RESEARCH   **Open Access**

Learning log-based automatic group formation: system design and classroom implementation study

Changhao Liang[1*], Rwitajit Majumdar[2] and Hiroaki Ogata[2]

*Correspondence:
bluster3a@gmail.com
[1]Graduate School of Informatics,
Kyoto University, Yoshida
Nihonmatsu-cho, Sakyo-ku,
606-8501 Kyoto, Japan
Full list of author information is
available at the end of the article

Abstract

Collaborative learning in the form of group work is becoming increasingly significant in education since interpersonal skills count in modern society. However, teachers often get overwhelmed by the logistics involved in conducting any group work. Valid support for executing and managing such activities in a timely and informed manner becomes imperative. This research introduces an intelligent system focusing on group formation which consists of a parameter setting module and the group member visualization panel where the results of the created group are shown to the user and can be graded. The system supports teachers by applying algorithms to actual learning log data thereby simplifying the group formation process and saving time for them. A pilot study in a primary school mathematics class proved to have a positive effect on students' engagement and affections while participating in group activities based on the system-generated groups, thus providing empirical evidence to the practice of Computer-Supported Collaborative Learning (CSCL) systems.
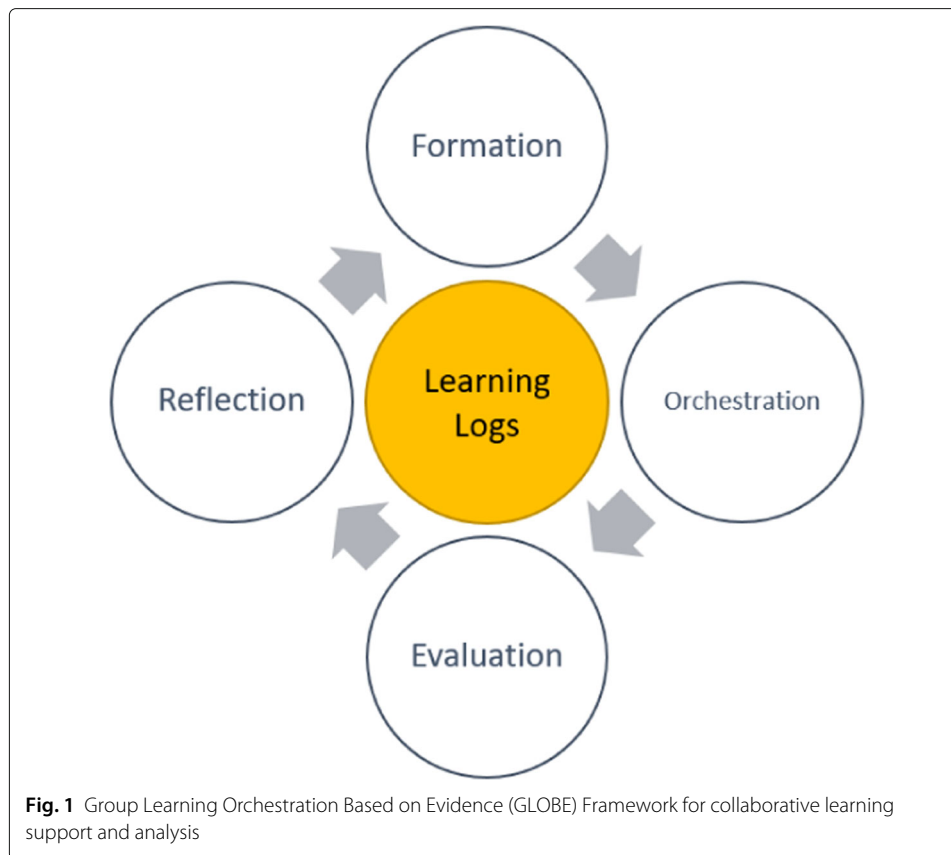
Keywords: Group formation, Computer-Supported Collaborative Learning (CSCL), Group work evaluation

Introduction

Collaborative learning is becoming increasingly prominent in educational activities since not only cognitive knowledge but also interpersonal skills such as critical thinking, problem-solving, and reasoning count in modern society (Stahl et al. 2006). In collaborative learning, students work together to complete a task or to reach team goals. (Dillenbourg 1999).

A framework organizing the research for collaborative learning support and analysis is put forward in Fig. 1. It is a circle composed of group formation, group work orchestration, group work evaluation, and reflection. For successful in-class collaborative learning, group formation is the fundamental component that determines the quality of group work (Wessner and Pfister 2001).

However, several obstacles might hinder the execution of in-class group work activities. For one group activity, a teacher needs to envision the lesson, enable collaboration,

Fig. 1 Group Learning Orchestration Based on Evidence (GLOBE) Framework for collaborative learning support and analysis

encourage students, ensure learning, and evaluate achievements (Urhahne et al. 2010). Just to form groups appropriately, teachers usually take more than 1 h on this trivial work and might get overwhelmed when using computer-supported tools. When it comes to evaluation, teachers need real-time support to get the performance of each group in a real-time manner. The problems of social loafing and free riding also bother instructors to give a fair evaluation to each student. Increasing self-assessment and peer-assessment methods are adopted since a teacher cannot monitor the whole class while the students participate in group work (Forsell et al. 2020).

Fortunately, the development of information facilities and increasing learning log data provide an opportunity. In recent years, learning analytics (LA) is introduced to measure, collect, analyze, and report data about learners and their contexts for improvement of their learning environment (Siemens 2012). Utilizing previous student-produced learning log data, we can do predictive analytics in educational settings (Ferguson 2012) thus affecting their performance (Macfadyen and Dawson 2012) and learning outcomes (Archer et al. 2014).

Given the above issues, valid support for executing and managing such activities in a timely and informed manner becomes imperative. In this research, we present a system that provides a solution to support teachers in group formation and analytics based on learning log data from BookRoll learning system (Flanagan and Ogata 2018). Furthermore, we implement the system to assist the teacher in conducting their group-based classroom activity in a school context. In the study, we examine the effectiveness of the

system by investigating the primary impact on the engagement and affective states of students. The specific research questions are as follows:

RQ1. How do the computer-formed groups affect the students' engagement in in-class group work?

RQ2. How do the computer-formed groups affect the students' affective states during in-class group work?

In the following sections, first, we review related works and position our research. Then, we introduce the architecture and functions of our technical support, followed by an empirical experiment in a real-school context. Finally, we provide the discussion and general implication for teachers and conclude.

Literature review

Computer-Supported Collaborative Learning

Computer-Supported Collaborative Learning (CSCL) is an emerging branch of learning sciences concerned with studying how people learn together with the help of computers (Stahl et al. 2006). For teachers, support from computers enables them to glimpse into students' performance instantly and give targeted guidance (van Leeuwen 2015). For students, CSCL facilitates peer discussion, leading to metacognitive, co-regulation, and social-emotional activities occurring to enhance learning effectiveness (Splichal et al. 2018). The application of CSCL runs through a broad variety of contexts throughout the process from creating groups, group regulation, in-group interaction to group evaluation, and reflection. For instance, kit-map generation is a typical activity where CSCL is frequently utilized for brainstorming and knowledge building (Manske and Hoppe 2016). Workshop such as programming projects is another application (Moreno et al. 2012) where students harvest collaboration skills. With the booming of online courses in recent years, CSCL has been applied to mobile learning and web-based contexts to promote communication (Boticki et al. 2020), especially for primary education (Dlab et al. 2020). Though several studies focus on the real-time application during group work, the group creation and evaluation is also critical and deserves our attention. This research will provide practice evidence of the CSCL implementation on group formation and evaluation for in-class activities.

Systems and algorithms for group creation

Collaborative learning with properly formed groups is found to outperform traditional teaching (Kyndt et al. 2013), while improper group formation parameters may raise several problems that lead to failure (Wang 2010). Therefore, forming a group that collaboratively learns is one of the most challenging tasks in the CSCL context. The characteristics of group members, the context of the group work, and the techniques used to form the group(s) are three main issues (Maqtary et al. 2019).

As for personal characteristics, knowledge and skill is the most commonly used attribute considered in group formation because of its direct effects on the final output (Abnar et al. 2012). Other attributes such as learning styles and personality (Zheng and Pinkwart 2014) are also used in previous research. Also, social issues such as relationships and roles are highlighted in recent studies (Yannibelli and Amandi 2011).

Regarding the context of learning, it is pointed out that the heterogeneity of learning groups differs with different pedagogical contexts (Manske et al. 2015). Studies indicated

that homogeneous grouping performs better in inquiry learning context (Lee Jensen and Lawson 2011) while learning effectiveness of heterogeneous grouping proves to outperform that of homogeneous one in didactic learning (Schneider and Blikstein 2015). In addition, the duration of group tasks is another attribute that affects the group formation process (Huang et al. 2009) and there is a division of static and dynamic groups for various contexts (Srba and Bielikova 2015).

