An Exact Algorithm for Group Formation to Promote Collaborative Learning

Zhiru Sun
zhiru@sdu.dk
Department of Design and Communication, University of Southern Denmark
Kolding, Denmark

Marco Chiarandini
marco@imada.sdu.dk
Department of Mathematics and Computer Science, University of Southern Denmark
Odense, Denmark

ABSTRACT

Collaborative learning has been widely used to foster students' communication and joint knowledge construction. However, the classification of learners into well-structured groups is one of the most challenging tasks in the field. The aim of this study is to propose a novel method to form intra-heterogeneous and inter-homogeneous groups based on relevant student characteristics. Such a method allows for the consideration of multiple student characteristics and can handle both numerical and categorical characteristic types simultaneously. It assumes that the teacher provides an order of importance of the characteristics, then it solves the grouping problem as a lexicographic optimization problem in the given order. We formulate the problem in mixed integer linear programming (MILP) terms and solve it to optimality. A pilot experiment was conducted with 29 college freshmen considering three general characteristics (i.e., 13 specific features) including knowledge level, demographic information, and motivation. Results of such an experiment demonstrate the validity and computational feasibility of the algorithmic approach. Large-scale studies are needed to assess the impact of the proposed grouping method on students' learning experience and academic achievement.

CCS CONCEPTS

• **Applied computing** → *Operations research*; **Collaborative learning**.

KEYWORDS

Group formation; Computer-supported collaborative learning (CSCL); Mixed integer linear programming (MILP); student-project assignment

ACM Reference Format:
Zhiru Sun and Marco Chiarandini. 2021. An Exact Algorithm for Group Formation to Promote Collaborative Learning. In *LAK21: 11th International Learning Analytics and Knowledge Conference (LAK21), April 12–16, 2021, Irvine, CA, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3448139.3448196

LAK21, April 12–16, 2021, Irvine, CA, USA

ACM ISBN 978-1-4503-8935-8/21/04…$15.00
https://doi.org/10.1145/3448139.3448196

1 INTRODUCTION

Collaborative learning is known to be an effective teaching method to engage learners through communication with other group members and the exchange of diverse ideas for joint knowledge construction [Liu and Tsai 2008]. It has been widely implemented in both traditional and online contexts [Radkowitsch et al. 2020]. However, simply putting learners in groups cannot guarantee the success of collaborative learning. Thus, group formation is considered the most important step in planning collaborative activities. It influences how learners approach learning tasks, perceive the environment, interact with peers, and take part in the collaborative learning processes [Hadwin et al. 2018]. However, the classification of learners into well-functioning groups is one of the most challenging tasks in the field of collaborative learning.

A line of research regarding group composition has categorized collaborative groups into two major types based on the within-group composition, which is *homogeneous group* (i.e., learners within a group having similar ability levels) and *heterogeneous group* (i.e., learners within a group having dissimilar ability levels) [Murphy et al. 2017]. Various studies have compared the two types of groups on learners' achievement and social interaction and suggested a result favoring the heterogeneous group as the best choice [Murphy et al. 2017]. Researchers believe that, compared to homogeneous groups, learners in heterogeneous groups tend to coordinate and create common ground faster and easier because the diverse skills and characteristics of group members might be complementary to each other and the group as a whole [Lou et al. 1996; Manske et al. 2015]. In addition, from the economics perspective, the level of equality or fairness in heterogeneous groups is higher than that in homogeneous groups because resources (e.g., time, knowledge) are more likely to be equally distributed to each group member instead of being collected by a single or limited number of group members [Fallucchi et al. 2018].

In practice, there are three major methods to form groups: *random grouping* (i.e., assigning learners in the groups by chance), *self-selected grouping* (i.e., the learners choose with whom they want to work with), and *controlled grouping* (i.e., assigning learners in the groups by instructors or computing systems based on certain criteria) [Chan et al. 2010; Chen and Kuo 2019; Hilton and Phillips 2010]. Among them, the random grouping method can generate either homogeneous or heterogeneous groups since the groups are formed by chance. Thus, it may result in unequal participation such as learners in the same group working at a different pace, or "segregated" groups in which all members exhibit some desirable or undesirable characteristics [Huxham and Land 2000]. In other terms, random grouping may lead to groups that are very different from each other promoting a certain degree of unfairness. The self-selected grouping method tends to form homogeneous groups, in which group members share common interests, get along well with each other, and are motivated to work together. However, learners working in such groups are often less task-oriented and engage in various off-task behaviors [Hilton and Phillips 2010]. The controlled grouping method is actively employed nowadays to promote the formation of heterogeneous groups, which as mentioned above seem to work best. In addition, the controlled grouping method aims at achieving a certain similarity among the groups and thus limit the pitfall of unfairness, which may arise with the random grouping method. However, achieving all these goals makes the assignment task complicated both in terms of formalization and in terms of finding optimal solutions. Thus, research has been focusing also on the algorithmic approach to achieve controlled grouping.

