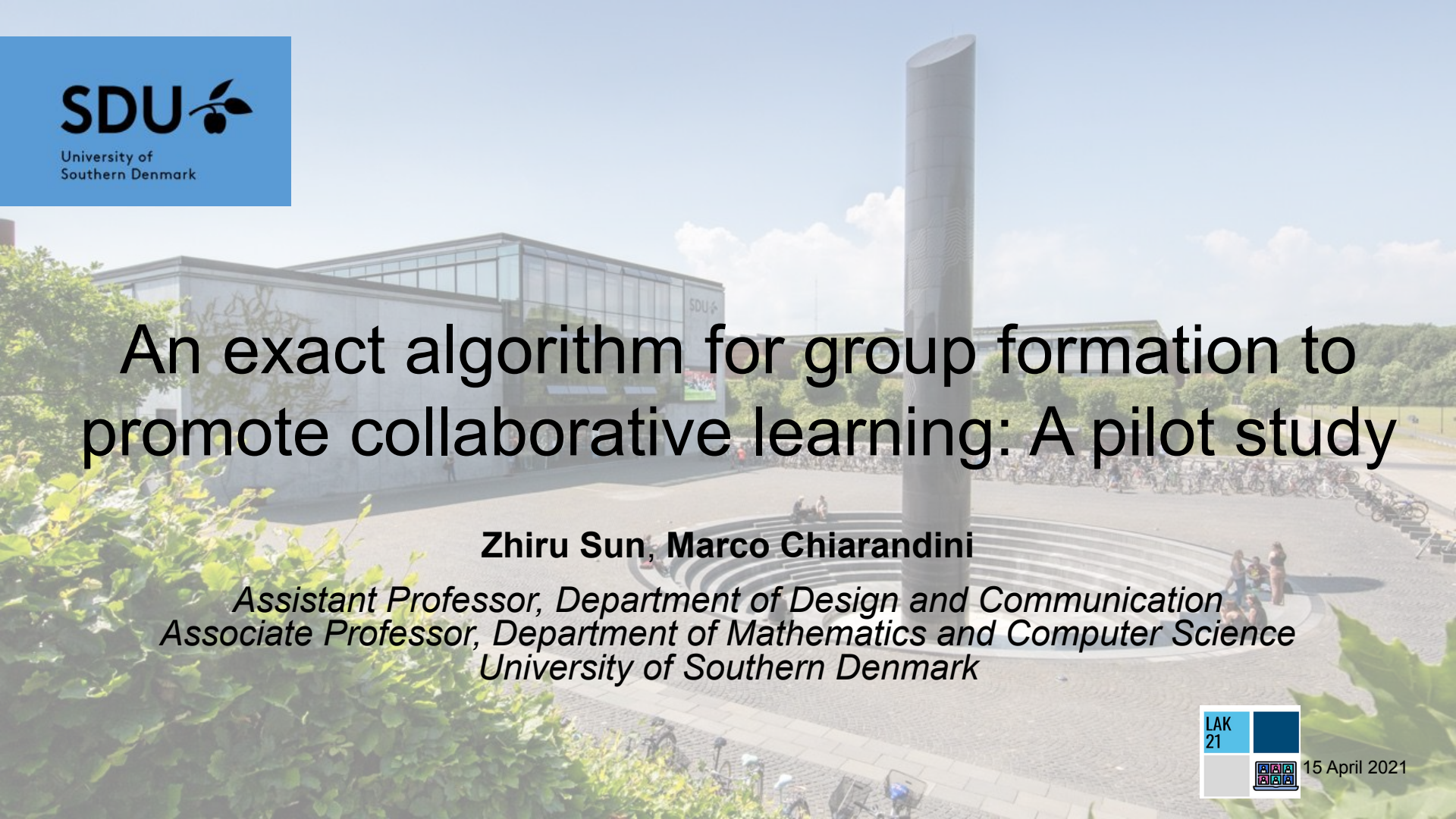


An exact algorithm for group formation to promote collaborative learning: A pilot study