Further, groups are formed by different techniques as is summarized in Table 1. The algorithm based on clustering using simple Euclidean distance measurements is the popular one and can fit various group formation purposes in both homogeneous (Christodoulopoulos and Papanikolaou 2007) and heterogeneous contexts. For example, research was conducted to form homogeneous groups in mobile collaboration using the K-means algorithm that put students in the same cluster together (Maqtary et al. 2019).

The semantic method is another idea that aims to form groups using semantic extraction from learner-generated content to create heterogeneous learning groups in terms of knowledge diversity based on textual similarity (Manske and Hoppe 2016). It can induce heterogeneity in semantic level which is hard for any methods using pure scores (Manske and Hoppe 2017). Furthermore, a semantic framework is presented to represent the interaction data of learners (Ounnas et al. 2007).

Evolutionary algorithms such as Genetic algorithm is a powerful solution which can compute multiple parameters by machine learning. An iterative process based on a genetic algorithm is done in the group formation process which is flexible to the number and type of the attributes (Abnar et al. 2012; Moreno et al. 2012). These researches model a fitness function with fairness and equity in terms of members' performance to ensure fair formation.

While multiple works discuss intelligent group formation algorithms in different contexts, few researchers integrate multiple algorithms into the same system and use data from multiple sources that are synchronized with that system. An integrated system that is designed for multiple contexts is introduced in this study.

Evaluating groups during their activity

The evaluation of group work is of necessity as well. In a data-rich environment nowadays, formative assessment (Strijbos 2011) is adopted since the computer promotes instant feedback and enriched information about people and context. Teachers can monitor the collaborative learning processes and gather information about individual performance

Table 1 Group formation techniques and its contexts

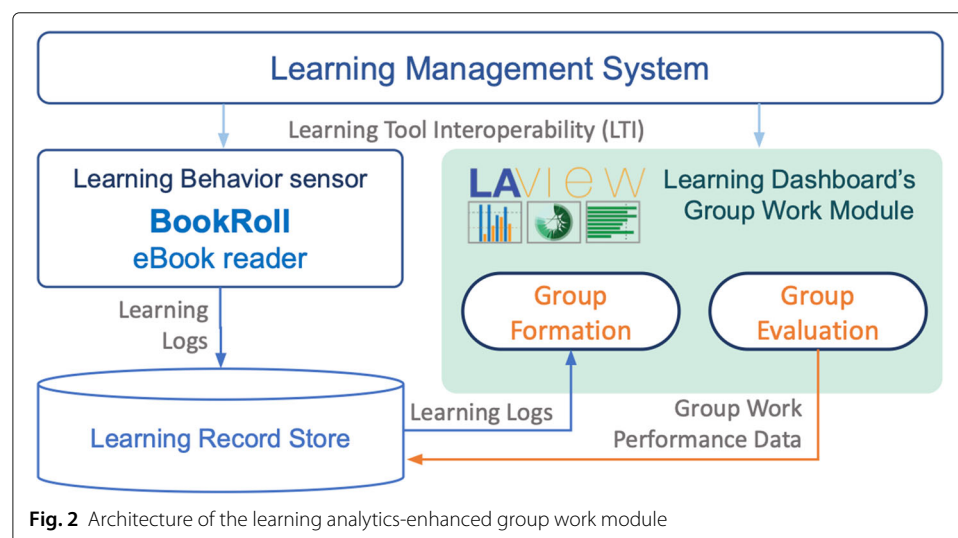| Technique	Context	Parameters
Clustering and distance method (Manske et al. 2015)	Inquiry-based learning scenario with concept map and text writing	Artifacts, particularly learning objects and the assessment of motivational scores
Discrete particle swarm optimization (Zheng and Pinkwart 2014)	Online classes of which the size is up to several hundred	Gender and personality type
Semantic method (Manske and Hoppe 2016)	Kit-map generation	Learner-generated artifacts, texts, concept maps and hypotheses
Evolutionary and genetic method (Abnar et al. 2012)	Peer discussion of dynamic programming problem	Programming skills, personality, learning style
Multi-objective optimization genetic method (Moreno et al. 2012; Abnar et al. 2012)	Course subject discussion of a computer programming	Student knowledge levels, student communicative skills, student leadership skills

and contributions to the group work (van Leeuwen 2015). Since we conduct the study in a face-to-face context, we focus on real-time evaluations. Oral communication is of vital importance as well as Speech Activity Detection (SAD) of collaborative behaviors is used to predict the quality of small-group collaboration (Kim et al. 2020). Features capturing information about the number, duration, and location of the speech regions are used to evaluate collaborative activities (D'angelo et al. 2019). These researches show high potentials of students' utterances during group work to collect and glimpse basic ideas about engagement in a real-time manner.

Except for traditional evaluation that highlights personal knowledge, affective parameters also need attention in the collaborative context (Milton 1965). How the participants feel about the activity is another real-time indicator, which is measured as affective states (D'Mello et al. 2008). Positive affections like joy and vitality under the speaking indicate favorable affective states for the group work, while negative affections like anger and calmness may indicate low affective states within the group. These affective states can be detected by monitoring conversational cues, gross body language, and facial features (D'mello and Graesser 2010).

Group formation module within a learning analytics framework

System overview

Figure 2 depicts the Learning Evidence Analytics Framework (LEAF) that lays the foundation for the system in this study (Ogata et al. 2018). The LA Dashboard fetches the learning log data, visualizes data, and models them for analysis (Majumdar et al. 2019). As a part of this LA dashboard illustrated in Fig. 2, the group work support module acquires student model data from LRS that covers learning log data from behavior sensors such as the BookRoll system and Moodle platform via LTI. The group formation system uses these data as input parameters to generate groups (Boticki et al. 2019), and in turn, the group formation results work as input to the LRS database. Users can access the group module from the LA dashboard and get visualized multiple student model data as well.

Fig. 2 Architecture of the learning analytics-enhanced group work module

Table 2 General components of group formation and evaluation system
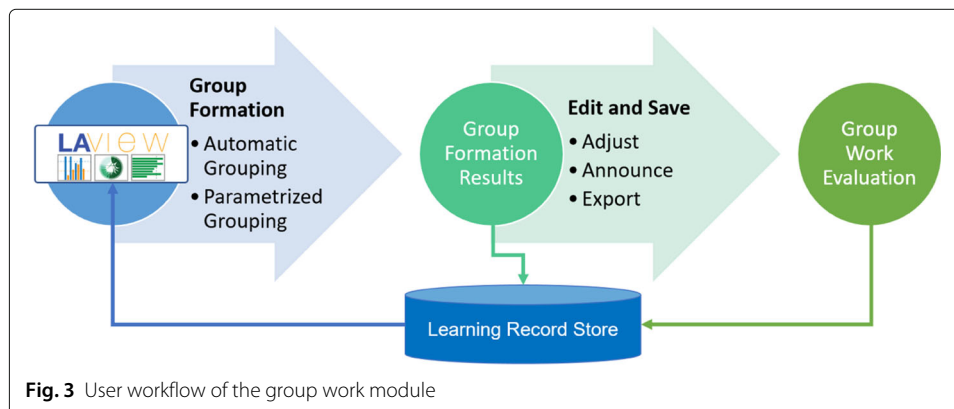
Component	Need and requirement
Group Creation Parameter Setting	Requires users to get students information; know the available data sources, algorithms, and parameters; conduct settings; and collect purpose.
Created Group Member Listing	Requires visualized listing of group formation result, editing of group information, manual adjustment of group members and export channels.
Group Work Evaluation	Requires indication of previous performance of stunted for reference, the easy-use group evaluation interface and performance prediction.

Based on the investigation of previous work, we present three main functional components required in a Group Formation system in Table 2. Following the order of general components, we introduce the three modules of the system with the interfaces.

Figure 3 shows the workflow of the user. Teachers can enter the group formation module in the LA dashboard and start by choosing either automatic or parameterized grouping. The automatic grouping will generate the group using the default heterogeneous algorithm based on engagement parameters and directly get results. For parameterized grouping, the teacher needs to decide and set the group formation parameters that best suit the specific learning activity such as group size, group algorithm, and parameters from different data sources in LRS. Once the group formation results are generated, the teacher can manually adjust group members and export results into CSV files. During or after group work, the teacher can grade the performance of group work and give feedback to the students. Meanwhile, the group configurations and performance data graded by the teacher are synchronized into LRS for further learning analytics.