In the past decade, a variety of algorithm-based grouping methods has been proposed to form controlled groups, that is, to create groups that are as similar among themselves as possible (*inter-homogeneous*), while maximizing the learners' individual differences within such groups (*intra-heterogeneous*) [Moreno et al. 2012a]. The majority of group formation algorithms are based on population-based metaheuristics such as Ant Colony Optimization [Graf and Bekele 2006], Particle Swarm Optimization [Lin et al. 2010], and Genetic Algorithm [Chen and Kuo 2019]. Additionally, local search-based heuristics such as Variable Neighborhood Search (VNS) are also employed to form collaborative groups [Takači et al. 2017]. However, all these approaches calculate a solution to the formalized problem in a heuristic way, that is, they do not guarantee to find the optimal solution. Thus, they add a confounding factor in the analysis of how effective a formulation is with respect to student performance, namely, the degree of approximation of the optimal solutions. Moreover, the existing algorithms are majorly based on one single learner characteristic to form groups such as ability level [Zheng et al. 2018]. A few studies attempted to form groups with consideration of a few more characteristics such as learning styles [Moreno et al. 2012b], personality traits [Wang et al. 2007], and social interaction [Chen and Kuo 2019]. However, [Lei et al. 2010] claimed that there are at least six major factors that significantly impact the collaborative learning process including gender, ethnicity, familiarity among members, ability, motivation, and source. To our best knowledge, existing studies fail to take into account the motivation factor in the group formation algorithm, which is also confirmed in Borges and his colleagues' review of group formation in CSCL [Borges et al. 2017].

In this paper, we propose a novel formalization of the problem of finding an inter-homogeneous and intra-heterogeneous grouping that differs from all the previous approaches. The new formalization is based on differences among the characteristics that are maximized within groups and minimized among the groups. It can work with any number of student characteristics and also with both numerical type (e.g., a real number from [0, 1]) and categorical type (e.g., the nationality). Moreover, the approach can be easily extended to contain constraints of the type "two persons cannot be in the same group" and similar. Instead of finding the whole set of Pareto optimal solutions, we then solve a lexicographic optimization problem where the characteristics are considered in the order of importance expressed by the user. Note that with a Pareto optimization approach the user is left with the task of expressing this order after the solution process has terminated. Our approach moves, instead, this task before the solution process. We show that our new formulation is practicable and solvable to proved optimality by mixed-integer linear programming (MILP).

Therefore, the present study has two-fold purposes. Firstly, we present a novel algorithm-based approach to form controlled groups based on MILP that can handle multiple characteristics of the students. Secondly, we examine the effectiveness of the algorithm-based grouping method in a pilot study with consideration of students' three general characteristics including knowledge level, gender, as well as key motivation constructs (i.e., expectancy-value and achievement goals) by comparing it with the self-selection grouping method in terms of students' achievements and emotions experienced during the group work.

2 PROBLEM FORMULATION

We want to group a set S indexed by s of n students. Each student is characterized by a set of characteristics (or factors or features) $F = \{1..m\}$ indexed by f. Some of these characteristics, $F^q \subseteq F$, are quantitative or numerical, that is, they take values in $\mathbb{R}$; others, $F^c \subseteq F$, are categorical and can be mapped to take values in $\mathbb{N}_0$. For example, the gender of a person can be mapped into the integer numbers 0 and 1. A categorical characteristic $f \in F^c$ takes values from a finite set of categories (or levels) $L_f = \{1..v_f\} \subset \mathbb{N}_0$ indexed by ℓ. Thus, a student $s \in S$ is characterized by a vector $\vec{c}(s) = [c_{s1}, \ldots, c_{sm}]$ with $c_{sf} \in \mathbb{R}$ for $f \in F^q$ and $c_{sf} \in \mathbb{N}_0$ for $f \in F^c$. Further, let $\pi : F \to F$ be a permutation of the characteristics such that $(\pi(1)..\pi(m))$ is the one-line representation of the permutation, indicating a strict order of decreasing importance given to the characteristics.

We aim at grouping the students in S into a set of groups $\mathcal{G} \subset 2^S$. We will denote such a grouping as a mapping $\sigma : S \to \mathcal{G}$. Thus, $\sigma(s) = G$, if the student $s \in S$ is assigned to group $G \in \mathcal{G}$. We want the grouping to be a partition of $\mathcal{G}$, that is, $G_1 \cap G_2 = \emptyset$ for any $G_1, G_2 \in \mathcal{G}$ and $\bigcup_{G \in \mathcal{G}} G = S$, and such that the size of each group G in $\mathcal{G}$ under σ is $\{\lfloor |S|/|G| \rfloor, \lceil |S|/|G| \rceil\}$, that is, as equal as possible. Among all groupings satisfying these requirements, Σ, we want to find the ones that maximize intra-group heterogeneity and inter-group homogeneity concerning the characteristics under the order π.