Zhiru Sun, Marco Chiarandini

Assistant Professor, Department of Design and Communication
Associate Professor, Department of Mathematics and Computer Science
University of Southern Denmark

15 April 2021

Background

- Collaborative learning is considered an effective teaching method.
  - Supported by zone of proximal development (ZPD) theory (Vygotsky, 1978)
  - The teacher asks two or more people to collaborate because it is expected that they will learn efficiently through
    - Engaging learners through communication with other group members
    - Exchanging of diverse ideas for joint knowledge construction (Liu & Tsai, 2008)
- Collaboration can benefit learning from both theoretical and empirical perspectives
  - Positive effects on engagement (Northey et al., 2017), achievement and attitude (Kyndt et al., 2013).

Motive

- Collaborative situation is a kind of "*social contract*" that specifies conditions under which certain types of desired interactions *may* occur and further lead to positive learning outcomes (Dillenbourg, 1999).
- However, simply putting students together for collaborative work will not guarantee desired interactions and maximized learning outcomes.
- The success of collaborative learning requires the satisfaction of certain conditions.

"Social Contract"

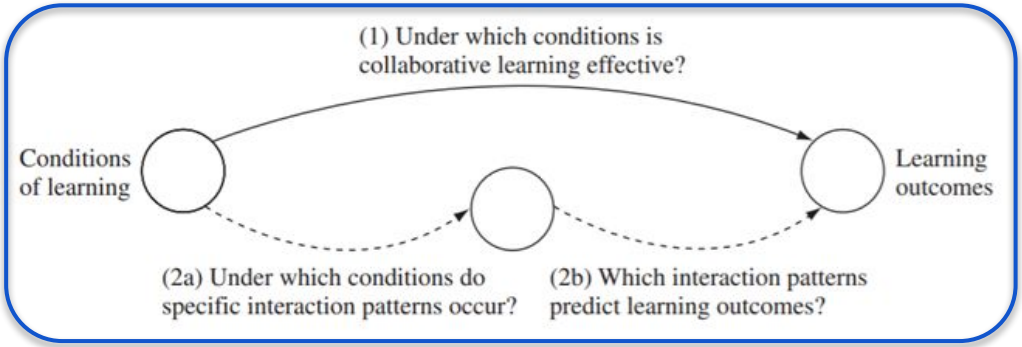
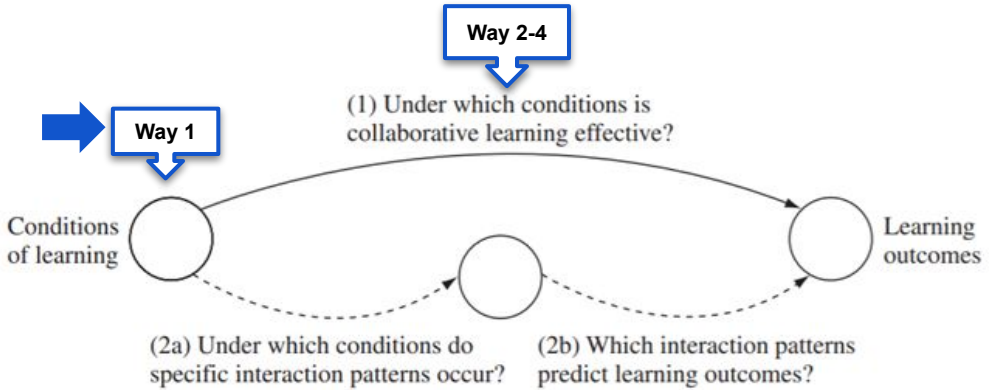


Image by Dillenbourg et al., 2009

Ways to meet the conditions

1. To set up initial conditions
2. To over-specify the “collaboration” contact with a scenario based on roles
3. To scaffold productive interactions by encompassing interaction rules in the medium
4. To monitor and regulate the interactions (Dillenbourg, 1999)



Set-up of initial conditions

1. What is the optimal group size?

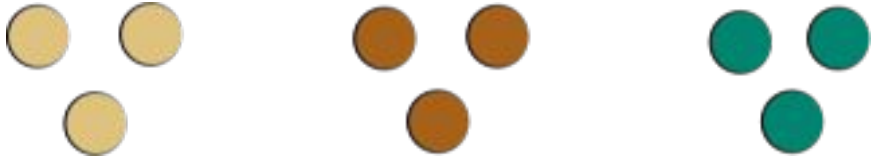
2. How to form the group?