Group creation parameters and algorithms

The parameters and algorithms used are key parts of the group formation module of the system. Teachers can use the group formation parameter console to set parameters and algorithms listed in Tables 3 and 4 respectively. Even if there is no data, a random algorithm is available. The homogeneous and heterogeneous algorithms used in the system adopt genetic algorithms with the fitness function of the minimum square. Using relationship data, the algorithm enables students with good relationships (type 1) to be assigned to the same group. Conversely, the negative relationship (type 2) will be considered to separate students. Figure 4 shows an example of relationship data. In line with this data,

Fig. 3 User workflow of the group work module

Table 3 Parameters used in the group formation process

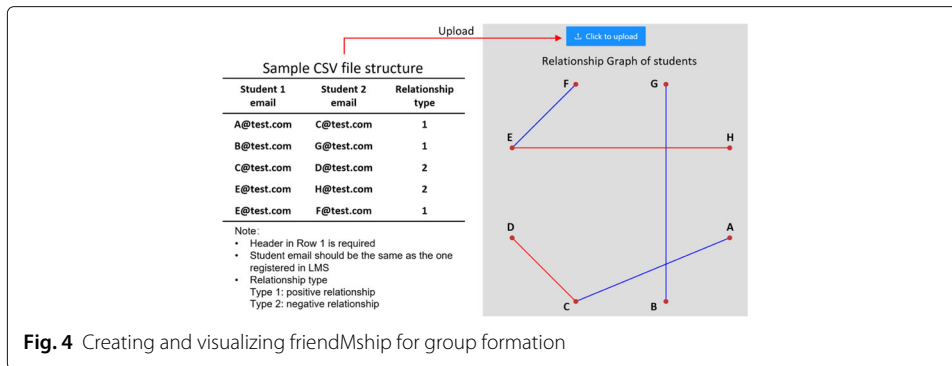| Parameter	Data source	Description
Engagement	Learning log data	The parameter records the time spent, number of notes, and completion rate using the learning platform.
Course Score	Dashboard	The parameter records scores from uploaded CSV tables.
BookRoll Score	BookRoll quiz	The parameter records the summation of selected BookRoll quiz scores.
Moodle Score	Moodle quiz	The parameter records scores of Moodle quiz.
Group Score	Group formation system	The parameter reflects students' previous performance in group, gathered as part of group grading.
Friendship	CSV files uploaded	The parameter describes the close or bad relationship between students.

student A and C, student B and G, and student E and F will be given priority to be together while student C and D and E and H will be separated. Once the relationship data indicating positive and negative relations between students is uploaded, a graph shown in Fig. 4 will be visualized. The red lines indicate pairs with poor relationships and blue lines indicate that with good relations. Each red dot represents a student and the name will be displayed with the mouse moves on it.

In the jigsaw algorithm which focuses on multiple scores, it distributes students with different ranks in different score columns into the same group to heterogeneity. As is illustrated in Fig. 5, students are ranked by each score respectively, where the students are selected evenly from those who have high ranks. Take groups of 2 members as an example, students H, A, F, and G are selected and assigned to groups 1, 2, 3, and 4 (cells with orange background). Then, students B, C, E, and D are successively selected in the second round (cells with blue background) and so on. If the student has already been grouped, it will jump to the next highest score holder in the corresponding column (cells with yellow background). The result of the groups is shown in Fig. 6.

Figure 7 is the parameter setting page for teachers to choose the grouping strategy depending on different purposes. Teachers can see the existing student's roaster and adjust the list to consider in the grouping process, choose grouping algorithms and

Table 4 Algorithms used in the group formation process

Algorithm	Algorithm description	Available input parameters
Homogeneous	The algorithm groups students with similar values of a parameter.	Engagement (Learning time for preparation, achievement rate, number of markers and memos), average scores of quizzes from different LRS before group study, group score (data entered for evaluation)
Heterogeneous	The algorithm groups students with different values of a parameter.	Engagement, average score of quizzes from different LRS before group study, group score
Jigsaw	Heterogeneous grouping based on multiple scores	Multiple scores of quizzes from different LRS before group study
Friendship	The algorithm groups students who are friends as identified by the teacher. Uploaded CSV file is required.	Friendship (obtained from past group learning data or teachers' input in CSV)
Random	The algorithm groups students randomly.	None

Fig. 4 Creating and visualizing friendMship for group formation

student model parameters in this step. An automatic grouping parameter is set as a default.

Figure 8 shows the list of previous group formation records list. The list provides the group formation name, purpose, and time and different icons represent different algorithms adopted. Teachers can browse and search for group formation to explore their previous group formation settings and the group grading for next group work planning.

Group member visualization

When the groups are formed in line with the selected parameters and algorithm, the result page will be visualized for teachers. The student list is intuitively organized by groups with the color indication of his previous performance in group works. Teachers can adjust the results by moving students and score the group performance here. The parameters and algorithm used for group formation are available and teachers can change group formation name and purpose as well. Meanwhile, the group formation results can be exported as an Excel file for offline use.

Figure 9 shows an example of a created group member list that depicts the results of a heterogeneous grouping algorithm operation. Traffic-light colors are used to give an indication of previous group work referring to perfect, good, and poor performance. If there is no data for previous group work performance, the color will be white.

Group work evaluation

As for the group evaluation module, the metrics of the three indicators are listed as follows:

Student	Score 1	Score 2	Score 3	Score 1 rank	Score 2 rank	Score 3 rank
A	1	8	3	H	A	F
B	2	7	3	G	B	G
C	3	6	4	F	C	H
D	4	5	4	E	D	C
E	5	4	2	D	E	D
F	6	3	6	C	F	A
G	7	2	5	B	G	B
H	8	1	5	A	H	E

Fig. 5 Demonstration of an example of jigsaw algorithm

Fig. 6 A sample result of jigsaw algorithm

- Collaboration quality: Frequency of interaction and communication occurring during group work, participation of members, and rational division of labor.
- Speed and efficiency: Whether each sub-task is finished on time and reasonable time management.
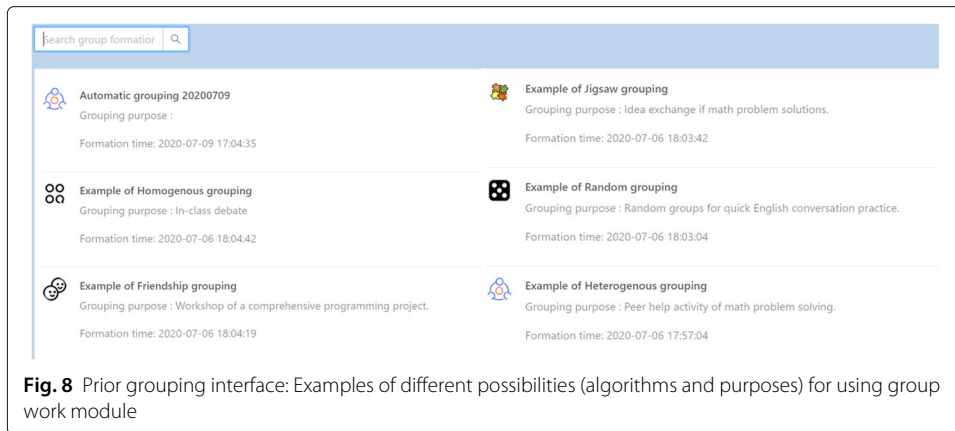- Final output: The quality of final outputs and artefacts of group work.

Not only summative but also formative indicators such as collaboration quality are considered in the system. Since the evaluation should be based on the whole group's work to avoid social loafing and free riding, the grade is given to the whole group, not individuals. Teachers should rate each group's performance in three indicators. In turn, the scores in three indicators are stored as part of the group user model giving an overall estimation of students' previous group work performance, which can work as an input parameter for the next group formation.