We formulate the preference criterion above in the following way. For a grouping σ, let $\delta_{f,p,G}$ be the absolute difference in the values of the characteristic f for any pair of students $p = (s, r)$ in G, that is, $\delta_{f,p,G} = |c_{sf} - c_{rf}|$ for all $f \in F^q$, $G \in \mathcal{G}$ and $\{p = (s, r) \mid \sigma(s) = \sigma(r) = G\}$. Then, let $\underline{\theta}_f$ and $\overline{\theta}_f$ for $f \in F^q$ be the smallest and the largest of these differences among the groups, that is, for $f \in F^q$

$$\underline{\theta}_f = \min_{G \in \mathcal{G}, p \in G} \delta_{f,p,G}$$

$$\overline{\theta}_f = \max_{G \in \mathcal{G}, p \in G} \delta_{f,p,G}$$

Similarly, for a grouping σ, let $\mu_{f,G}$ be the number of categories of the characteristic $f \in F^c$ and let $\underline{\eta}_f^c$ and $\overline{\eta}_f^c$ for $f \in F^c$ be, respectively, the smallest and largest number of categories present in any $G \in \mathcal{G}$, that is, for $f \in F^c$

$$\underline{\eta}_f = \min_{G \in \mathcal{G}} \mu_{f,G}$$
$$\overline{\eta}_f = \max_{G \in \mathcal{G}} \mu_{f,G}$$

We use $\underline{\eta}_f$ and $\underline{\theta}_f$ as measures of intra-group heterogeneity, and maximize them and $\overline{\theta}_f$ and $\overline{\eta}_f$ as measures of inter-group homogeneity and minimize them. In other terms, for categorical factors we maximize the smallest number of categories in the groups, thus promoting intra-heterogeneity and minimize the largest number of categories thus promoting a small range between minimum and maximum number of categories among the groups and consequently favouring inter-homogeneity. While for quantitative factors we maximize the smallest value of differences within the groups thus promoting intra-heterogeneity and minimize the smallest value of the differences within the groups thus aiming at the smallest range between these values and consequently promoting inter-homogeneity.

This *multiobjective optimization problem* can be solved by lexicographic optimization using the strict order π of importance on the characteristics. Formally,

$$\operatorname{lex}\max_{\sigma \in \Sigma} (\varphi_1(\sigma), \ldots, \varphi_{2m}(\sigma))$$

where

$$\varphi_i(\sigma) = \begin{cases} \underline{\theta}_f & \text{if } i = 2\pi(f) - 1 \text{ and } f \in F^q \\ -\overline{\theta}_f & \text{if } i = 2\pi(f) \text{ and } f \in F^q \\ \underline{\eta}_f & \text{if } i = 2\pi(f) - 1 \text{ and } f \in F^c \\ -\overline{\eta}_f & \text{if } i = 2\pi(f) \text{ and } f \in F^c \end{cases} \quad \text{for } i = 1..2m.$$

This means that we consider first the characteristic that is first in the order, that is, $f \in F$ such that $\pi(f) = 1$, and maximize the value $\underline{\theta}_f$ or $\underline{\eta}_f$ depending on whether f is a quantitative or a categorical factor, respectively. Once the optimal grouping with respect to this objective has been found, we set that objective as a constraint and maximize $-\overline{\theta}_f$ or $-\overline{\eta}_f$, which corresponds to minimize $\overline{\theta}_f$ or $\overline{\eta}_f$. Then, we consider the next characteristic in the order, i.e., $f \in F$ such that $\pi(f) = 2$, and repeat the process while keeping all previously optimized objectives as constraints. We proceed in this way until all characteristics are considered.Each optimization problem can be formulated as a mixed integer linear programming problem and solved with one of the available general-purpose solvers. Further artificial restrictions on the set Σ of feasible groupings, such as "student s cannot be in the same group as student r" can be easily added within the same formalism.

Consider the example of Figure 1. We have four students s_1, s_2, s_3, s_4 described by two categorical characteristics C_1 and C_2 and four numerical characteristics, C_3, C_4, C_5, C_6. The order of importance of the characteristics is $\pi = (1, 2, 3, 4, 5, 6)$. We want to group the students in two groups G_1, G_2. The table on the top shows the values of the characteristics for the four students with columns in the same order as the order of importance.

The assignment made by the algorithm is shown on the bottom table. It corresponds to $\sigma(s_1) = \sigma(s_2) = G_1$ and $\sigma(s_3) = \sigma(s_4) = G_2$.

	C1	C2	C3	C4	C5	C6
s_1	1	2	0.3	0.4	0.2	0.3
s_2	0	3	0.5	0.3	0.7	0.8
s_3	1	4	0.1	0.7	0.3	0.2
s_4	0	2	0.8	0.9	0.4	0.5

Group 1	C1	C2	C3	C4	C5	C6
s_1	1	2	0.30	0.40	0.20	0.30
s_2	0	3	0.50	0.30	0.70	0.80
Group 2	C1	C2	C3	C4	C5	C6
s_3	1	4	0.10	0.70	0.30	0.20
s_4	0	2	0.80	0.90	0.40	0.50
Discrepancies	C1	C2	C3	C4	C5	C6
min	2	2	0.20	0.10	0.10	0.30
max	2	2	0.70	0.20	0.50	0.50

Figure 1: A numerical example. On the top, the input data for a case with four students (on the rows) and six characteristics (on the columns), of which the first two categorical. On the bottom, the grouping produced by the algorithm.

