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

Exploring Predictive Indicators of Reading-Based Online Group Work for Group Formation Teaching Assistance

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

Abstract: Using digital systems to group students according to their indicators provides opportunities for better group work implementation. However, how these indicators can affect group work performance remains unclear. Teachers tend to feel confused about which indicators should be considered when creating groups using learning log data. Capitalized on the data-driven environment under GLOBE, we conducted a preliminary study to explore predictive indicators for algorithmic group formation in a reading-based group learning context. This study presented our effort to explore the key factors that correlated to a desirable group work via factor analysis and correlation analysis. We found that reading engagement and previous peer rating scores suggest a higher potential to predict desirable group work performance in the reading-based online group work, which aims to help teachers set appropriately in future student model data-based group formation.

Keywords: data-driven group formation, reading-based group work, CSCL, factor analysis, correlation analysis

1. Introduction

Group learning is progressively adapted in pedagogical practice and prevalent online courses nowadays raise impetus to the demand for such interactive activities. To initial a group work, teachers should align students appropriately according to different learning contexts (Urhahne et al., 2010). With the scaffold of Computer-supported collaborative learning (CSCL) (Stahl et al., 2006) and learning analytics (LA) (Siemens, 2012), tools and enriched data became available to support group work implementation such as algorithmic group formation systems (Liang et al., 2019).

Meanwhile, teachers face difficulties when using digital systems (Austin et al., 2010). As the group formation system provides multiple input variables, teachers may feel confused about what each variable conveys and get overwhelmed by the settings. To enable teachers to smoothly form desirable groups, we conducted this study based on a data-driven environment. Using the log data from reading-based online group work, we made a preliminary analysis to estimate the potential of existing indicators to predict group work performance. This study aims to assist teachers by achieving the following goals:

- Present more explainable group formation input variables for teachers.
- Enrich experience by parameter-setting recommendations for similar group learning contexts.

2. Research background

There exist multiple issues that deserve nuanced consideration when executing collaborative learning activities (Urhahne et al., 2010). Janssen & Kirschner (2020) categorized these indicators into three attributes: antecedent, process and consequence. Since indicators of antecedents can pose an effect on processes and consequences of collaboration (see Figure 1), to deal with antecedent attributes, studies on group formation techniques using these attributes are highlighted in the CSCL field to help teachers set groups appropriately before the group work (Maqtary et al., 2019).

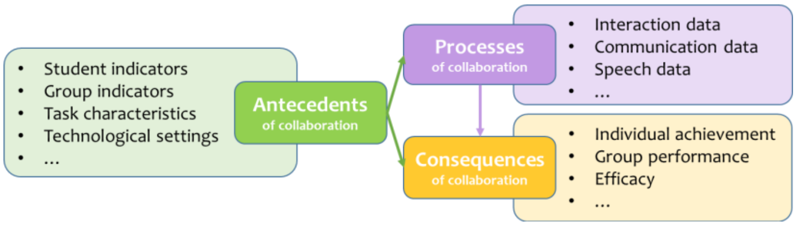

Figure 1. Indicators among Collaborative Process Attributes (Janssen & Kirschner, 2020).

As for antecedent, Janssen & Kirschner (2020) listed several indicators based on what it describes. Student and group indicators are frequently-used (Saqr et al., 2020) but vary from group work tasks. Student indicators can include intellectual abilities, experiences, preferences and personalities (Sánchez et al., 2021). Group indicators depict higher-level aspects such as social relationships and heterogeneity of certain student-level indicators within a group, hence Cress (2008) underscored the necessity to simultaneously take into account student indicators as well. In terms of process attributes, behavior logs such as students' interaction logs (Saqr et al., 2020) as well as utterance (Liang et al., 2021) are highlighted to reflect an overall look of the group work engagement. In terms of group work implementation, process data such as students' interaction logs are highlighted to reflect an overall look of the group work engagement (Saqr et al., 2020). Besides, when it comes to consequences, individual achievement, perceived efficacy and group work performance are used to evaluate the consequence of group work (Janssen & Kirschner, 2020).