Intra-Group composition: HOM vs. HET

Group type
Intra-Homogeneous groups (*i.e., group members with similar characteristics*)
Intra-Heterogeneous groups (*i.e., group members with diverse characteristics*)

Murphy et al., 2017

Methods of group formation

Methods	Possible Groups
Random grouping (*i.e., assigning students to groups by chance*)	
Self-selected grouping (*i.e., students self-selecting their group members*)	
Controlled grouping (*i.e., assigning students to groups by teachers or computers based on certain criteria*)	

Controlled grouping method

Elements	Teacher-based	Algorithm-based
How to form groups?	Teachers manually assign groups.	Algorithms automatically assign groups.
Class size	Small class size	Both small and large class size
Criteria	Limited criteria	Various criteria
Group composition	Intra-HET and Inter-HOM	Intra-HET and Inter-HOM
Optimal solution	Possible, but not guaranteed Lack of the computational resources	Mathematically possible

Review of existing algorithm-based grouping methods

1. Algorithm approach:
   a. Population-based metaheuristics
      i. Ant Colony Optimization (Graf & Bekele, 2006)
      ii. Particle Swarm Optimization (Line et al., 2010)
      iii. Genetic Algorithm (Chen & Kuo, 2019)
   b. Local search-based heuristics
      i. Variable Neighborhood Search (VNS) (Takaci et al., 2017)

2. Possible limitations:
   a. No guarantee to find the optimal solution due to the heuristic approach
   b. Characteristics:
      i. Limited number of criteria considered in the algorithm
         1. Ability level (Zheng et al., 2018), Learning styles (Moreno et al., 2012), Personality traits (Wang et al., 2007), Social interaction (Chen & Kuo, 2019)
      ii. Lack of characteristics such as gender, ethnicity, motivation etc. (Borges et al., 2017)
   c. Disregard priorities on the characteristics

Research purpose

- We propose a novel method based on Mixed-Integer Linear Programming (MILP) to form intra-heterogeneous and inter-homogenous groups based on three major student characteristics - *student demographic information, domain knowledge level, and motivation*.
- We use a design-based approach to examine the effectiveness of the algorithm-based grouping method by comparing it with the self-selected grouping method in terms of students achievement and affect during the collaborative work.

Problem Formulation (1)

Given:

- a set of students
- a set of characteristics:
  - some are categorical and can be mapped to the natural numbers
  - some are quantitative or numerical and take real numbers
- a value for each student in each characteristic
- a priority order on the characteristics (decreasing order of importance)
- the number of groups to create

	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
s_5	1	4	0.3	0.2	0.3	0.1
s_6	0	2	0.4	0.8	0.5	0.3

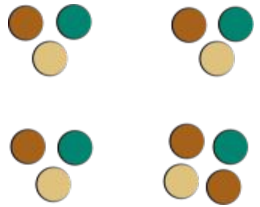
Problem Formulation (2)

Task:

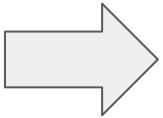
Find a partition of students into groups of size as equal as possible that maximizes:

- intra-group heterogeneity
- inter-group homogeneity

with respect to the characteristics under the order



	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
s_5	1	4	0.3	0.2	0.3	0.1
s_6	0	2	0.4	0.8	0.5	0.3



Group 1	C1	C2	C3	C4	C5	C6
s_1	1	2	0.3	0.4	0.2	0.3
s_3	1	4	0.1	0.7	0.3	0.2
s_5	1	4	0.3	0.2	0.3	0.1
Group 2	C1	C2	C3	C4	C5	C6
s_2	0	3	0.5	0.3	0.7	0.8
s_4	0	2	0.8	0.9	0.4	0.5
s_6	0	2	0.4	0.8	0.5	0.3

Measures for Homogeneity/Heterogeneity (1)