The last two rows show the measures of discrepancy among the groups for each characteristic. The minimum values represent $\underline{\delta}_f$ and $\underline{\theta}_f$, depending on whether the characteristic is quantitative or categorical, respectively. Similarly, the maximum values represent $\overline{\delta}_f$ and $\overline{\theta}_f$.

3 EXPERIMENTAL DESIGN

We designed an experiment as a proof of concept for our algorithm-based grouping method. We also collected preliminary data to investigate the impact of the proposed method on students' academic achievement and emotions experienced during collaborative learning in comparison with the self-selection grouping method.

3.1 Participants and context

Twenty-nine students from an introductory statistics course in a large public European university voluntarily participated in the experiment. Among the participants, around 59% of them were female and the average age was 20.4 years old. The course was required for all first-year undergraduate students in the program of Business and Administration, in which students were expected to understand a series of statistical concepts and apply them to solve complex statistical problems.

The experiment was conducted in the week when topic "continuous random variable and distribution" was introduced and discussed. Since the in-class exercise of the topic required students to enact advanced statistical skills (e.g., knowledge application, statistical analysis, and evaluation) to solve real-life problems, it is suitable for employing group learning activities in class [Retnowati et al. 2017].

Table 1: Summary statistics and reliability coefficients for measures in both Survey I and II.

Survey I			Survey II		
Measure	α	Mean (SD)	Measure	α	Mean (SD)
Expectancy	0.88	4.73 (0.91)	Relief	0.87	3.41 (1.35)
Value_Stats	0.81	4.99 (0.95)	Anger	0.51	1.69 (0.83)
Value_Group	0.92	4.26 (1.30)	Anxiety	0.87	2.18 (1.00)
Interest	0.88	4.64 (1.14)	Enjoy	0.92	4.08 (1.07)
Master	0.64	5.08 (0.74)	Pride	0.64	4.33 (0.85)
Perform_app	0.90	4.21 (1.22)	Boring	0.82	2.55 (0.94)
Perform_avo	0.89	4.34 (1.12)	Hopeless	0.55	1.93 (1.00)

3.2 Considered characteristics

Based on the literature discussed above [Borges et al. 2017; Lei et al. 2010] , three general student characteristics with 13 specific features were used as criteria to build the group formation algorithm, including i. an estimate of the level of student statistical knowledge, ii. student demographic information (i.e., gender and nationality), and iii. estimates of student motivation in learning statistics (i.e., expectancy-value, and goal orientation). To obtain such student characteristics, an online survey (survey I) was conducted before the in-class group work.The survey took each student around 5-10 min to complete. The survey measures included demographic information, statistical knowledge level, expectancy-value of learning statistics and engaging in group work, and goal orientations of learning statistics. Summary statistics and reliability coefficients of the measures are presented in Table 1.

Student demographic information. Students reported a list of demographic information including gender, age, study program, nationality and native language at beginning of the survey.

Statistical knowledge level. The information was obtained by asking students rating their perceived statistical knowledge level ranging from 1 (extremely low level) to 6 (extremely high level).

Expectancy-value. [Wigfield and Eccles 2000] expectancy-value survey instrument was adapted to measure students' expectancy and value of learning statistics and engaging in group work. Items were measured on a 7-point Likert Scale ranging from 1 ("strongly disagree") to 7 ("strongly agree"). One sample item for learning statistics is "I value the statistics class". One sample item for engaging in group work is "I think group work is useful in the statistics class".

Goal orientations. [Elliot and Murayama 2008] 2×2 goal orientation survey instrument was adapted to measure students' adoption of goal orientations in the statistics course. Items were measured on a 7-point Likert Scale ranging from 1 ("strongly disagree") to 7 ("strongly agree"). One sample item is "My aim is to completely master the material presented in the statistics class."

The characteristics of each individual student are reported in Table 2. Except for the first column providing an anonymous identifier, the other columns are used by our algorithm in order of decreasing importance. For each column we also emphasize whether the characteristic has been treated as a numerical or categorical criterion.

3.3 Output and performance of our group formation algorithm

We solved the MILP formulation of the problem formalized in Section 2 on the data from Table 2. As solver we used the commercial solver Gurobi that can handle lexicographic optimization automatically [Gurobi Optimization 2021]. The optimal solution was determined in less than five seconds on a normal contemporary laptop. The results are reported in Table 3. The order of the columns of Table 3 indicates the order of importance π used in the optimization process.

Analyzing the results was not a trivial task. The characteristic "statslevel" had several students with value 3.00 hence a smallest difference within the groups of 0.00 might be unavoidable. If the largest difference was high then we might risk having some group very heterogeneous with respect to this characteristic while another group, like Group 4, having all students with the same value for the characteristic. Similar observations can be made with the other characteristics later in the order, although for them it might be more complicated because what can be done in their regard depends on what is done on the characteristics earlier in the order.