Since antecedents tend to affect processes and consequences, researchers have presented the impact of some antecedent indicators on the process and consequence of group work in different scenarios. For example, the previous knowledge and task experience proved to be closely related to the group work performance in the collaborative searching tasks (Jiang et al., 2022). The collaborative tendency can also affect group work outcomes (Zhang et al., 2022). Besides student indicators, groups with different sizes, intimacy levels (Block, 2015) and compositions (Sánchez et al., 2021) make a difference in the consequence. Though former researchers inspected the effect of specific indicators under controlled experiments, less study comprehensively looks at these modeling antecedents, and which input plays a more important role remains unclear.

3. Group work indicators in a data-driven environment

Group Learning Orchestration Based on Evidence (GLOBE) provides a data-driven environment (Liang et al., 2021) to study the indicators of group work mentioned in the previous section. As is illustrated in Figure 2, the data-driven support via GLOBE systems covers four phases: group formation, orchestration, evaluation, and reflection.

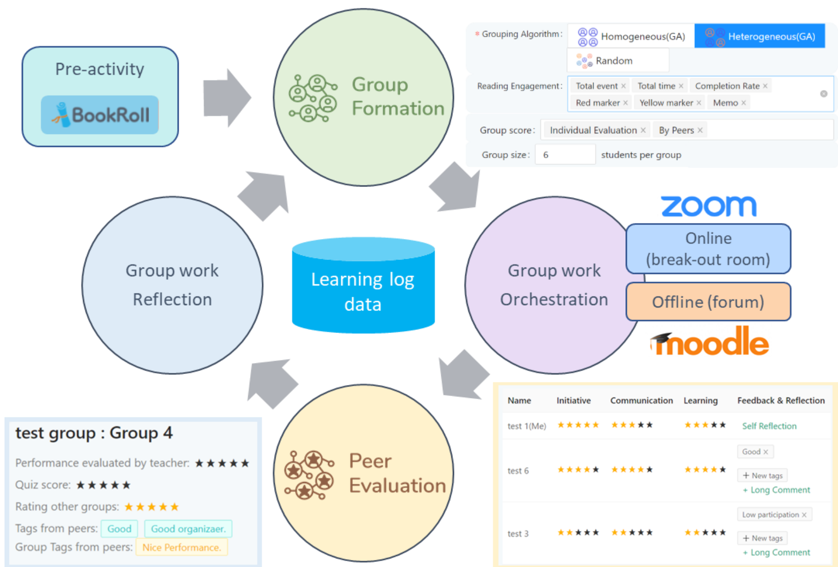

Figure 2. Systems used in different steps in the GLOBE framework (Liang et al., 2021).

3.1 *Conducting group work using GLOBE*

To conduct group work under GLOBE, teachers can select any number of indicators to create either homogeneous or heterogeneous groups via an algorithmic group formation system (Flanagan et al., 2021). These student indicators come from learning log data from the reading attributes computed from BookRoll logs (Flanagan & Ogata, 2018), and external sources such as test scores uploaded to the LAView dashboard (Majumdar et al. 2019). During the group work, the teacher can overview the group work progress and get prompts of intervention for undesirable groups with the help of orchestration data analyzers and dashboards. After the group work, peer and teacher evaluations with ratings and feedback are provided in the evaluation modules.

Throughout the group formation, orchestration and evaluation phases, GLOBE covers multiple indicators presenting antecedent, process and consequences attributes. Such group work-related data are used and generated within different phases, hence ensuring a systematic and continuous usage of data to buttress subsequent group work with the same students.