Classroom implementation

Learning context

The study was conducted in a primary school maths problem-solving class covering several topics. For two different classes, two different teachers conducted the class respectively but the topic is the same. Two classes firstly underwent activities 1 to 3 with teacher-formed groups as baseline conditions. Then, the group formation was changed and done according to the system, and activities 4 to 7 were conducted as experiment class. Each class is of the same length and the topics are in the same order. It maintains that data from each class are comparable. The main data for analysis of the research comes from the voice records throughout the class via USB headsets and microphones. In total, 13,462 pieces of voice data which cover text and affective scores (6030 pieces for class 1 and 12,767 pieces for class 2) were collected. After data cleaning, the data for analysis

Fig. 7 Parameter setting options in Group Formation Module in LAView

Fig. 8 Prior grouping interface: Examples of different possibilities (algorithms and purposes) for using group work module
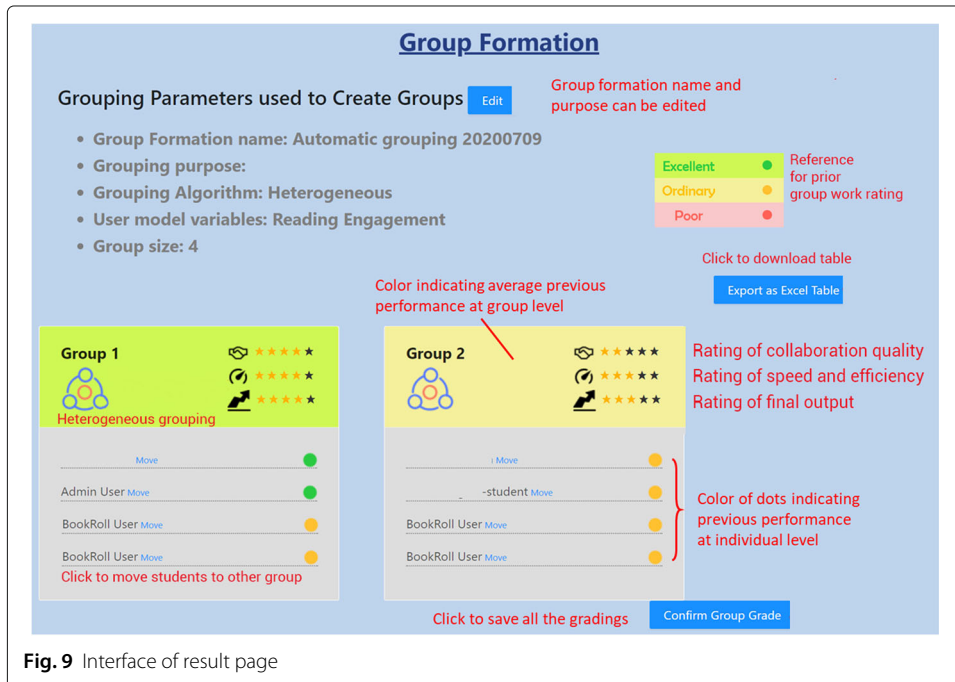
covers 7 lecture topics of 11 in-class activities (see Table 5, "TG" means groups formed by the teacher, and "CG" means groups formed by computer).

Participant

The experiment was conducted in a primary school in two grade 5 classes. There are 32 students in 12 groups for class 1 and 33 students in 12 groups for class 2. However, not all of the 65 students participated all the class due to uncontrollable issues.

Learning design

The in-class group work adopts the "jigsaw learning method" consisting of two different phases (Fig. 10). Each student will work in a "knowledge exploration phase" and a "knowledge exchange phase" during one class, which corresponds to two different group combinations. In the knowledge exploration phase, students work on a solution with the

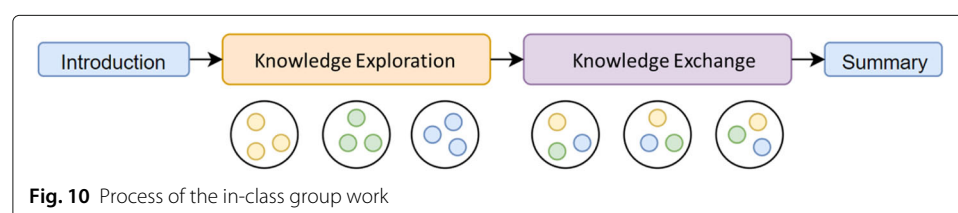**Fig. 9** Interface of result page

Table 5 Summary of data collection

Activity	Topic	Type	# of students	# of voice data
A1 (Initialization)	Square	TG	32	1611
A2 (Trial)	Square	TG	65	1480
A3 (Baseline)	Multiplication	TG	65	2325
A4 (Trial)	Quantity per unit	CG	65	5005
A5 (Intervention)	Percentage	CG	65	3041
A6	Percentage	CG	33	2084
A7	Percentage	CG	33	2223

same idea. They discuss and check their solutions with members within the knowledge exploration group and illustrate ideas to each other. After that, students from different knowledge exploration groups go to knowledge exchange groups and explain the idea to those who solved the problem differently. In the knowledge exchange phase, students exchange ideas and talk about different solutions.

Take the topic "the square of a trapezoid" as an example, the system firstly collects data from different sources and then forms groups accordingly. A pre-test about the estimation of triangle squares is conducted at the BookRoll system to confirm the level of understanding of these learned items. The test results are used as input parameters of the group formation. Meanwhile, course scores from the LA view dashboard indicating communication skills and performance data of previous performance scores relating to topic "Square" are extracted to conduct group formation in the system. Besides, relationship data are created by teachers and uploaded in the tab of "relationship" on the group formation parameter setting page. In this context, the system first uses the friendship algorithm to group students with positive relationships, then groups the rest of the students using the jigsaw algorithm as is illustrated in chapter 3.

Before the class starts, the tablets and headset microphones are prepared and set in the classroom. At the commencement of the class, the teacher writes the goal of the class "Square of a trapezoid" on the blackboard and puts forward a specific problem of calculating the square of a trapezoid. The problem is to be solved throughout the class thus motivating students to learn. Then, the group work activity starts and the utterances are recorded for each student respectively. For the topic of "the square of trapezoid", each knowledge exploration group will be asked to discuss either of the following solutions: making a parallelogram, dividing into two triangles, dividing into a triangle and a parallelogram. And in the knowledge exchange phase, students in one group will share all three solutions with other members so that all students know the three solutions. Finally, the teacher gives the summary of the whole class and students write down three ways of calculating the square of the trapezoid on the blackboard. After the class, a feedback seminar

Fig. 10 Process of the in-class group work

is conducted where teachers reflect on their teaching experience and share their doubts and feelings.

System usage

In the implementation, input parameters from three data sources were considered based on related works and teachers' opinions. The jigsaw algorithm was applied using the following parameters:

- Bookroll quiz scores: The pre-test indicating the pre-knowledge of the learning subject was done on online textbook Bookroll using its quiz function and the quiz scores are acquired as an important input source of the group formation.
- Course skill scores: Communication skills, way of thinking, and academic skills are provided as scores by teachers and uploaded in the LA view dashboard.
- Friendship data: The friendship data indicating both positive and negative relationships of students is uploaded in the group formation tool since the teacher stressed that students with negative relationships should not be grouped together.

Research study

In this study, we mainly focus on the primary impact on the engagement and affective states of students in the groups formed by the system. We explored the difference between the group work based on teacher-formed groups and computer-formed groups by practical experiment.

Experiment design

To make a comparison between groups formed by the teacher and by the system, we adopted a within-subjects design (A-B design). We conduct the study with a single cohort of primary school students in grade 5; however, the indicators observed are at a group level that keeps changing based on teacher-generated and computer-generated grouping, the A and B conditions. Activities A2 and A4 is the first attempt for each condition; to reduce the novice effect, we choose activity A3 (applied problems of multiplication) and A5 (applied problems of percentage) for both classes 1 and 2 for the data analysis in this research. We assume activities A3 and A5 are similar and comparable since both of them focus on math problem-solving in similar topics.

Data collection

For the utterance data indicating students' engagement, the duration of each speaking was recorded and then the speech data was textualized by speech-to-text API. We divided the text into tokens (meaningful words) by Node.js TinySegmnter API for Japanese tokenization (Kudo 2016). Then, the words are counted as the number of tokens. The teachers' speech data was filtered before the analysis as well.

The affective scores data indicating affective states are transformed from utterance data as well by pattern recognition API. Four affective states, joy, vitality, anger, and calmness, were computed into scores for each piece of utterance. Joy indicates the student works in a positive mood. Vitality denotes how active the student performs in the group work. Anger

implies conflict within group members. Calmness represents low engagement and low motivation. Each affective score is standardized into the range of 0 to 1 before analysis.

Data analysis

To explore the difference of the knowledge exchange phase between teacher-formed groups and computer-formed groups and answer research question 1, we do analysis at both group level and individual level. Comparing overall mean provided a group level aggregation of engagement, we look into the effect of intervention condition (CG) in three indicators: times of utterance, duration of utterance, and the number of tokens. Since the data of the three indicators do not satisfy the normal distribution according to the Shapiro–Wilk test ($p < 0.05$) (Shapiro and Wilk 1965). We adopt non-parametric tests to measure the significance of the difference. Mann-Whitney U test is conducted and the effect size is calculated respectively for the three engagement indicators.

Further analysis was done to understand transitions of cohorts of specific engaged students within phases of one activity or across activities. Individual learner's engagement category, based on their speaking duration, was considered to do this analysis. The transitions in engagement categories were looked at from two different perspectives. One perspective is between two activities for each phase and overall. Such analysis was afforded by the iSAT tool which could visualize transition patterns across phases with SAT Diagram (Majumdar and Iyer 2014).

