. A. Chinn, C. K. K. Chan, & A. M. O'Donnell (Eds.), *The international handbook of collaborative learning* (pp. 124–137). Routledge.

- Jeong, H., Hmelo-Silver, C. E., & Jo, K. (2019). Ten years of computer-supported collaborative learning: A meta-analysis of CSCL in STEM education during 2005–2014. *Educational Research Review*, 28, 100284.
- Joksimović, S., Poquet, O., Kovanović, V., Dowell, N., Mills, C., Gašević, D., Dawson, S., Graesser, A. C., & Brooks, C. (2018). How do we model learning at scale? A systematic review of research on MOOCs. *Review of Educational Research*, 88(1), 43–86.
- Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), 20150202.
- Jong, B., Wu, Y., & Chan, T. (2006). Dynamic grouping strategies based on a conceptual graph for cooperative learning. *IEEE Transactions on Knowledge and Data Engineering*, 18(6), 738–747.
- Krouska, A., & Virvou, M. (2020). An enhanced genetic algorithm for heterogeneous group formation based on multi-characteristics in social networking-based learning. *IEEE Transactions on Learning Technologies*, 13(3), 465–476.
- Kumar, M., Husain, D. M., Upreti, N., & Gupta, D. (2010). Genetic algorithm: Review and application. *Social Science Electronic Publishing*. <https://doi.org/10.2139/ssrn.3529843>
- Lambić, D., Lazović, B., Djenić, A., & Marić, M. (2018). A novel metaheuristic approach for collaborative learning group formation. *Journal of Computer Assisted Learning*, 34(6), 907–916.
- Lavy, S. (2017). Who benefits from group work in higher education? An attachment theory perspective. *Higher Education*, 73(2), 175–187.
- Lika, B., Kolomvatsos, K., & Hadjiefthymiades, S. (2014). Facing the cold start problem in recommender systems. *Expert Systems with Applications*, 41(4), 2065–2073.
- Lin, Y., Chang, Y., & Chu, C. (2016). Novel approach to facilitating tradeoff multi-objective grouping optimization. *IEEE Transactions on Learning Technologies*, 9(2), 107–119. <https://doi.org/10.1109/TLT.2015.2471995>
- Lin, Y.-T., Huang, Y.-M., & Cheng, S.-C. (2010). An automatic group composition system for composing collaborative learning groups using enhanced particle swarm optimization. *Computers & Education*, 55(4), 1483–1493.
- Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B., & d'Apollonia, S. (1996). Within-Class grouping: A meta-analysis. *Review of Educational Research*, 66(4), 423–458.
- Ludvigsen, S. (2016). CSCL: Connecting the social, emotional and cognitive dimensions. *International Journal of Computer-Supported Collaborative Learning*, 11(2), 115–121. <https://doi.org/10.1007/s11412-016-9236-4>
- Malmberg, J., Järvelä, S., & Järvenoja, H. (2017). Capturing temporal and sequential patterns of self-, co-, and socially shared regulation in the context of collaborative learning. *Contemporary Educational Psychology*, 49, 160–174. <https://doi.org/10.1016/j.cedpsych.2017.01.009>
- 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.
- Ouyang, F. (2021). Using three social network analysis approaches to understand computer-supported collaborative learning. *Journal of Educational Computing Research*, 59(7), 1401–1424. <https://doi.org/10.1177/0735633121996477>
- Ouyang, F., & Chang, Y. H. (2019). The relationship between social participatory role and cognitive engagement level in online discussions. *British Journal of Educational Technology*, 50(3), 1396–1414. <https://doi.org/10.1111/bjet.12647>
- Ouyang, F., Chen, S., & Li, X. (2021a). Effect of three network visualizations on students' social-cognitive engagement in online discussions. *British Journal of Educational Technology*, 52(6), 2242–2262. <https://doi.org/10.1111/1467-8535.13126>
- Ouyang, F., Ling, T., & Jiao, P. (2021b). Development of group cognition in online collaborative problem-solving processes. *Journal of Educational Computing Research*. <https://doi.org/10.1177/07356331211047784>
- Ouyang, F., & Scharber, C. (2017). The influences of an experienced instructor's discussion design and facilitation on an online learning community development: A social network analysis study. *The Internet and Higher Education*, 35, 34–47. <https://doi.org/10.1016/j.iheduc.2017.07.002>