For a categorical characteristic

1. Determine the number of different categories in each group
2. Determine the smallest and largest of these values over all groups

- max intra-heterogeneity = max smallest value
- max inter-homogeneity = min the largest value

Group 1	C1	C2	C3	C4	C5	C6
s_1	1	2	0.3	0.4	0.2	0.3
s_3	1	4	0.1	0.7	0.3	0.2
s_5	1	4	0.3	0.2	0.3	0.1
#	1	2				
Group 2	C1	C2	C3	C4	C5	C6
s_2	0	3	0.5	0.3	0.7	0.8
s_4	0	2	0.8	0.9	0.4	0.5
s_6	0	2	0.4	0.8	0.5	0.3
#	1	2				
min	1	2				
max	1	2				

Measures for Homogeneity/Heterogeneity (2)

For a quantitative characteristic

1. Determine the absolute differences between all pairs of students in each groups
2. Determine the smallest and largest of these values over all groups

- max intra-heterogeneity = max smallest value
- max inter-homogeneity = min largest value

Group 1	C1	C2	C3	C4	C5	C6
s_1	1	2	0.3	0.4	0.2	0.3
s_3	1	4	0.1	0.7	0.3	0.2
s_5	1	4	0.3	0.2	0.3	0.1
$\lvert s_1 - s_3 \rvert$			0.2	0.3	0.1	0.1
$\lvert s_3 - s_5 \rvert$			0.2	0.5	0.0	0.1
$\lvert s_1 - s_5 \rvert$			0.0	0.2	0.1	0.2
Group 2	C1	C2	C3	C4	C5	C6
s_2	0	3	0.5	0.3	0.7	0.8
s_4	0	2	0.8	0.9	0.4	0.5
s_6	0	2	0.4	0.8	0.5	0.3
$\lvert s_2 - s_4 \rvert$			0.3	0.6	0.3	0.3
$\lvert s_4 - s_6 \rvert$			0.4	0.1	0.1	0.2
$\lvert s_2 - s_6 \rvert$			0.1	0.5	0.2	0.5
min			0.0	0.1	0.0	0.1
max			0.4	0.6	0.3	0.5

The Overall Algorithm

Let C be the set of constraints (initially, disjoint and equal size groups)

For each characteristic x in decreasing order of priority:

- maximize the intra-heterogeneity wrt to x subject to C ⇨ MILP
- add to C a constraint on intra-heterogeneity wrt x
- maximize the inter-homogeneity wrt x subject to C ⇨ MILP
- add to C a constraint on intra-heterogeneity wrt x