3.2 *Modeling for key input indicators for prediction*

Table 1 shows the student indicators available in GLOBE systems with their corresponding Collaborative Process Attribute (Janssen & Kirschner, 2020). These indicators convey certain student indicators or skills mentioned in recent studies. Given the set of indicators, while using the system some of the previous indicators can be used as antecedent indicators in subsequent group formation, which shows an advantage of the system that maintains the history of learners' group work activities. However, feature selection remains a problem for teachers, especially those who are not familiar with CSCL tools. We start from the analysis of student indicators that describe individual-level characteristics. A two-step analysis is presented to answer the following questions:

- Q1: What are the principal factors that can prune the antecedent indicators?
- Q2: What are relationships among indicators in different Collaborative Process Attributes?

Table 1. *Student indicators available in GLOBE systems*

Collaborative Process Attribute	Indicators within GLOBE framework	Data source	Proxy for construct (what does it convey)
Antecedents of collaboration	reading time	BookRoll	Academic performance (Juncoet al., 2015)
	operation times	BookRoll	
	completion rate	BookRoll	
	red marker (highlight)	BookRoll	Active reading skills (Toyokawa et al., 2021)
	yellow marker (difficult)	BookRoll	
	*memo	BookRoll	
	*quiz score	LAView	Prior knowledge
Processes of collaboration	forum view	Moodle forum	Engagement and active interactions (Fidalgo-Blanco et al., 2015)
	forum post	Moodle forum	
Consequences of collaboration	*teacher's rating	Peer evaluation	Group work experience and task experience (Janssen & Kirschner, 2020)
	peer rating (individual)	Peer evaluation	
	peer rating (group)	Peer evaluation	
	final course grade	LAView	Academic performance

* not used in this study

4. Dataset and analysis results

The data of a university course in which 30 students were enrolled is selected for this study. Reading-based online group work was conducted several times across the 15-week semester, where students were required to read and summarize research articles as after-class group work and present their reading outcomes during the online lecture. We used the data from the last group work in week 11 for analysis, since it contains abundant data and students had got familiar with the process. Five groups with six members were formed by a heterogeneous algorithm with checked antecedent indicators in Table 1. Two students were excluded from the dataset due to their absence.

4.1 *Factor analysis for input parameters*

To explore the underlying factors from the antecedent indicators, principal component factor analysis with varimax rotation was implemented. The Kaiser–Meyer–Olkin Measure of Sampling Adequacy (MSA) value for the correlation matrix of the dataset is 0.64, showing medium appropriation for factor analysis (Ferguson & Cox, 1993). The F value of Bartlett's Test is significant as well.

As a result, three factors were identified with initial eigenvalues greater than 1, which can account for 81.85 percent of the indicators. Table 2 shows the percentage variance explained by the antecedent indicators and Table 3 presents the rotated component matrix of factors. The communalities of the input variable range from 0.58 to 0.95, showing that all of the indicators can be expressed by these factors. A cutoff for statistical significance of the factor loadings of 0.5 was used (Muilenburg & Berge, 2005). Each item loaded distinctively on one factor; the highest factor loading was separated from its next nearest loading by at least 0.4.

Table 2. *Total variance explained by factors*

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
Factor 1: engagement	3.364	48.056	48.056
Factor 2: active reading tendency	1.339	19.134	67.191
Factor 3: previous peer ratings for group	1.027	14.667	81.858

Table 3. *Rotated component matrix of factors*

	Component		
	Factor 1: engagement	Factor 2: active reading tendency	Factor 3: previous peer ratings for group
completion rate	0.908		
operation times	0.907		
reading time	0.807		
previous peer rating (individual)	0.693		
red marker		0.909	
yellow marker		0.888	
previous peer rating (group)			0.973

We named factor 1 as “engagement” since it covers the reading engagement indicators. The previous peer ratings scores are loaded into this factor as well. We can infer that the peer rating of online group work is partially based on the engagement which is the most obvious to the groupmates, hence members with high reading engagement tend to show high group work engagement as well. Factor 2 was named as “active reading tendency” according to Toyokawa et al. (2021), and factor 3 as “previous group work output”. These factors provide a more understandable integration of raw input variables from the group formation system, hence also throw light on the optimization of the system.

