Iyer, S. et al. (Eds.) (2022). Proceedings of the 30th International Conference on Computers in Education.
Asia-Pacific Society for Computers in Education

Learning Log-Based Group Work Support: GLOBE Framework and System Implementations

Changhao LIANG[a*], Izumi HORIHOSHI[a], Rwitajit MAJUMDAR[a] & Hiroaki OGATA[a]
[a] *Kyoto University, Japan*
*liang.changhao.84c@st.kyoto-u.ac.jp

Abstract: Group work activities can promote the interpersonal skills of learners. To support the teachers in facilitating such activities, we suggested a learning analytics-enhanced technology framework, Group Learning Orchestration Based on Evidence (GLOBE) with data-driven approaches. We designed and implemented a group formation system using genetic algorithms to form groups using learning log data. Even if there is no existing data, we presented a paradigm of continuous data-driven support for the whole group learning process, incorporating the peer and teacher evaluation results as input to subsequent groupings. Further, utilizing accumulated group learning evidence in such an ecosystem, we aim to explore predictive group formation indicators which can lead to automatic group formation based on teachers' purpose in different contexts for desirable performance in subsequent group learning phases.

Keywords: CSCL, learning analytics, group formation, group work implementation, data-driven systems, predictive modeling

1. Research background and goals

Collaborative learning is progressively adapted in various pedagogical contexts where participants work together to share ideas, help each other or accomplish team goals (Dillenbourg, 1999), which benefits their soft skills development. Nowadays, computer-supported collaborative learning (CSCL) (Stahl et al., 2006) and learning analytics (LA) (Siemens, 2012) provides digital tools and data support, bringing immense opportunities to scaffold group work activities.

However, obstacles to providing valid CSCL support still exist. In terms of group formation, teachers tend to resort to random grouping or just pairing neighboring students owing to difficulties to do it in a real-time manner (Salihoun et al., 2017). Students from traditional classrooms seldom use digital tools, which leads to a cold start problem for the lack of enough learning logs to create learner models for group allocation (van der Velde et al., 2021). Even with CSCL support, teachers can get overwhelmed if they do not know how to use CSCL tools for orchestration. Currently, researchers focus on LA tools during the orchestration phase of the group work, while valid support for other phases in a data re-usage perspective deserves further attention as well.

In this research, we put forward the Group Learning Orchestration Based on Evidence (GLOBE) framework to support group work with data-driven systems. This thesis will talk about the data-driven system design and empirical implementations surrounding the phases of GLOBE with iterative data flow. We shall conduct several empirical studies to investigate the impact of LA-enhanced systems in different learning contexts, and in turn figure out predictors of successful group work from these inputs to orchestrate an ecosystem. The main research questions goes as follows.

Q1: How to design a group formation tool using multiple attributes of student model data?
Q2: How to support the whole group learning process with continuous group work data?
Q3: How to orchestrate an ecosystem for automatic group formation with predictive indicators?

2. Research framework

2.1 Group Learning Orchestration Based on Evidence (GLOBE)

Group Learning Orchestration Based on Evidence (GLOBE) provides a framework for group learning support with data-driven approaches in the learning analytics-enhanced environment (Liang et al., 2021a). As illustrated in Figure 1, the data-driven workflow covers four phases: group formation, orchestration of group work, evaluation of group work, and reflection after group work. Currently, the algorithmic group formation system and the peer evaluation system instantiate the GLOBE framework as two organic components of a Learning Analytics Dashboard (Majumdar et al. 2019).

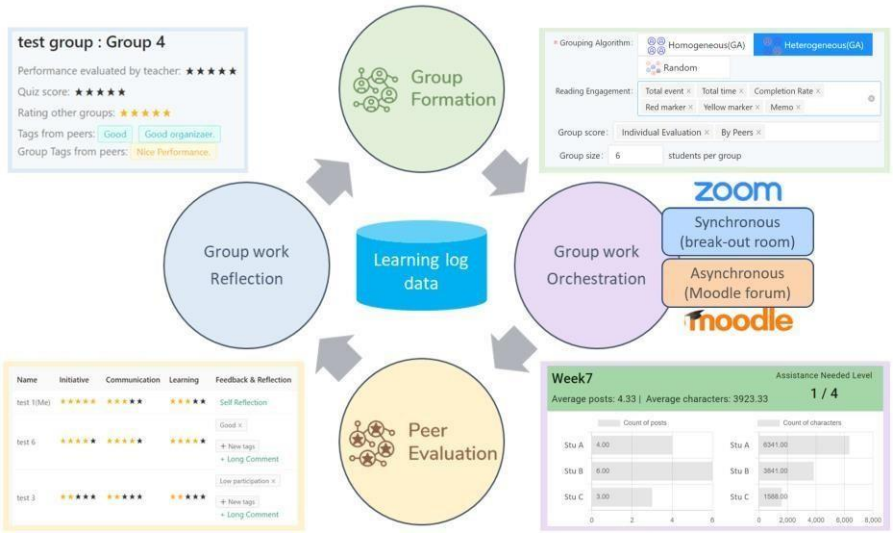

Figure 1. Systems used in different steps in the GLOBE framework (Liang et al., 2021a).