The affective scores of two independent samples are compared by independent t-test to answer research question 2. Since the Shapiro–Wilk test of affective indicators ($p = 0.053 > 0.05$ for anger, $p = 0.299 > 0.05$ for calmness, and $p = 0.511 > 0.05$ for joy) shows normal distribution except vitality ($p < 0.05$), an independent T-test is done on three indicators and a Mann-Whitney U test to vitality score. The null hypothesis establishes that the means of the affective scores are of equivalence, and correspondingly, the alternative hypothesis establishes that the means are of difference.

Result and inferences

Engagement

Knowledge exchange phase

As shown in Table 6, all three indicators of group work engagement suggest significant improvement in the intervention condition (CG). Groups formed by the system have more times of utterance (M = 110.4, SD = 58.85), longer utterance duration (M = 734.02, SD = 375.52), and also more meaningful tokens (M = 609, SD = 340.89) in comparison with groups formed by teachers: T (M = 35.46, SD = 13.98); D (M = 230.11, SD = 108.03); N (M = 256.38, SD = 89.33). The effective size of the three indicators are 0.452, 0.405, and 0.437 respectively, which indicates a medium to large effect (Cohen 1988).

Figure 11 shows the transition graph of utterance duration indicator in the knowledge exchange phase between two conditions.
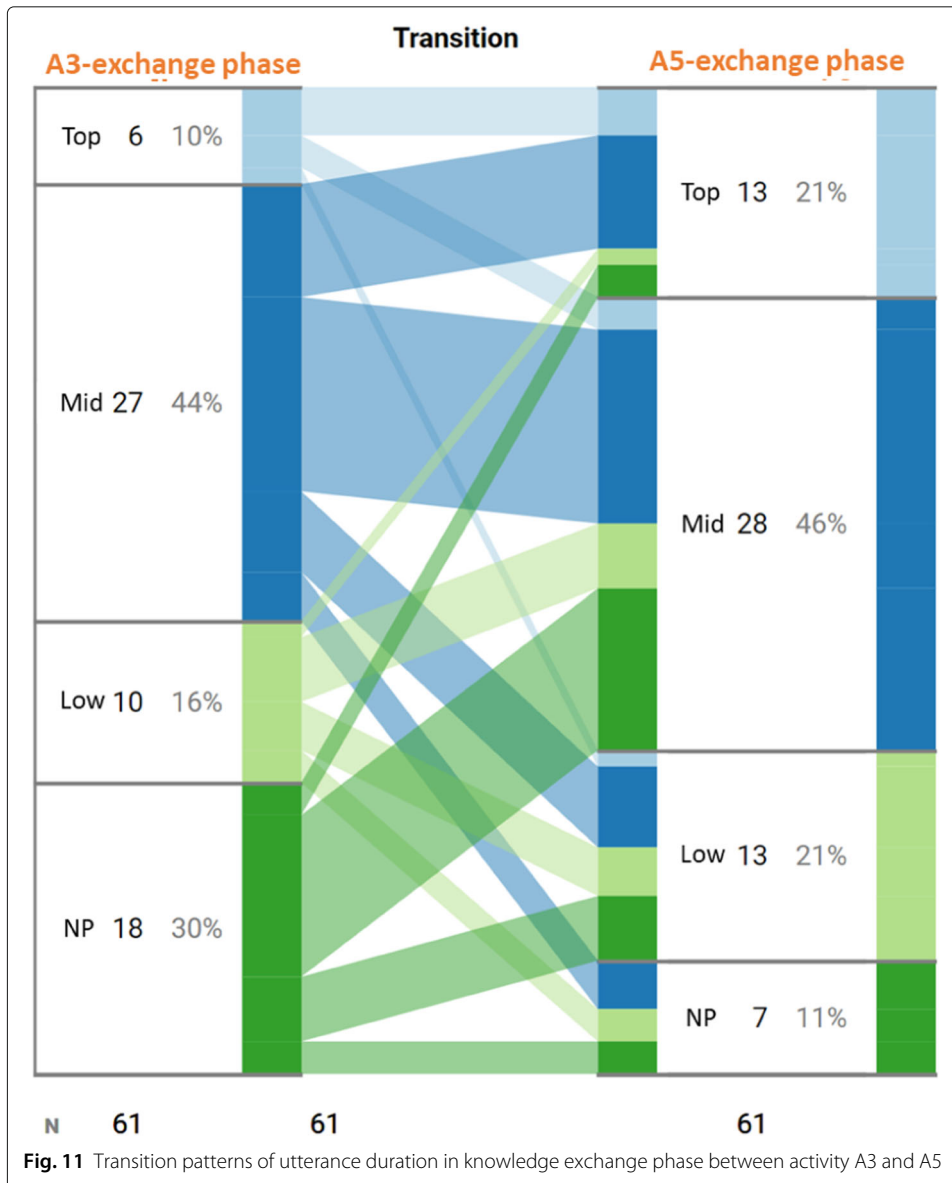
In the transition graph, three strata (Top, Mid, and Low) are defined for each phase independently and presented in Table 7. The Top-Mid cutoff is delimited using mean plus standard deviation and Mid-low cutoff by mean minus standard deviation. NP (Not-participate) layer indicates absence in this phase. We can see more students start to

Table 6 Difference in engagement indicators for knowledge exchange phase

		N	Mean	SD	p	Effect size
Times of utterance (T)	TG	24	35.46	13.98	0.000***	0.452
	CG	20	110.4	58.85		
Duration of utterance (D)	TG	24	230.11	108.03	0.000***	0.405
	CG	20	734.02	375.52		
Number of tokens (N)	TG	24	256.38	89.33	0.000***	0.437
	CG	20	609	340.89		

***$p < .001$.

participate in discussion in computer-formed groups since the transition from NP to Top and Mid account for 19% for the knowledge exchange phase. Meanwhile, computer-formed groups encourage active students to even speak more than the baseline condition. It is indicated that more students' utterance duration reaches a high level in A5 activity which is based on computer-formed groups.

Fig. 11 Transition patterns of utterance duration in knowledge exchange phase between activity A3 and A5

Table 7 Cutoff of three strata of the utterance duration transition graph

Strata	A3-knowledge exchange	A5-knowledge exchange
Top-mid cutoff	194.25 s	492.80 s
Mid-low cutoff	62.62 s	50.92 s

Knowledge exploration phase

Table 8 shows the result of the Mann-Whitney U test for idea exploration group work on this regrouping activity at the group level. Converse to the knowledge exchange phase, it is indicated that for the engagement indicators, teacher-formed groups perform better in this context in all three indicators with small effect sizes of 0.319, 0.322, and 0.303 respectively.

A simple observation of transition of the duration of utterance is also implemented in the reshuffled group (Fig. 12). We found that still, 15% of students from Mid, Low and, NP layers in teacher-formed groups come to Top layer in knowledge exploration activity, which makes the percentage for Top layer increase in computer-formed groups. However, 13% of students in Mid layer kept silent without any utterance in the computer-formed groups.