4.2 *Correlational analysis for predictive antecedents*

Besides dimension reduction, we explored predictive antecedents using correlational analysis to find predictive indicators for novice teachers to choose from in a similar context. Pearson correlational examination was calculated among indicators of three phases and Figure 4 presents the result.

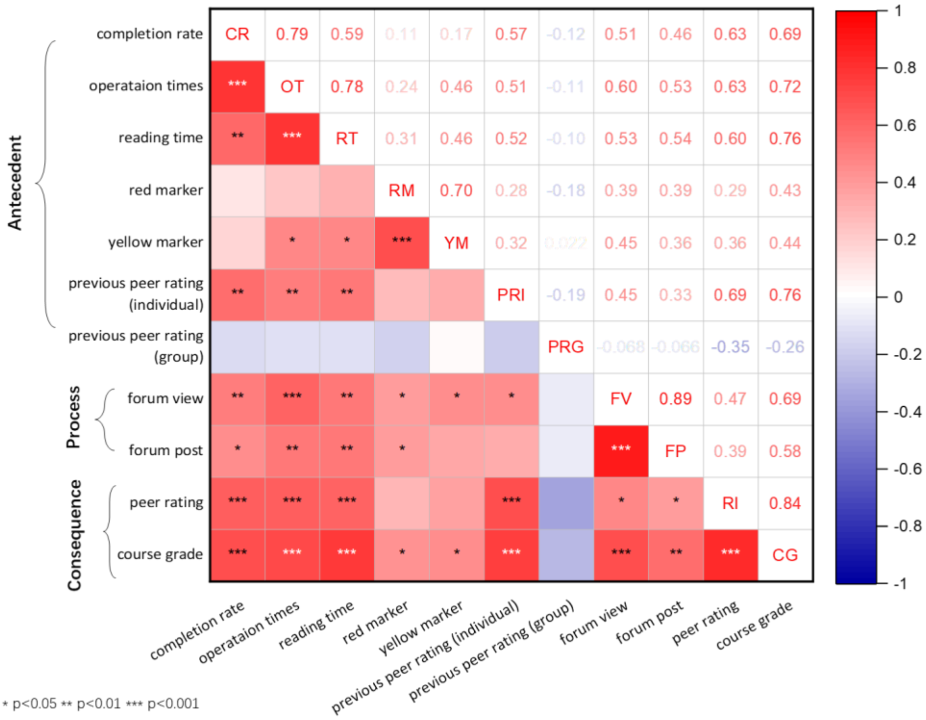

Figure 4. Results of correlation analysis of all indicators in group work

Reading engagement indicators (completion rate, operation times and reading time) suggest significant positive pertinence to all process and consequence attributes. Students with high engagement in reading tasks are inclined to not only get better academic achievements (Juncoet al., 2015) but also take an active part in group discussion and get positive peer feedback. Meanwhile, previous peer rating scores for group members show similar relevance to the reading engagement indicators with high predictive potential, which proves the impact of cyclically accumulated group work data under GLOBE (Liang et al., 2021). The results also agree with the idea that prior knowledge and task experience can influence the group work consequence (Janssen & Kirschner, 2020). Besides, annotation behaviors are positively related to the forum views as well as final course grades, and red markers suggest positive relation to the forum post behaviors which indicates active participation (Fidalgo-Blanco et al., 2015). Such finding indicates a promising connection between acting reading behaviors (Toyokawa et al., 2021) and its corresponding forum-based group work participation. This result is also meaningful to inform and assist the ways teachers create groupings for future reading-based group work.

For peer ratings of previous group work output, there appears no significant relation to any other indicators, indicating a low predictive weight. As this indicator derives from a group-level rating, social loafing and free-riding could decrease its reliability, it should be less considered in following group creations for teachers.