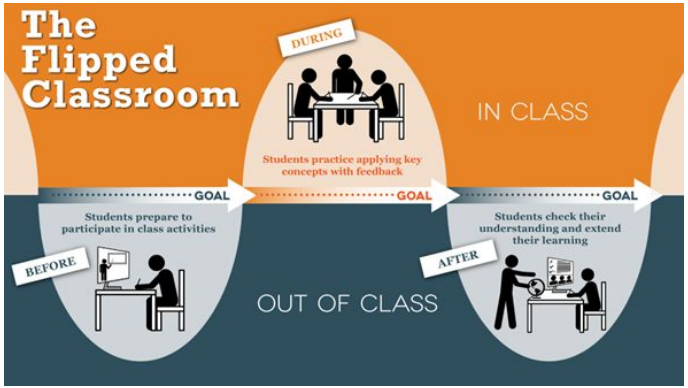
The Instance of the Feasibility Study

ID	statslevel (numerical)	gender (categorical)	value_groupwork (numerical)	interest (numerical)	program (categorical)	nationality (categorical)	age (categorical)	expectancy (numerical)	value (numerical)	perform_app (numerical)	perform_avo (numerical)	master_goal (numerical)	English (categorical)
1	2	Female	6,00	6,00	EU Studies	Italian	20	4,67	6,00	4,33	4,33	6,00	No
2	3	Female	4,33	6,00	BA	Moldovan	19	5,67	6,00	4,67	6,00	6,00	No
3	4	Female	1,67	5,33	BA	Slovak	20	5,67	6,00	4,00	4,00	5,00	No
4	3	Female	4,67	2,67	BA	Danish	25	3,67	4,00	3,33	3,67	4,33	No
5	4	Male	3,00	5,00	BA	Romanian	24	5,00	5,00	5,33	3,67	4,67	No
6	5	Male	3,67	6,00	BA	Danish	19	6,00	5,67	6,00	6,00	5,67	No
7	1	Female	5,67	5,00	BA	UK	24	5,00	5,00	5,00	3,67	5,33	Yes
8	2	Male	3,67	4,00	EU Studies	German	22	4,33	3,67	3,33	2,67	5,33	No
9	1	Female	6,00	4,00	EU Studies	Danish	22	5,00	4,67	2,00	3,33	4,67	No
10	3	Female	2,67	5,00	EU Studies	Danish	22	4,33	5,00	4,00	3,33	5,67	No
11	3	Female	4,00	3,33	BA	Slovenian	20	5,00	4,00	4,33	5,33	4,00	No
12	3	Female	4,33	5,33	BA	Danish	23	5,00	6,00	4,67	5,00	5,33	No
13	1	Female	6,00	2,33	EU Studies	Danish	20	2,67	3,33	2,33	2,33	5,33	No
14	2	Male	5,00	5,00	EU Studies	Danish	21	3,33	5,67	3,67	4,33	5,00	No
15	3	Female	2,33	5,00	BA	Ukrainian	33	5,00	5,00	5,00	5,33	5,33	No
16	3	Male	5,67	5,33	BA	Germany	22	6,00	5,67	5,67	5,67	5,67	No
17	3	Male	4,00	5,00	BA	Hungarian	20	5,00	5,00	2,67	3,00	5,33	No
18	3	Female	4,67	5,67	EU Studies	American	19	5,00	5,67	4,67	5,00	6,00	Yes
19	3	Female	3,67	3,00	BA	germany	21	3,00	3,00	3,33	3,00	3,00	No
20	1	Female	5,00	3,67	EU Studies	German	20	3,33	3,67	2,67	4,33	4,33	No
21	3	Female	3,33	4,67	BA	China	18	5,00	5,33	4,67	5,00	5,00	No

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

Participation and context

- Twenty-nine freshmen in the program of Business and Administration and European Studies in a large public university voluntarily participated in the design-based study.
- 59% of participants were female, and the average age was 20.4 years old.
- The experiment was conducted in one of the flipped introductory statistics lectures where in-class was devoted for collaborative learning.



Instance features

- Two surveys were conducted before and after the group work (i.e., Survey I and II).
- Characteristics/criteria in the algorithm in Survey I
  - Demographic information: Gender, nationality, native language
  - Domain knowledge level: Self-reported statistical level ranging from 1(extremely low level) to 6 (extremely high level)
  - Motivation constructs:
    - Expectancy and value: Expectancy-value survey (Wigfield & Eccles, 2000)
    - Achievement goal: 2x2 goal orientation survey (Elliot & Murayama, 2008)