Affective states

Figure 13 depicts the result of the test on the affective scores at group level and the mean of each standardized effective score for each group is labeled on the bars. As is indicated in the figure, the joy and vitality affection present the same pattern that the experiment class where groups are formed by the system has a higher score of these positive affections. On the contrary, regarding negative affections, calmness and anger denote the opposite result, with the control group higher scores. However, only the difference in joy proves to be at a significant level in the statistics ($t(24)=0.004 >0.05$) and the null hypothesis is rejected. For calmness ($t(24)=0.143 <0.05$), anger ($t(24) = 0.777 >0.05$), and vitality ($p=0.066 <0.05$, effect size=0.079, indicating very low effect), the null hypothesis cannot be rejected within a confidence level.
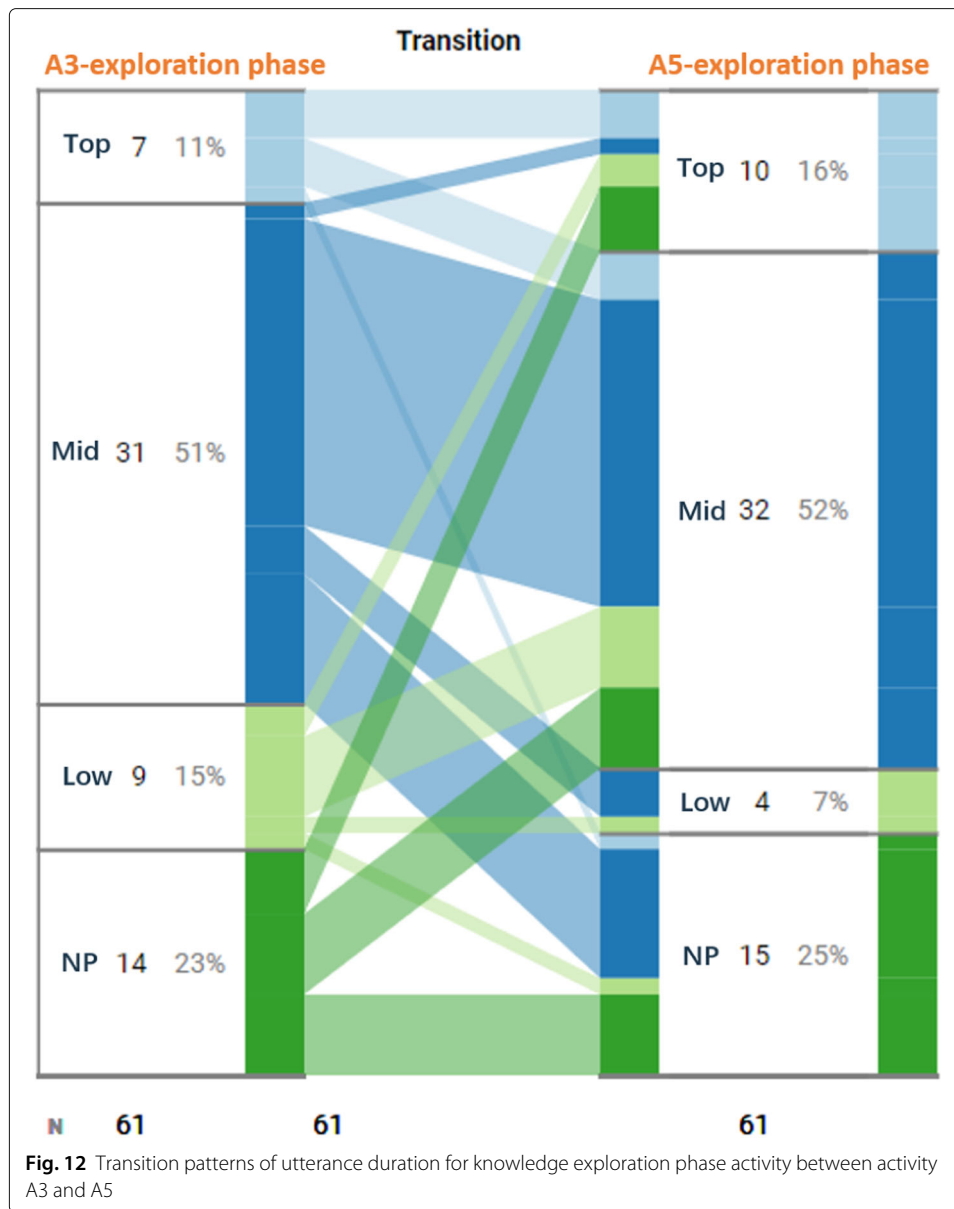
Discussion and conclusion

RQ1: How does the computer-formed groups affect the students' engagement of in-class group work?

The results show the difference in the process of the group work between groups formed by teachers' experience and by evidence data using the system. Generally, each group speaks more, and the duration of utterance increases in the computer-based groups. This

Table 8 Difference in engagement indicators for knowledge exploration phase activity
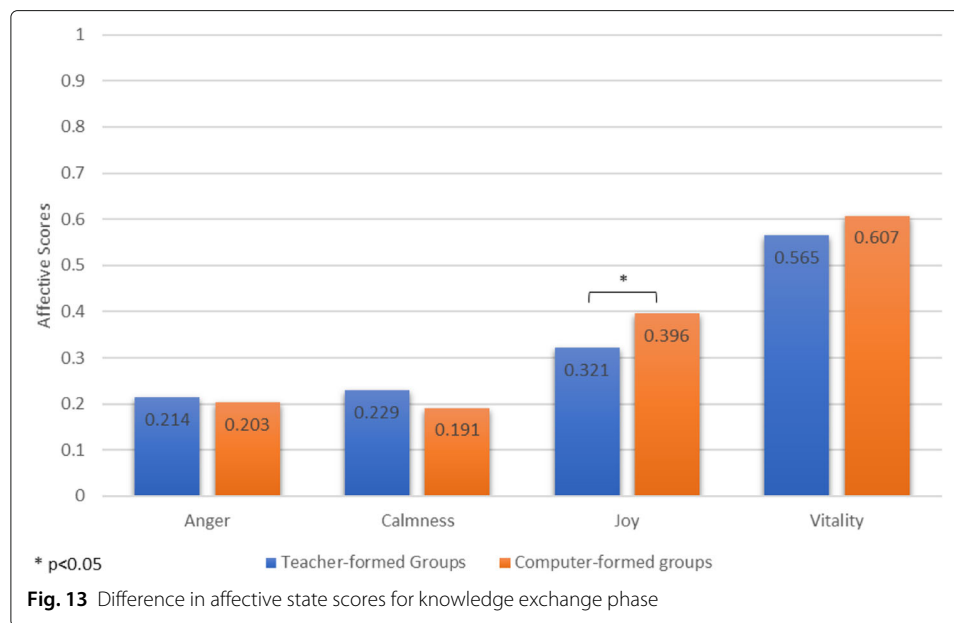
		N	Mean	SD	*p*	Effect size
Times of utterance (T)	TG	18	53.94	24.08	0.000***	0.319
	CG	20	26.6	24.89		
Duration of utterance (D)	TG	18	352.77	135.41	0.000***	0.322
	CG	20	177.69	132.49		
Number of tokens (N)	TG	18	312.17	203.80	0.000***	0.303
	CG	20	136.5	135.89		

*** $p < .001$.

Fig. 12 Transition patterns of utterance duration for knowledge exploration phase activity between activity A3 and A5

finding supports the superiority of the system for idea exchange activity to arouse motivation and facilitate engagement of students. The parameters for group formation may be a key factor that determines this phenomenon. That is to say, the diversity of communication skills, pre-knowledge of the learning topic, and previous academic performance catalyze the atmosphere and facilitate interaction for idea exchange within heterogeneous groups. It is also grounded in the research in the area of the Zone of Proximal Development (ZPD) and potentially promotes the construction of knowledge and an elevated level of the mutual understanding of a topic (Nyikos and Hashimoto 1997). The finding also agrees with the recent work that presents the effectiveness of heterogeneity of the student cohort in workshop group activities (Sivaloganathan et al. 2020). Besides, we can see that the difference reverses in the reshuffled groups for knowledge exploration phase activity. On the one hand, it supports the effectiveness of the system and parameter settings in

Fig. 13 Difference in affective state scores for knowledge exchange phase

the knowledge exchange condition. On the other hand, we cannot deny the fact that the system is still short of flexibility in the regrouping context.

In terms of the transition graph, we can infer that the new combination of group members encourages active students to even speak more and in turn facilitate low-performance students to participate. Even for the reshuffled group in a regrouping context, the percentage of top-level students increases in the computer-formed groups, which can be partially attributed to the work of friendship data.

RQ2: How does the computer-formed groups affect the students' affective states during in-class group work?

As for affective states, students act more positively in the groups formed by the system where their utterances showed more positive affective states such as joy and vitality. Also, students performed less reserved and less irritated in the experiment groups as is indicated in the scores of calmness and anger. The difference of joy affection reaches a significant level, we can infer that the computer-formed groups bring about more happiness for students, thus promoting the initiative of utterance and high engagement in the group work. According to the teachers' feedback, it is indicated that the novelty of the new group combination motivates students to speak more and participate more actively. We can also conclude that friendship-priority grouping strategy utilizing friendship data reduces the conflict within group members because trust relationship and the group's willingness to handle group work challenge was positively related to individual student's group work self-efficacy (Du et al. 2019). However, since the difference in vitality, calmness, and anger do not reach a significant level, the effect of the new group composition on these affections is limited.