- Pliakos, K., Joo, S.-H., Park, J. Y., Cornillie, F., Vens, C., & Van den Noortgate, W. (2019). Integrating machine learning into item response theory for addressing the cold start problem in adaptive learning systems. *Computers & Education*, 137, 91–103.
- Qiu, M., & McDougall, D. (2015). Influence of group configuration on online discourse reading. *Computers & Education*, 87, 151–165.
- Reis, R. C. D., Isotani, S., Rodriguez, C. L., Lyra, K. T., Jaques, P. A., & Bittencourt, I. I. (2018). Affective states in computer-supported collaborative learning: Studying the past to drive the future. *Computers & Education*, 120, 29–50.
- Sadeghi, H., & Kardan, A. A. (2016). Toward effective group formation in computer-supported collaborative learning. *Interactive Learning Environments*, 24(3), 382–395.
- Sawyer, R. K. (Ed.). (2005). *The Cambridge handbook of the learning sciences*. Cambridge University Press.
- Schein, A. I., Popescul, A., Ungar, L. H., & Pennock, D. M. (2002). Methods and metrics for cold-start recommendations. In M., Beaulieu, R., Baeza-Yates, S. H., Myaeng, & K., Jarvelin (Eds.), *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information retrieval* (pp. 253–260). ACM.
- Schneider, B., Dowell, N., & Thompson, K. (2021). Collaboration analytics—current state and potential futures. *Journal of Learning Analytics*, 8(1), 1–12.
- Soller, A., Goodman, B., Linton, F., & Gaimari, R. (1998). Promoting effective peer interaction in an intelligent collaborative learning system. In B. P. Goettl, H. M. Half, C. L. Redfield, & V. J. Shute (Eds.), *Intelligent Tutoring Systems* (pp. 186–195). Springer. https://doi.org/10.1007/3-540-68716-5_24
- Srba, I., & Bielikova, M. (2015). Dynamic group formation as an approach to collaborative learning support. *IEEE Transactions on Learning Technologies*, 8(2), 173–186. <https://doi.org/10.1109/TLT.2014.2373374>
- Stahl, G. (2009a). Introduction to CSCL. In: U. Cress, C. Rosé, A. F. Wise, & J. Oshima (Eds.), *Proceedings of the 9th International Conference on Computer Supported Collaborative Learning* (pp.211–211). Springer.
- Stahl, G. (2009c). Introduction to CSCL. In: *Proceedings of the 9th International Conference on Computer Supported Collaborative Learning - CSCL'09*, 2, 211–211.
- Stahl, G. (2009b). Introduction to CSCL. In: *Proceedings of the 9th International Conference on Computer Supported Collaborative Learning* (pp. 211–211). Rhodes, Greece: ACM.
- Stahl, G., Koschmann, T., & Suthers, D. (2014). Computer-supported collaborative learning. In R. Keith Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (pp. 479–500). Cambridge University Press. <https://doi.org/10.1017/CBO9781139519526.029>
- Sukstrienwong, A. (2017). A genetic-algorithm approach for balancing learning styles and academic attributes in heterogeneous grouping of students. *International Journal of Emerging Technologies in Learning (IJET)*, 12(03), 4–25. <https://doi.org/10.3991/ijet.v12i03.5803>
- Suthers, D., Lund, K., Rosé, C. P., Teplov, C., & Law, N. (2013). *Productive multivocality in the analysis of group interactions*. Springer.
- Takači, Đ, Marić, M., Stankov, G., & Djeniđ, A. (2017). Efficiency of using VNS algorithm for forming heterogeneous groups for CSCL learning. *Computers & Education*, 109, 98–108. <https://doi.org/10.1016/j.compedu.2017.02.014>
- Uto, M., Nguyen, D.-T., & Ueno, M. (2020). Group optimization to maximize peer assessment accuracy using item response theory and integer programming. *IEEE Transactions on Learning Technologies*, 13(1), 91–106. <https://doi.org/10.1109/TLT.2019.2896966>
- Van der Laan Smith, J., & Spindle, R. M. (2007). The impact of group formation in a cooperative learning environment. *Journal of Accounting Education*, 25(4), 153–167. <https://doi.org/10.1016/j.jacce.2007.09.002>
- Wang, D.-Y., Lin, S. S. J., & Sun, C.-T. (2007). DIANA: A computer-supported heterogeneous grouping system for teachers to conduct successful small learning groups. *Computers in Human Behavior*, 23(4), 1997–2010. <https://doi.org/10.1016/j.chb.2006.02.008>
- Wang, S.-L., & Hwang, G.-J. (2012). The role of collective efficacy, cognitive quality, and task cohesion in computer-supported collaborative learning (CSCL). *Computers & Education*, 58(2), 679–687. <https://doi.org/10.1016/j.compedu.2011.09.003>

- Yannibelli, V., & Amandi, A. (2012). A deterministic crowding evolutionary algorithm to form learning teams in a collaborative learning context. *Expert Systems with Applications*, 39(10), 8584–8592. <https://doi.org/10.1016/j.eswa.2012.01.195>
- Yilmaz, R., Karaoglan Yilmaz, F. G., & Keser, H. (2020). Vertical versus shared e-leadership approach in online project-based learning: A comparison of self-regulated learning skills, motivation and group collaboration processes. *Journal of Computing in Higher Education*, 32(3), 628–654. <https://doi.org/10.1007/s12528-020-09250-2>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.