Note that if a user is unsatisfied with the outcome, the user can easily try alternative orders thus interacting with the solution process. Computation time does not seem to be an issue although we did not test how the method scales with growing size of the problem.

3.4 Impact on learning

We attempted to compare the algorithm-based and self-selection grouping methods in terms of two types of outcome measures, which were academic achievement and emotions experienced during the group work.

Academic achievement. Exam scores of the questions related to the "continuous random variable and distribution" topic were obtained from the final exam serving as an indicator of students' cognitive learning outcome.

Emotions. The Achievement Emotions Questionnaire [Peixoto et al. 2015] (survey II) was adapted and administrated immediately after the in-class group work to measure students' emotions during the group work. Specifically, seven types of emotions were measured including relief, anger, anxiety, enjoyment, pride, boring, and hopelessness. Items were measured on a 7-point Likert Scale ranging from 1 ("strongly disagree") to 7 ("strongly agree"). The sample item is "I enjoy doing the group work".

3.5 Experiment procedure

The first step was to form groups based on the proposed algorithm with students' responses to survey I as criteria. Twenty-one students responded to survey I and were assigned to six groups with three to four students per group before the lecture. In the day of conducting the experiment, thirteen of the twenty-one students showed up. Since the grouping information was provided to students on the experimental day, thus we considered the missing students were random, which may affect but not confound with the results. Based on the algorithm method, the thirteen students resulted in five groups, among which two groups had only two students and the rest

Table 2: The characteristics of the students. Columns are not sorted by type of caracteristic, not by order of importance in the optimization process.

ID	gender (categorical)	age (categorical)	nationality (categorical)	English (categorical)	program (categorical)	statslevel (numerical)	expectancy (numerical)	value (numerical)	master_goal (numerical)	perform_app (numerical)	perform_avo (numerical)	interest (numerical)	value_groupwork (numerical)
1	Female	20	Italian	No	EU Studies	2	4.67	6.00	6.00	4.33	4.33	6.00	6.00
2	Female	19	Moldovan	No	MBA	3	5.67	6.00	6.00	4.67	6.00	6.00	4.33
3	Female	20	Slovak	No	MBA	4	5.67	6.00	5.00	4.00	4.00	5.33	1.67
4	Female	25	Danish	No	MBA	3	3.67	4.00	4.33	3.33	3.67	2.67	4.67
5	Male	24	Romanian	No	MBA	4	5.00	5.00	4.67	5.33	3.67	5.00	3.00
6	Male	19	Danish	No	MBA	5	6.00	5.67	5.67	6.00	6.00	6.00	3.67
7	Female	24	UK	Yes	MBA	1	5.00	5.00	5.33	5.00	3.67	5.00	5.67
8	Male	22	German	No	EU Studies	2	4.33	3.67	5.33	3.33	2.67	4.00	3.67
9	Female	22	Danish	No	EU Studies	1	5.00	4.67	4.67	2.00	3.33	4.00	6.00
10	Female	22	Danish	No	EU Studies	3	4.33	5.00	5.67	4.00	3.33	5.00	2.67
11	Female	20	Slovenian	No	MBA	3	5.00	4.00	4.00	4.33	5.33	3.33	4.00
12	Female	23	Danish	No	MBA	3	5.00	6.00	5.33	4.67	5.00	5.33	4.33
13	Female	20	Danish	No	EU Studies	1	2.67	3.33	5.33	2.33	2.33	2.33	6.00
14	Male	21	Danish	No	EU Studies	2	3.33	5.67	5.00	3.67	4.33	5.00	5.00
15	Female	33	Ukrainian	No	MBA	3	5.00	5.00	5.33	5.00	5.33	5.00	2.33
16	Male	22	Germany	No	MBA	3	6.00	5.67	5.67	5.67	5.67	5.33	5.67
17	Male	20	Hungarian	No	MBA	3	5.00	5.00	5.33	2.67	3.00	5.00	4.00
18	Female	19	American	Yes	EU Studies	3	5.00	5.67	6.00	4.67	5.00	5.67	4.67
19	Female	21	germany	No	MBA	3	3.00	3.00	3.00	3.33	3.00	3.00	3.67
20	Female	20	German	No	EU Studies	1	3.33	3.67	4.33	2.67	4.33	3.67	5.00
21	Female	18	China	No	MBA	3	5.00	5.33	5.00	4.67	5.00	4.67	3.33

Table 3: The groups created by our group formation algorithm. Rows represent students that have been re-indexed by group, except the last two rows that represent the measures of discrepancies as discussed in the example of Table 1. Columns are sorted by order of decreasing importance in the optimization process. Categorical variables have been encoded in natural numbers.