As for the group formation module, underpinned by the first version that formed groups based on the simple ranking of each characteristic, a genetic algorithm was applied to strengthen the flexibility to multiple data sources (Flanagan et al., 2021). For each student, there is a corresponding vector covering multiple characteristics of the student for the calculation of fitness value (see Figure 2). These characteristics come from user model variables (Boticki et al., 2019) such as online reading logs, quiz scores from the LMS, and previous rating data from the peer evaluation module. Employing the fitness value, we can determine homogeneous groups that have similar members, or heterogeneous groups that are made up of dissimilar group members.

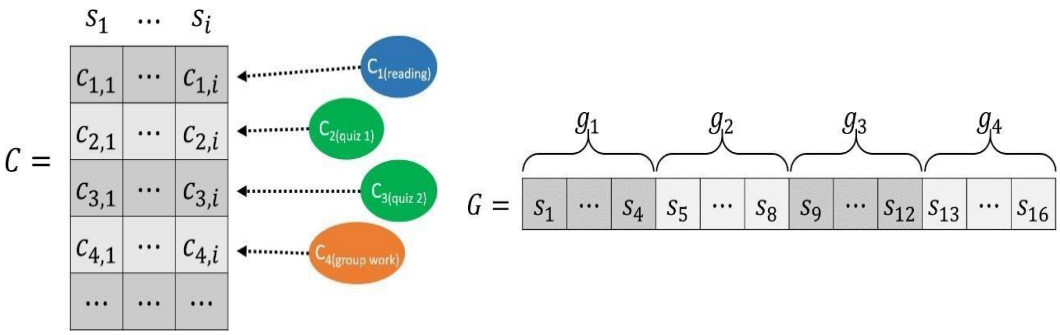

Figure 2. The mapping of student model variables to the characteristic representation matrix and group representation (Flanagan et al., 2021)

2.2 Continuous data-driven group work support

Figure 3 summarizes the continuous data-driven support throughout GLOBE that features the novelty of this study (Liang et al., 2022). This data flow provides an example of how to start with no existing learning logs in the student model initially and then incorporate the group work evaluations data cyclically for eventual group formation. Simple randomized grouping followed by using the evaluation score for subsequent grouping provides a feasible solution to the cold start problem in data-driven research (van der Velde et al., 2021). Such data iteration can also be used to identify students who may need special attention in the current group learning beforehand (Bukowski et al., 2017) in the detail panel of group formation results. Additionally, to determine the reliability of each evaluator's peer ratings, the student model attributes existing in the group formation system can be utilized as performance indicators according to Piech et al., (2013).

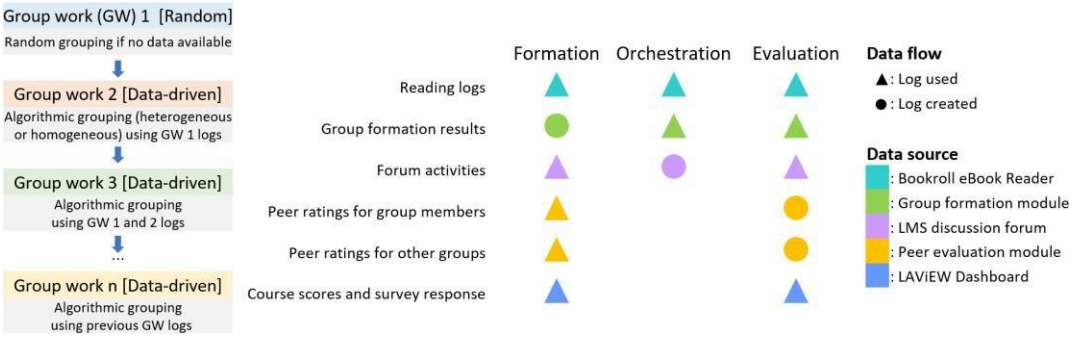

Figure 3. Continuous data feedforward of GLOBE (Liang et al., 2022).

3. Contributions to the field of CSCL and learning analytics research

This study provides a low threshold for the teacher to adapt the workflow thus promoting the use of a data-driven environment in actual class activities in several empirical studies. Previous implementations of GLOBE systems showed that it could reduce the time for teachers and students from trivial works of group formation and evaluation. Though we only used a few student model data in current studies, it disclosed further opportunities for the LA-enhanced group work orchestration following the continuous data flow, even in classroom-based contexts where no initial data existed. GLOBE implementation in other in-class learning contexts could be done with similar workflow, such as math problem-solving (Liang et al., 2021a) and English reading.

Though there have been studies discussing the impact of group compositions in actual group work (Janssen & Kirschner, 2020), this study contributes to digitizing this issue by introducing the heterogeneity value of each group which derives from the fitness value of genetic algorithm (Flanagan et al., 2021). We suggested a new perspective to explore how group composition on diverse student model data tends to make a difference in performance of subsequent phases. By investigating the specific student model variables for group formation, we can inspect the heterogeneity among which characteristics weigh more to affect the subsequent group work process and outcome.