Implication for teaching

Due to the busy schedule of the teachers, an informal interview with them was conducted to gather feedback after they used the system. The overall impression was positive.
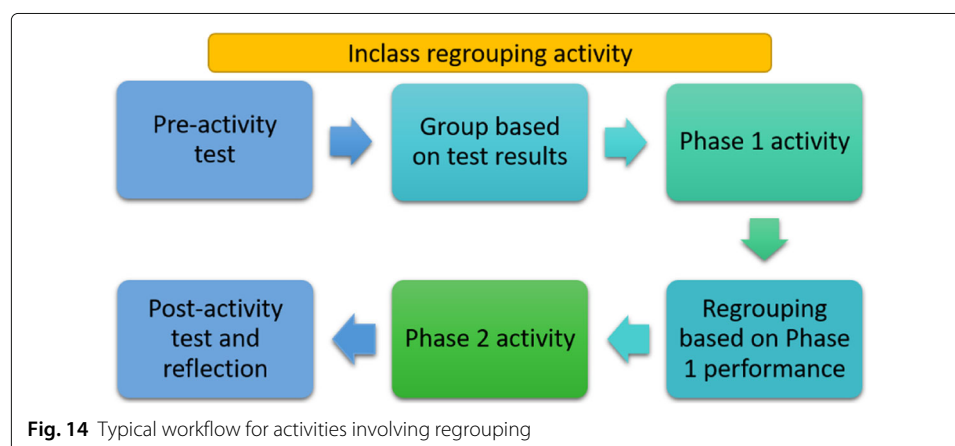
Teachers mentioned that unexpected combinations of students which broke the teachers' prototypes were discovered. Furthermore, teachers found new qualities about students and some students demonstrated leadership which is not found in ordinary classes, though they still have some doubts and as well. Nevertheless, there is a possibility that the parameters provided not enough or not suitable for all the contexts of group formation. Therefore, it is imperative to discuss implementation potentials in further context.
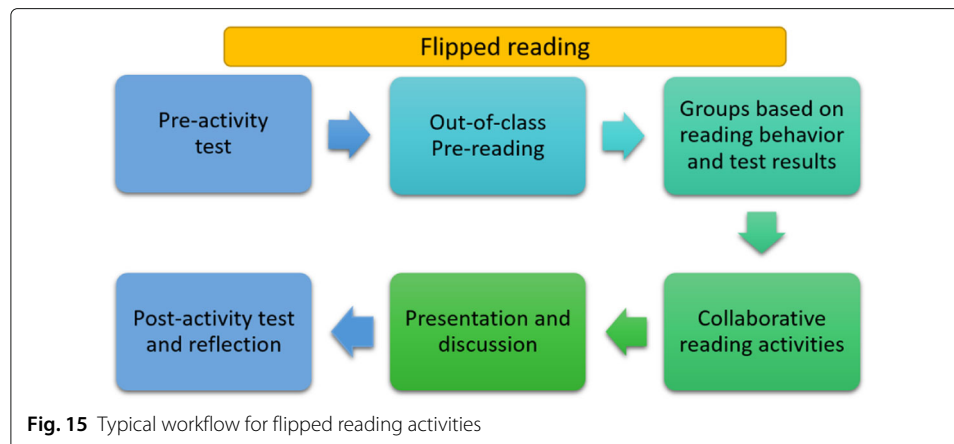
The system can be applied to broader pedagogical scenarios where teachers can use the tool. For example, the system can support more complicated group work activities like multi-phase in-class regrouping activities beyond the one illustrated in this study (Fig. 14). Before the class, the teacher can assign an online pre-test to students and then form groups based on prior knowledge indicated in the test. Since the system can form groups in seconds, it is convenient for teachers to create groups just in class for different phases of activity for multi-phases activities, even utilizing the performance data of the previous phase. The workflow can be applied not only in the maths problem-solving, but to other forms of collaborative problem solving (CPS) (Pöysä-Tarhonen et al. 2018).

Flipped reading is another example. Using learning logs from reading behaviors and records from LRS, the teacher can conduct flipped reading classes using the system (Fig. 15). Since rich learning logs are indicating the reading skills, preference of students, the integration of reading data makes it easy for teachers to generate homogeneous or heterogeneous groups using data regarding reading logs. The teacher can group students with similar reading habits or preferences within the group work. During the class, there can be multiple collaborative reading activities such as kit-build concept map (Hirashima et al. 2015), peer help of reading comprehension, and topic-based collaborative writing (Bremner 2010).

Limitation

Some limitations are identified in the present study for consideration. Regarding the system development, the reshuffle method proved to be of low performance for regrouping activities, which calls for improvement using different strategies. As for the experiment design, the learning topic is not perfectly identical, so the result may be affected by the topic of the class activity. Some students did not speak even a word through the whole class or across an activity phase, which makes it hard to explain the results. Case studies may be of necessity to inspect the reason behind their silence.

Fig. 14 Typical workflow for activities involving regrouping

Fig. 15 Typical workflow for flipped reading activities

Besides, the precision of the transition from voice collected in class to textual data (textualized data divided by all entries of utterance record) is between 40 and 50%, which limits the deeper analysis of the specific content of the utterance. With the available data, we conducted a basic analysis of the sound features to get an initial indicator of participant's motivation and engagement in the learning activity. However, the pattern recognition API directly coded the emotions and did not require tokenized words from the speech. Anyways in our specific context, the words were mostly limited to nouns and digits. This restricted further semantic analysis of the utterances in our current study. To make further investigation of speech signals, not only the overall duration of speech but also the spurts (Smith et al. 2016), defined as regions of uninterrupted speech, should be considered for deeper analysis. Also, more synchronized multi-modal signals are expected to catch more accurate features. For instance, the Collaboration Literacy Feedback framework including body posture and facial features provides an instructive reference for related research (Kim et al. 2020). As for the interview for teachers, we could only conduct an informal one over a group video call online due to time and access limitations to directly contact them during the period of the pandemic. Finding reasons related to ease of use by the teachers deserve further investigation, which is part of our future agenda.

For the evaluation module, which is not used in this experiment, we adopted the group assessment that only relies on the teacher's assessment. The disadvantage is obvious that it is hard to track each member's contribution and real-time performance, thus causing social loafing and free riding. The trivial way for teachers to grade the performance group by group is not user-friendly enough. A combination of teacher evaluation and peer evaluation will provide a solution which is recommended as other researchers' work (Forsell et al. 2020).

Contribution and future work

The paper provides a feasible solution to conducting in-class group work by helping teachers divide students into groups efficiently for better group work performance. It makes an instructive technical contribution to the research on group work support systems in the CSCL field as well. An experiment to primarily test its performance was conducted as a scientific investigation, thus providing empirical evidence to the practice

of CSCL systems. By using its visualization support, teachers can compare students' performance in group work and make more informed group formation decisions in their subsequent learning designs. Compared to related work, the system proves novelty in that it integrates multiple algorithms into one same system and combines data from multiple sources that is synchronized with that system, which is designed for application in multiple contexts.

In the further study, the implementation of the system will be extended to different activities and contexts such as regrouping activities and flipped reading mentioned in the discussion part. Besides in-class practice in school, contexts like university courses and remote education level a field for group work researches as well. Meanwhile, a more intelhligent reshuffling method will be imported to enhance the flexibility to more contexts. As is pointed out in the first chapter, the research for group work support is not only confined to group formation but also covers orchestration, evaluation, and reflection. As for the orchestration phase, real-time evaluation based on speech-to-text API is under investigation. To evaluate the performance of the group work, a system that combines the teachers' evaluation and peer evaluation are on schedule. Furthermore, in the reflection phase, utilizing the accumulating group formation and performance data, group work analytics and machine learning for optimized algorithm recommendation become possible.

Abbreviations
CSCL: Computer-supported collaborative learning; LA: Learning analytics; LMS: Learning management system; LRS: Learning record store; LTI: Learning tools interoperability; TG: Teacher-formed groups; CG: Computer-formed groups; iSAT: Interactive stratefied attribute tracking

Research involving human participants and/or animals
Not applicable

Authors' contributions
LC designed and developed the system, drafted the initial manuscript, and performed data analysis. RM provided insight and contributed to editing of the manuscript. HO provided supervision of the research. All authors read and approved the final manuscript.

Availability of data and materials
Not applicable

Declarations

Consent for publication
All the students participated in the study as the part of their regular mathematics course and the protocol of study was approved by the ethical standards of the school for academic research reporting.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Graduate School of Informatics, Kyoto University, Yoshida Nihonmatsu-cho, Sakyo-ku, 606-8501 Kyoto, Japan. [2]Ogata Laboratory, Academic Center for Computing and Media Studies, Kyoto University, Yoshida Nihonmatsu-cho, Sakyo-ku, 606-8501 Kyoto, Japan.

References

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

Archer, E., Chetty, Y.B., Prinsloo, P. (2014). Benchmarking the habits and behaviours of successful students: A case study of academic-business collaboration. *The International Review of Research in Open and Distributed Learning*, *15*(1), 62–83.

Boticki, I., Akçapınar, G., Ogata, H. (2019). E-book user modelling through learning analytics: the case of learner engagement and reading styles. *Interactive Learning Environments*, *27*, 754–765.

Boticki, I., Uzelac, N., Dlab, M.H., Hoić-Božić, N. (2020). Making synchronous CSCL work: a widget-based learning system with group work support. *Educational Media International*, *57*(3), 187–20.

Bremner, S. (2010). Collaborative writing: Bridging the gap between the textbook and the workplace. *English for Specific Purposes*, *29*(2), 121–132.

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, 2nd ed. Hillsdale, NJ: Lawrence Earlbaum Associates.