Group 1	statslevel	gender	value_groupwork	interest	program	nationality	age	expectancy	value	perform_app	perform_avo	master_goal	English
0	1.00	0	5.67	5.00	0	12	6	5.00	5.00	5.00	3.67	5.33	1
1	1.00	0	6.00	4.00	1	3	4	5.00	4.67	2.00	3.33	4.67	0
2	1.00	0	5.00	3.67	1	4	2	3.33	3.67	2.67	4.33	4.33	0
Group 2	statslevel	gender	value_groupwork	interest	program	nationality	age	expectancy	value	perform_app	perform_avo	master_goal	English
0	3.00	0	4.67	2.67	0	2	7	3.67	4.00	3.33	3.67	4.33	0
1	3.00	0	2.33	5.00	0	13	8	5.00	5.00	5.00	5.33	5.33	0
2	3.00	0	4.33	5.33	0	3	5	5.00	6.00	4.67	5.00	5.33	0
3	3.00	0	3.67	3.00	0	14	3	3.00	3.00	3.33	3.00	3.00	0
Group 3	statslevel	gender	value_groupwork	interest	program	nationality	age	expectancy	value	perform_app	perform_avo	master_goal	English
0	2.00	1	5.00	5.00	1	3	3	3.33	5.67	3.67	4.33	5.00	0
1	2.00	1	3.67	4.00	1	4	4	4.33	3.67	3.33	2.67	5.33	0
2	1.00	0	6.00	2.33	1	3	2	2.67	3.33	2.33	2.33	5.33	0
Group 4	statslevel	gender	value_groupwork	interest	program	nationality	age	expectancy	value	perform_app	perform_avo	master_goal	English
0	3.00	0	2.67	5.00	1	3	4	4.33	5.00	4.00	3.33	5.67	0
1	3.00	0	4.33	6.00	0	8	1	5.67	6.00	4.67	6.00	6.00	0
2	3.00	0	3.33	4.67	0	1	0	5.00	5.33	4.67	5.00	5.00	0
3	3.00	0	4.00	3.33	0	11	2	5.00	4.00	4.33	5.33	4.00	0
Group 5	statslevel	gender	value_groupwork	interest	program	nationality	age	expectancy	value	perform_app	perform_avo	master_goal	English
0	3.00	0	4.67	5.67	1	0	1	5.00	5.67	4.67	5.00	6.00	1
1	2.00	0	6.00	6.00	1	7	2	4.67	6.00	4.33	4.33	6.00	0
2	3.00	1	4.00	5.00	0	6	2	5.00	5.00	2.67	3.00	5.33	0
3	3.00	1	5.67	5.33	0	5	4	6.00	5.67	5.67	5.67	5.67	0
Group 6	statslevel	gender	value_groupwork	interest	program	nationality	age	expectancy	value	perform_app	perform_avo	master_goal	English
0	4.00	1	3.00	5.00	0	9	6	5.00	5.00	5.33	3.67	4.67	0
1	5.00	1	3.67	6.00	0	3	1	6.00	5.67	6.00	6.00	5.67	0
2	4.00	0	1.67	5.33	0	10	2	5.67	6.00	4.00	4.00	5.00	0
Discrepancies	statslevel	gender	value_groupwork	interest	program	nationality	age	expectancy	value	perform_app	perform_avo	master_goal	English
min	0.00	1	0.33	0.33	1	2	3	0.00	0.00	0.00	0.33	0.00	1
max	1.00	2	2.33	2.67	2	4	4	2.00	3.00	3.00	2.67	2.33	2

three groups had three students. The rest of the class that consisted of sixteen students was self-grouped into six groups, among which one group had two students, and the remaining groups all had three students. In sum, thirteen students were assigned by the *algorithm-based grouping method* and sixteen students were grouped by the *self-selected grouping method*.

Following the group formation, the experiment entered the second stage. Students were invited to work in groups to solve two statistical problems that were associated with the "continuous random variable and distribution" topic. While students conducted collaborative work, the instructor walked around to provide guidance if needed.

After the completion of the group activities, we collected information concerning students' emotions during collaborative learning. A paper-pencil survey (survey II) was conducted immediately in class after the group work.

4 RESULTS

The results compared the two types of group formation methods, self-selected and algorithm-based grouping method, in terms of students' academic achievement on the chosen topic and the perceived emotions during the group work. As for the academic achievement, the results showed that students in the algorithm-based group (median: 13, average: 9.71, standard deviation 5.79) scored slightly higher than those in the self-selection group (median: 11, average: 9, standard deviation: 5.35) (see Figure 2). However, the differences are not statistically significant according to an independent t-test. With respect to students' emotion during the group work, the results showed that students in the algorithm-based group demonstrated slightly higher level of almost all types of emotions except boredom than those in the self-selection group. As shown in Figure 2, in general, students in both types of groups felt more positive (e.g. relief, pride, and enjoyment) while working in groups. When comparing with the self-selection group, we found that students in the algorithm-based group seemed to feel more enjoyable, pride, and relief working in the group, meanwhile also feel more anxious, hopeless, and anger during the group work. However, a multivariate analysis of variance (MANOVA) conducted to investigate the differences on the seven emotions measures indicated these differences was not statistically significant.