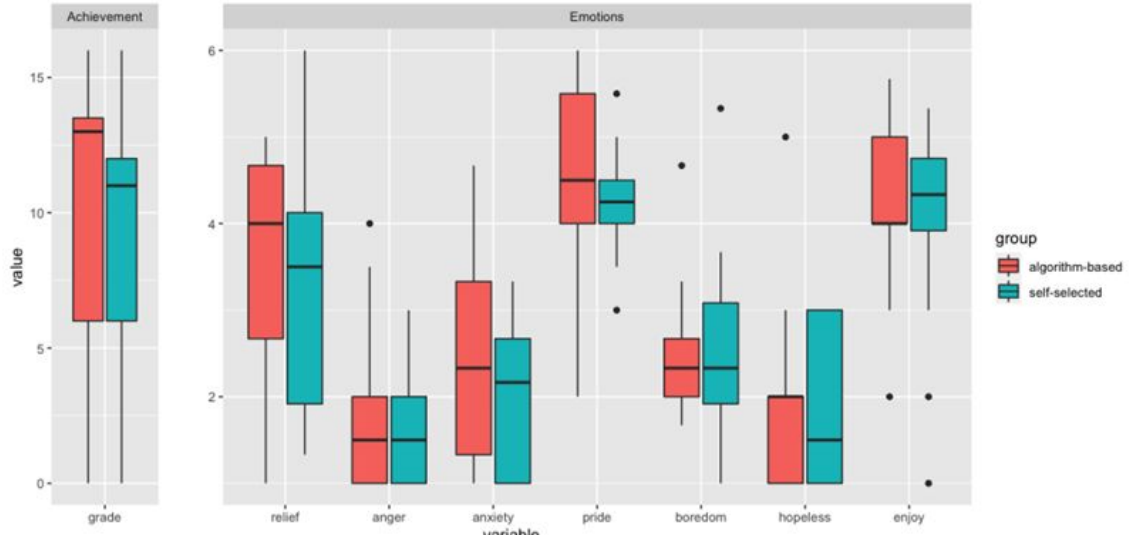
Learning outcome measures

- Academic achievement: Exam scores of the question related to the lecture
- Affect: Achievement Emotions Questionnaire (Peixoto et al., 2015) including relief, anger, anxiety, enjoyment, pride, boredom, and hopelessness

Experiment procedure

- Group formation:
  - 13 students who responded to survey I were assigned to 5 algorithm-based groups ---> ***algorithm-based groups***.
  - 16 students formed the group on their own with 2-4 students in a group ---> ***self-selected groups.***
- Collaborative work:
  - Students were required to solve two statistical questions in groups by answering scaffolded questions.
  - Teachers walked around to provide guidance if needed.

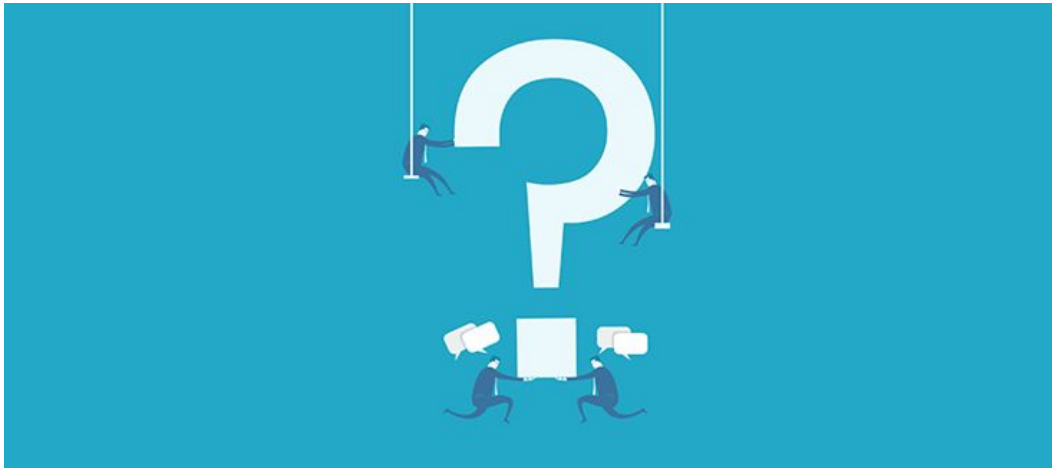
Results



- No sig. differences were found between algorithm-based and self-selected groups.
- On average, compared to students in the self-selected group, students in the algorithm-based group
  - achieved slightly higher
  - Felt a bit more pride and relieved after the group work, but also more anxious during the group work

Conclusion and Future improvement

- Conclusion:
  - The proposed algorithm provides a practical solution for the task that can
    - Formalize inter-homogeneous and intra-heterogeneous groups
    - Handle both categorical and numerical characteristic types simultaneously
    - Allows priorities on the characteristics
    - Obtain optimal solutions by using mixed integer linear programming (MILP)
  - The results of the design-based study demonstrated the computational feasibility of the algorithm-based grouping approach.
- Future work:
  - The experiment will be replicated in a large sample size to evaluate the effectiveness of the proposed grouping method.



THANK YOU & QUESTIONS?