D'angelo, C.M., Smith, J., Alozie, N., Tsiartas, A., Richey, C., Bratt, H. (2019). Mapping individual to group level collaboration indicators using speech data, In *13th International Conference on Computer Supported Collaborative Learning-A Wide Lens: Combining Embodied, Enactive, Extended, and Embedded Learning in Collaborative Settings, CSCL 2019*: International Society of the Learning Sciences (ISLS).

Dillenbourg, P. (1999). What do you mean by collaborative learning? *Collaborative-learning: Cognitive and Computational Approaches*, 1–19.

Dlab, M.H., Boticki, I., Hoic-Bozic, N., Looi, C.K. (2020). Exploring group interactions in synchronous mobile computer-supported learning activities. *Computers & Education*, *146*, 103735.

D'mello, S.K., & Graesser, A. (2010). Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features. *User Modeling and User-Adapted Interaction*, *20*(2), 147–187.

D'Mello, S., Jackson, T., Craig, S., Morgan, B., Chipman, P., White, H., Person, N., Kort, B., El Kaliouby, R., Picard, R.W., Graesser, A. (2008). AutoTutor detects and responds to learners affective and cognitive states, In *Workshop on emotional and cognitive issues at the international conference on intelligent tutoring systems* (pp. 306–308).

Du, J., Fan, X., Xu, J., Wang, C., Sun, L., Liu, F. (2019). Predictors for students' self-efficacy in online collaborative groupwork. *Educational Technology Research and Development*, *67*(4), 767–791.

Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, *4*(5/6), 304–317.

Flanagan, B., & Ogata, H. (2018). Learning analytics platform in higher education in Japan. *Knowledge Management and E-Learning*, *10*, 469–484.

Forsell, J., Forslund Frykedal, K., Hammar Chiriac, E. (2020). Group Work Assessment: Assessing Social Skills at Group Level. *Small Group Research*, *51*, 87–124.

Hirashima, T., Yamasaki, K., Fukuda, H., Funaoi, H. (2015). Framework of kit-build concept map for automatic diagnosis and its preliminary use. *Research and Practice in Technology Enhanced Learning*, *10*(1), 1–21.

Huang, Y., Zhu, M., Wang, J., Pathak, N., Shen, C., Keegan, B., Williams, D., Contractor, N. (2009). The formation of task-oriented groups: Exploring combat activities in online games, In *2009 International Conference on Computational Science and Engineering*, Vol. 4 (pp. 122–127). IEEE.

Kim, Y., D'Angelo, C., Cafaro, F., Ochoa, X., Espino, D., Kline, A., Hamilton, E., Lee, S., Butail, S., Liu, L., et al (2020). Multimodal data analytics for assessing collaborative interactions. In M. Gresalfi & I.S. Horn (Eds.), *Proceedings of the 14th International Conference on Learning Sciences*, Vol. 5 (pp. 2547–2554). International Society of the Learning Sciences.

Kudo, T. (2016). Tiny segmenter (in japanese). Available: http://chasen.org/~taku/software/TinySegmenter/.

Kyndt, E., Raes, E., Lismont, B., Timmers, F., Cascallar, E., Dochy, F. (2013). A meta-analysis of the effects of face-to-face cooperative learning. Do recent studies falsify or verify earlier findings? *Educational Research Review*, *10*, 133–149.

Lee Jensen, J., & Lawson, A. (2011). Effects of collaborative group composition and inquiry instruction on reasoning gains and achievement in undergraduate biology. *CBE Life Sciences Education*, *10*, 64–73.

Macfadyen, L.P., & Dawson, S. (2012). Numbers are not enough. Why e-learning analytics failed to inform an institutional strategic plan. *Journal of Educational Technology & Society*, *15*(3), 149–163.

Majumdar, R., Akçapınar, A., Akçapınar, G., Flanagan, B., Ogata, H. (2019). LAView: Learning analytics dashboard towards evidence-based education, In *9th International Conference on Learning Analytics and Knowledge* (pp. 386–387).

Majumdar, R., & Iyer, S. (2014). Using stratified attribute tracking (SAT) diagrams for learning analytics, In *IEEE 14th International Conference on Advanced Learning Technologies* (pp. 386–387).

Manske, S., Hecking, T., Chounta, I.A., Werneburg, S., Ulrich Hoppe, H. (2015). Using differences to make a difference: A study on heterogeneity of learning groups. In O. Lindwall, P. Häkkinen, T. Koschmann, P. Tchounikine, S. Ludvigsen (Eds.), *Exploring the material conditions of learning: the computer supported collaborative learning (CSCL) conference 2015*, Vol. 1 (pp. 182–189). Gothenburg: The International Society of the Learning Sciences.

Manske, S., & Hoppe, H.U. (2016). The "Concept cloud": Supporting collaborative knowledge construction based on semantic extraction from learner-generated artefacts, In *Proceedings - IEEE 16th International Conference on Advanced Learning Technologies* (pp. 302–306).

Manske, S., & Hoppe, H.U. (2017). Managing Knowledge Diversity: Towards Automatic Semantic Group Formation, In *Proceedings - IEEE 17th International Conference on Advanced Learning Technologies* (pp. 330–332).

Maqtary, N., Mohsen, A., Bechkoum, K. (2019). Group formation techniques in computer-supported collaborative learning: A systematic literature review. *Technology, Knowledge and Learning*, *24*(2), 169–190.

Milton, G.A. (1965). Enthusiasm vs effectiveness in group and individual problem-solving. *Psychological Reports*, *16*, 1197–1201.

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.

Nyikos, M., & Hashimoto, R. (1997). Constructivist theory applied to collaborative learning in teacher education: In search of ZPD. *The Modern Language Journal*, *81*(4), 506–51.

Ogata, H., Majumdar, R., Akçapinar, G., Hasnine, M.N., Flanagan, B. (2018). Beyond learning analytics: Framework for technology-enhanced evidence-based education and learning, In *26th International Conference on Computers in Education, Workshop Proceedings* (pp. 493–496).

Ounnas, A., Davis, H.C., Millard, D.E. (2007). Towards semantic group formation, In *Proceedings - The 7th IEEE International Conference on Advanced Learning Technologies* (pp. 825–827).

Pöysä-Tarhonen, J., Care, E., Awwal, N., Häkkinen, P. (2018). Pair interactions in online assessments of collaborative problem solving: case-based portraits. *Research and Practice in Technology Enhanced Learning*, *13*, 1–29.

Schneider, B., & Blikstein, P. (2015). Unraveling Students' Interaction Around a Tangible Interface using Multimodal Learning Analytics. *Journal of Educational Data Mining*, *7*(3), 89–116.

Shapiro, S.S., & Wilk, M.B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3/4), 591–611.

Siemens, G. (2012). Learning analytics: envisioning a research discipline and a domain of practice, In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 4–8).

Sivaloganathan, S., Al-Marzouqi, A., Zaneldin, E. (2020). Teaching conceptual design to a heterogeneous group: A workshop method. *2020 ASEE Virtual Annual Conference Content Access*.

Smith, J., Bratt, H., Richey, C., Bassiou, N., Alozie, N. (2016). Spoken interaction modeling for automatic assessment of collaborative learning, In *Speech Prosody* (pp. 277–281).

Splichal, J.M., Oshima, J., Oshima, R. (2018). Regulation of collaboration in project-based learning mediated by CSCL scripting reflection. *Computers & Education*, *125*, 132–145.

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

Stahl, G., Koschmann, T., Suthers, D.D. (2006). Computer-supported collaborative learning: An historical perspective [Electronic Version]. Retrieved 2007-06-07 from http://lilt.ics.hawaii.edu/lilt/papers/2006/CSCL_American_English.pdf.

Strijbos, J.W. (2011). Assessment of (computer-supported) collaborative learning. *IEEE Transactions on Learning Technologies*, *4*(1), 59–73.

Urhahne, D., Schanze, S., Bell, T., Mansfield, A., Holmes, J. (2010). Role of the teacher in computer-supported collaborative inquiry learning. *International Journal of Science Education*, *32*(2), 221–243.

van Leeuwen, A. (2015). Learning analytics to support teachers during synchronous CSCL: Balancing between overview and overload. *Journal of Learning Analytics*, *2*(2), 138–162.

Wang, Q. (2010). Using online shared workspaces to support group collaborative learning. *Computers & Education*, *55*(3), 1270–1276.

Wessner, M., & Pfister, H.R. (2001). Group formation in computer-supported collaborative learning, In *Proceedings of the 2001 International ACM SIGGROUP Conference on Supporting Group Work - GROUP '01* (pp. 24–31).

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

Zheng, Z., & Pinkwart, N. (2014). A discrete particle swarm optimization approach to compose heterogeneous learning groups, In *Proceedings - IEEE 14th International Conference on Advanced Learning Technologies* (pp. 49–51).

Publisher's Note