5 CONCLUSION AND DISCUSSION

Classification of learners into optimal learning groups is one of the most challenging tasks in the field of collaborative learning. Intra-heterogeneous and inter-homogeneous groups seem to be functioning best but their creation is a complicated task. The algorithm-based approach presented here provides a practicable solution for this task. In our study, we formalized the group formation problem in a novel way and used mixed integer linear programming (MILP) solvers to find provable optimal solutions to our formulation. Different from other algorithmic approaches, the method we proposed allows for the consideration of multiple student characteristics, handles both numerical and categorical characteristic types simultaneously, and enables an order of importance for the characteristics. We have shown that the method is computationally viable.[1]

We conducted a pilot study to illustrate how to use this new way of forming groups and its potential in improving collaborative learning. To the best of our knowledge, we have been the first to include motivation factors as criteria in the group formation algorithm. We made an attempt at comparing our approach with self-selection grouping but the small scale of our study does not allow us to draw reliable conclusions. Further research is needed to understand which student features are relevant to be used with our algorithm-based grouping method and whether it generates well functioning groups, and thus enhance students' gain from collaborative learning activities both academically and emotionally.

REFERENCES

Simone Borges, Riichiro Mizoguchi, Ig Ibert Bittencourt, and Seiji Isotani. 2017. Group Formation in CSCL: A Review of the State of the Art. In *Researcher Links Workshop: Higher Education for All*. Springer, 71–88.

Teyi Chan, Chien-Ming Chen, Yu-Lung Wu, Bin-Shyan Jong, Yen-Teh Hsia, and Tsong-Wuu Lin. 2010. Applying the genetic encoded conceptual graph to grouping learning. *Expert Systems with Applications* 37, 6 (2010), 4103–4118.

Chih-Ming Chen and Chi-Hsiung Kuo. 2019. An optimized group formation scheme to promote collaborative problem-based learning. *Computers & Education* 133 (2019), 94–115.

Andrew J Elliot and Kou Murayama. 2008. On the measurement of achievement goals: Critique, illustration, and application. *Journal of educational psychology* 100, 3 (2008), 613.

Francesco Fallucchi, Enrique Fatas, Felix Kölle, and Ori Weisel. 2018. Not all group members are created equal: Heterogeneous abilities in inter-group contests. *Luxembourg Institute of Socio-Economic Research (LISER) Working Paper Series* 14 (2018), 1–42.

Sabine Graf and Rahel Bekele. 2006. Forming heterogeneous groups for intelligent collaborative learning systems with ant colony optimization. In *International conference on intelligent tutoring systems*. Springer, 217–226.

LLC Gurobi Optimization. 2021. Gurobi Optimizer Reference Manual. http://www.gurobi.com

Allyson Hadwin, Sanna Järvelä, and Mariel Miller. 2018. Self-regulation, co-regulation, and shared regulation in collaborative learning environments. In *Educational psychology handbook series. Handbook of self-regulation of learning and performance*, Dale H. Schunk and Jeffrey A. Greene (Eds.). Routledge/Taylor & Francis Group.

Sandy Hilton and Fred Phillips. 2010. Instructor-assigned and student-selected groups: A view from inside. *Issues in Accounting Education* 25, 1 (2010), 15–33.

Mark Huxham and Ray Land. 2000. Assigning students in group work projects. Can we do better than random? *Innovations in Education and Training International* 37, 1 (2000), 17–22.

Simon A Lei, Bailey N Kuestermeyer, and Kara A Westmeyer. 2010. Group composition affecting student interaction and achievement: instructors' perspectives. *Journal of instructional psychology* 37, 4 (2010), 317.

Yen-Ting Lin, Yueh-Min Huang, and Shu-Chen Cheng. 2010. An automatic group composition system for composing collaborative learning groups using enhanced particle swarm optimization. *Computers & Education* 55, 4 (2010), 1483–1493.

Chen-Chung Liu and Chin-Chung Tsai. 2008. An analysis of peer interaction patterns as discoursed by on-line small group problem-solving activity. *Computers & Education* 50, 3 (2008), 627–639.

Yiping Lou, Philip C Abrami, John C Spence, Catherine Poulsen, Bette Chambers, and Sylvia d'Apollonia. 1996. Within-class grouping: A meta-analysis. *Review of educational research* 66, 4 (1996), 423–458.

Sven Manske, Tobias Hecking, Irene-Angelica Chounta, Sören Werneburg, and H Ulrich Hoppe. 2015. Using differences to make a difference: a study on heterogeneity of learning groups. In *Exploring the Material Conditions of Learning: The Computer Supported Collaborative Learning (CSCL) Conference 2015*, O. Lindwall, P. Haäkkinen, T. Koschman, P. Tchounikine, and S. Ludvigsen (Eds.), Vol. 1. International Society of the Learning Sciences, Inc.[ISLS]., Gothenburg, Sweden, 8.

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

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

[1]We maintain the scripts implementing the method publicly available at https://github.com/belzebuu/GroupFormation.