4. Proposed research methodology

4.1 Algorithmic Group formation system in different learning contexts.

To address RQ1, we plan to conduct empirical studies on group formation using different student model data in various contexts. An implementation in Japanese primary school context was conducted. Engagement and affective states extracted by speech record were operationalized as group work indicators and suggest effectiveness of the group formation system (Liang et al., 2021a).

In subsequent studies, we aim to examine the flexibility of the group formation system in different scenarios to underscore its pedagogical implications. We shall fine-tune the system algorithm based on demand from different contexts. We are currently working on active reading in English learning which requires common reading markers as group formation criteria, and math learning that requires knowledge mastery models.

4.2 Continuous data support for the whole group learning process

As an organic component of GLOBE, group learning evidence can be used for group formation in subsequent rounds. We conducted a first trial of such a data feedforward approach across four sessions of in-class group discussion in a national language class in a Japanese middle school context where no student model data is available at first. The peer evaluation system (Liang et al. 2021b) was used to generate student models on group work experience for next-round group work. Peer ratings and group work perception surveys were adopted for evaluation of the group work performance.

The data-driven flow can not only benefit the group formation phase. For the evaluation phase, previous learning evidence can be used to determine the reliability of each evaluator. This will lead to another topic that focuses on the cultivation of peer evaluation capabilities across several rounds.

4.3 Ecosystem for automatic group formation with predictive indicators

Using the data of GLOBE implementation in different contexts, we can further explore which indicators can predict and determine the weight of each student model variable in group formation for different contexts (Janssen & Kirschner, 2020), and form an ecosystem to automatically recommend these indicators for teachers. Currently, we adopted factor analysis and correlation analysis to discover the relationships of indicators in all phases of GLOBE and highlighted key issues for successful group work in specific contexts utilizing current implementations of GLOBE systems.

In the next stage, as a closure of the loop between AI support and practice, we plan to transform parameter-based group formation to context-based group formation where groups can be automatically created using recommended parameter selection in line with learning context assigned by the teacher.

Acknowledgements

This research was supported by the following grants: JSPS KAKENHI 20K20131, 20H01722, 22H03902, NEDO JPNP20006, JPNP18013, and JST SPRING JPMJSP2110.

References

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*(5-6), 754-765.

Bukowski, W. M., Castellanos, M., & Persram, R. J. (2017). The current status of peer assessment techniques and sociometric methods. *New directions for child and adolescent development*, *2017*(157), 75-82.

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

Flanagan, B., Liang, C., Majumdar, R., and Ogata, H. (2021). Towards explainable group formation by knowledge map-based genetic algorithm. In *2021 International Conference on Advanced Learning Technologies (ICALT)* (pp.370–372).

Janssen, J., & Kirschner, P. A. (2020). Applying collaborative cognitive load theory to computer-supported collaborative learning: towards a research agenda. *Educational Technology Research and Development*, *68*(2), 783-805.

Liang, C., Majumdar, R. & Ogata, H. (2021a). Learning log-based automatic group formation: system design and classroom implementation study. *Research and Practice in Technology Enhanced Learning*, *16*(1), 1–22.

Liang, C., Toyokawa, Y., Nakanishi, T., Majumdar, R., & Ogata, H. (2021b, August). Supporting Peer Evaluation in a Data-Driven Group Learning Environment. In *International Conference on Collaboration Technologies and Social Computing* (pp. 93-100). Springer, Cham.

Liang, C., Majumdar, R., & Ogata, H. (2022, June). Continuous Data-Driven Group Learning Support: Case Study of an Asynchronous Online Course, In *Proceedings of the 15th International Conference on Computer- Supported Collaborative Learning - CSCL 2022* (pp. 547-548).

Majumdar, R., Akçapınar, A., Akçapınar, G., Flanagan, B., & Ogata, H. (2019). Learning analytics dashboard widgets to author teaching-learning cases for evidence-based education. In *Companion Proceedings of the 9th International Conference on Learning Analytics and Knowledge* (pp.600-607).

Piech, C., Huang, J., Chen, Z., Do, C., Ng, A., Koller, D.: Tuned models of peer assessment in moocs. arXiv preprint arXiv:1307.2579 (2013)

Salihoun, M., Guerouate, F., Berbiche, N., & Sbihi, M. (2017). How to Assist Tutors to Rebuild Groups Within an ITS by Exploiting Traces. Case of a Closed Forum. *International Journal of Emerging Technologies in Learning*, *12*(3).

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).

Stahl, G., Koschmann, T., & Suthers, D. (2006). Computer-supported collaborative learning: An historical perspective. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 409–426). van der

Velde, M., Sense, F., Borst, J., & van Rijn, H. (2021). Alleviating the cold start problem in adaptive learning using data-driven difficulty estimates. *Computational Brain & Behavior*, *4*(2), 231-249.