5. Conclusions and Future Work

This study puts forward a foreground to modeling from log-based antecedent indicators to predict group work processes and consequences. Though we focused on reading-based group work context, we disclosed the opportunity for predictive analysis in different tasks under the GLOBE environment.

The indicators listed herein are still limited in the current stage and cannot fully depict student characteristics. We only examined student indicators in a specific group work context, while group-level indicators such as intimacy and heterogeneity were not covered. Since the dataset only covers one group work of the course, the results need validation within a larger sample in subsequent works.

In the next step, more process indicators such as forum interaction logs will be re-used as antecedents, and group-level indicators like heterogeneity will be examined. Capitalized on the integrated data-rich environment, predictive modeling such as regression would be done with a larger sample so that the specific weight of each antecedent indicator can be estimated. This preliminary analysis attempted to make parameter selection more understandable to teachers and presented parameter-setting recommendations for similar learning contexts, and will lead to system improvement as future work to lower the threshold for teachers.

Acknowledgements

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

References

Austin, R., Smyth, J., Rickard, A., Quirk-Bolt, N., & Metcalfe, N. (2010). Collaborative digital learning in schools: Teacher perceptions of purpose and effectiveness. *Technology, Pedagogy and Education*, *19*(3), 327-343.

Block, P. (2015). Reciprocity, transitivity, and the mysterious three-cycle. *Social Networks*, *40*, 163-173.

Cress, U.(2008).The need for considering multilevel analysis in CSCL research: An appeal for the use of more advanced statistical methods.*International Journal of Computer-Supported Collaborative Learning*,*3*,69-84.

Ferguson, E., & Cox, T. (1993). Exploratory factor analysis: A users' guide. *International journal of selection and assessment*, *1*(2), 84-94.

Fidalgo-Blanco, Á., Sein-Echaluce, M. L., García-Peñalvo, F. J., & Conde, M. Á. (2015). Using Learning Analytics to improve teamwork assessment. *Computers in Human Behavior*, *47*, 149-156.

Flanagan, B., & Ogata, H. (2018). Learning analytics platform in higher education in Japan. Knowledge Management & E-Learning: An International Journal, 10(4), 469-484.

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.

Jiang, X., Li, Y., Zhang, Y., Z. & Liu, C. (2022). Research on the Effect of Prior Knowledge Differences on Collaborative Information Search and Learning. *Library and Information Service*, *66*(8), 32-41.

Junco, R., & Clem, C. (2015). Predicting course outcomes with digital textbook usage data. *The Internet and Higher Education*, *27*, 54-63.

Liang, C., Boticki, I., & Ogata, H. (2019). Supporting teachers in group work formation and analytics for in-class group activities. In *the 27th International Conference on Computers in Education (ICCE 2019)*(pp.744-749).

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

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).

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.

Muilenburg, L. Y., & Berge, Z. L. (2005). Student barriers to online learning: A factor analytic study. *Distance education*, *26*(1), 29-48.

Sánchez, O. R., Ordóñez, C. A. C., Duque, M. Á. R., & Pinto, I. I. B. S. (2021). Homogeneous group formation in collaborative learning scenarios: An approach based on personality traits and genetic algorithms. *IEEE Transactions on Learning Technologies*, *14*(4), 486-499.

Saqr, M., Nouri, J., Vartiainen, H., & Malmberg, J. (2020). What makes an online problem-based group successful? A learning analytics study using social network analysis. *BMC medical education*, *20*(1), 1-11.

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).

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.

Toyokawa, Y., Majumdar, R., Lecailliez, L., & Ogata, H. (2021). A flipped model of active reading using a learning analytics-enhanced e-book platform. In *the 29th International Conference on Computers in Education (ICCE 2021)* (pp. 510-515).

Zhang, P., Wang, D. & Tang, Z.(2022).Research on the Online Knowledge Collaboration Behaviors: the Influence of Team Member Intimacy and Collaborative Tendencies. *Library and Information Service*, *66*(8), 21-31.