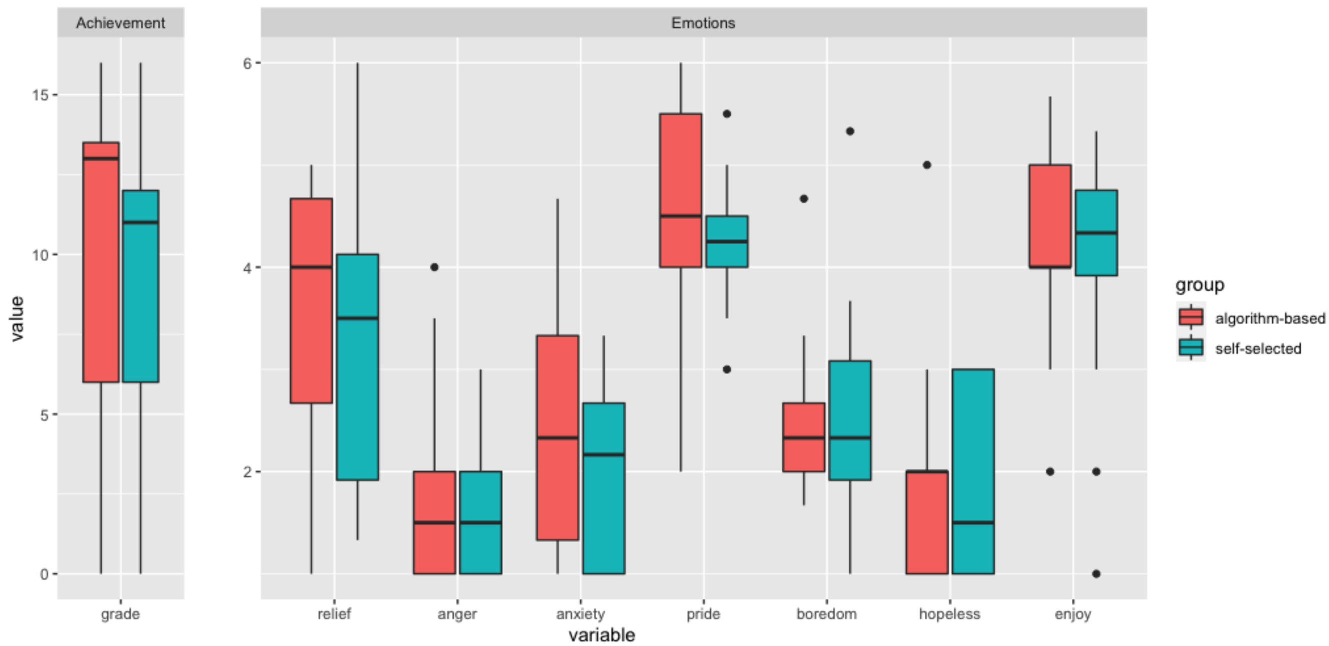

Figure 2: Comparison of algorithm-based grouping and self-selected grouping with respect to the outcomes academic achievement and emotions.

P Karen Murphy, Jeffrey A Greene, Carla M Firetto, Mengyi Li, Nikki G Lobczowski, Rebekah F Duke, Liwei Wei, and Rachel MV Croninger. 2017. Exploring the influence of homogeneous versus heterogeneous grouping on students' text-based discussions and comprehension. *Contemporary Educational Psychology* 51 (2017), 336–355.

Francisco Peixoto, Lourdes Mata, Vera Monteiro, Cristina Sanches, and Reinhard Pekrun. 2015. The achievement emotions questionnaire: Validation for pre-adolescent students. *European Journal of Developmental Psychology* 12, 4 (2015), 472–481.

Anika Radkowitsch, Freydis Vogel, and Frank Fischer. 2020. Good for learning, bad for motivation? A meta-analysis on the effects of computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning* 15, 1 (March 2020), 5–47.

Endah Retnowati, Paul Ayres, and John Sweller. 2017. Can collaborative learning improve the effectiveness of worked examples in learning mathematics? *Journal of educational psychology* 109, 5 (2017), 666.

Đurđica Takači, Miroslav Marić, Gordana Stankov, and Aleksandar Djenić. 2017. Efficiency of using VNS algorithm for forming heterogeneous groups for CSCL learning. *Computers & Education* 109 (2017), 98–108.

Dai-Yi Wang, Sunny SJ Lin, and Chuen-Tsai Sun. 2007. DIANA: A computer-supported heterogeneous grouping system for teachers to conduct successful small learning groups. *Computers in Human Behavior* 23, 4 (2007), 1997–2010.

Allan Wigfield and Jacquelynne S Eccles. 2000. Expectancy–value theory of achievement motivation. *Contemporary educational psychology* 25, 1 (2000), 68–81.

Yaqian Zheng, Chunrong Li, Shiyu Liu, and Weigang Lu. 2018. An improved genetic approach for composing optimal collaborative learning groups. *Knowledge-Based Systems* 139 (2018), 214–225.