Using Differences to Make a Difference: A Study on Heterogeneity of Learning Groups

Sven Manske, University Duisburg-Essen, manske@collide.info
Tobias Hecking, University Duisburg-Essen, hecking@collide.info
Irene-Angelica Chounta, University Duisburg-Essen, chounta@collide.info
Sören Werneburg, Otto-Hahn-Gymnasium Dinslaken, soeren.werneburg@gmail.com
H. Ulrich Hoppe, University Duisburg-Essen, hoppe@collide.info

Abstract: There is an increasing interest in student-centered teaching methods with small group learning as an important ingredient. In this paper, we present a study in which the performance of heterogeneous and homogeneous learning groups has been compared in a technology-enhanced classroom setting in the area of STEM learning. The group formation was based on learning analytics results that were considered in a semi-automatic formation process. The analytic methods used incorporated different artefact-related characteristics, but also motivational features as input. We observed that the heterogeneous groups outperformed the homogeneous ones in different ways. The results of the study are analysed using quantitative and qualitative approaches on both the individual and the group level.

Keywords: group learning, heterogeneity, inquiry-based learning, online experimentation

Introduction

There is an ongoing debate about the future of Europe being threatened by a downward trend in science education due to a lack of attraction for young people to become researchers or to work in fields of science and technology (Rocard, et al., 2007; Gago, Ziman, Caro, Constantinou, & Davies, 2004). Many initiatives point out that there is a need for more scientists and thus for fostering students' interest in science (Global Science Forum, 2006). Large scale assessments, such as PISA, show that student-centered and supportive teaching methods are needed. These aim to convey key competencies to learners by activating them, which opens new opportunities in contrast to teacher-centered approaches (Hannafin, Hill, Land, & Lee, 2014). This has been mainly underpinned by a change from teacher-centric and deductive to inquiry-based methods. Initiatives like the Go-Lab project enforce the idea of inquiry-based science education with online experimentation. Students dive into the role of researchers to investigate the big ideas of science. They receive guidance in the different steps of inquiry, e.g., through explicit phase names, scaffolds and other guidance mechanisms. Typical scenarios in Go-Lab consist of different tools and learning resources, aligned to an experiment with the Go-Lab portal serving as a general infrastructure and access point to a variety of online labs. As such, Go-Lab does neither directly support student collaboration nor classroom orchestration on the part of teachers. While collaborative online tools, especially in the context of virtual and remote experimentation, are not generally available, blended learning scenarios offer the opportunity to induce off-line collaboration in the classroom. Thus, teachers are required to set up group learning scenarios and take care of the classroom and group orchestration (Dimitriadis, Prieto, & Asensio-Pérez, 2013; Roschelle, Dimitriadis, & Hoppe, 2013; Dillenbourg, 2013).

To support this kind of classroom orchestration, we employ methods from learning analytics to create student groups with specific performance characteristics. To that end, we propose a multidimensional, clustering schema that takes into consideration students' motivation towards science and students' activity style with respect to artefact-based measures in a computer-supported learning environment. With the use of the aforementioned schema, we created heterogeneous and homogeneous groups that carried out a learning activity through the Go-Lab portal. Our main objective was to compare the performance of those groups and explore whether group heterogeneity had an effect on the overall learning outcome. Usually it is expected that weak learners who are members of heterogeneous groups or medium ability learners forming homogeneous teams achieve the maximum knowledge gain (Lou, Abrami, & d'Apollonia, 2001). We argue that heterogeneous groups coordinate and create common ground faster and easier than the homogeneous groups since the different characteristics of the individuals might complement each other. This is also depicted on the quality of the activity's outcome and the knowledge gain of students.

We evaluated this hypothesis on heterogeneity in a real classroom setting where the students had to carry out an inquiry-based learning-task. In order to evaluate the knowledge gain, the students took pre- and post-knowledge tests and the outcome of the activity (concept maps, short descriptive texts and group reports) was assessed by the teacher. We used the log files of the Go-Lab portal to analyze the activity of students. In addition, two experts were asked to observe the activity and keep notes in the form of activity transcripts in order to assess the collaborative practices of groups.

In the next sections we present a short overview of the state of the art, the experimental setting of the study and our conceptual framework. We present the analysis and evaluation of the results, a discussion on the finding of the study and we conclude with the outcome and future work.

Background and related work

Group formation is a key aspect of CSCL because it can affect the way people work together towards a common goal and eventually the learning outcome itself. Collaborative activities are expected to promote learning through common knowledge building and the social interaction among users (Stahl, Koschmann, & Suthers, 2006). However, collaboration alone does not ensure knowledge gain or successful practice (Jermann, Soller, & Muehlenbrock, 2001). Usually the task of group formation is carried out by the teacher who uses his experience on pre-defined criteria that may refer to students' social skills, gender, motivation or knowledge background (Ounnas, Davis, & Millard, 2009). This complicated process requires time and does not always lead to success.

Based on the availability of student performance data in computerized learning environments, (semi-) automatic or algorithmic approaches to group formation have been suggested. E.g., Balmaceda, Schiaffino, & Pace (2014), define group formation as a weighted constraint satisfaction problem (WCSP) depending on the characteristics of students such as personality traits, team roles, and social relationships. Also network analysis techniques have been employed for analyzing the interaction of users through a learning platform and clustering students based on their similarity (Sadeghi & Kardan, 2014). As one of the most sophisticated technical solutions so far, the GroupAL algorithm (Konert, Burlak, & Steinmetz, 2014) allows for optimizing group composition according to a variety of features, with the option of choosing between homogeneity and heterogeneity for each of these features.

The role of group homogeneity in collaborative classroom activities has been a subject of various studies. There are indications that heterogeneity of knowledge is beneficial for group performance (Webb, Nemer, & Zuniga, 2002; Kizilcec, 2013). However a certain baseline of background knowledge appears to be required for the collaboration to be beneficial (Gijlers & De Jong, 2005). In our own prior work we had also seen positive effects of diversity on the performance of learning groups (Chounta, Giemza, & Hoppe, 2014).

Experimental setting

This section covers the experimental setting of the study. First, we describe the Go-Lab platform which will be explained in line to the implemented scenario. Apart from the technical system, we explain the didactical goals and the production of learning objects by the students during this scenario. These artefacts are used for the assignment of groups and the assessment of performance characteristics.

The learning activity was split up into two phases: the first phase consists of individual student work for the initial assessment of performance characteristics. The tasks that the students had to carry out involved writing a short text to describe a simulation, and creating a concept map from different learning resources. A motivational questionnaire captured their interest and motivation in science. These characteristics served as an input for the group formation, which was used in the second phase of the study. In the second phase, the students performed an inquiry-based learning task in groups. The task objective was the online experimentation with a virtual lab of an osmotic power plant.

The outcome of the second phase was a concept map and a written report per group. The concept map should describe the parameter model of the power plant, and for the report the students had to formulate a short summary about their findings. This should include a critical reflection about the usefulness of osmotic power under different aspects, e.g. sustainability, effectiveness and dependence of the location. Four explicit assignments guided the students through the scenario and provided a scaffold for the report. However, no formal structure was given in order to promote an open-ended range of possible solutions.

Go-Lab learning environment

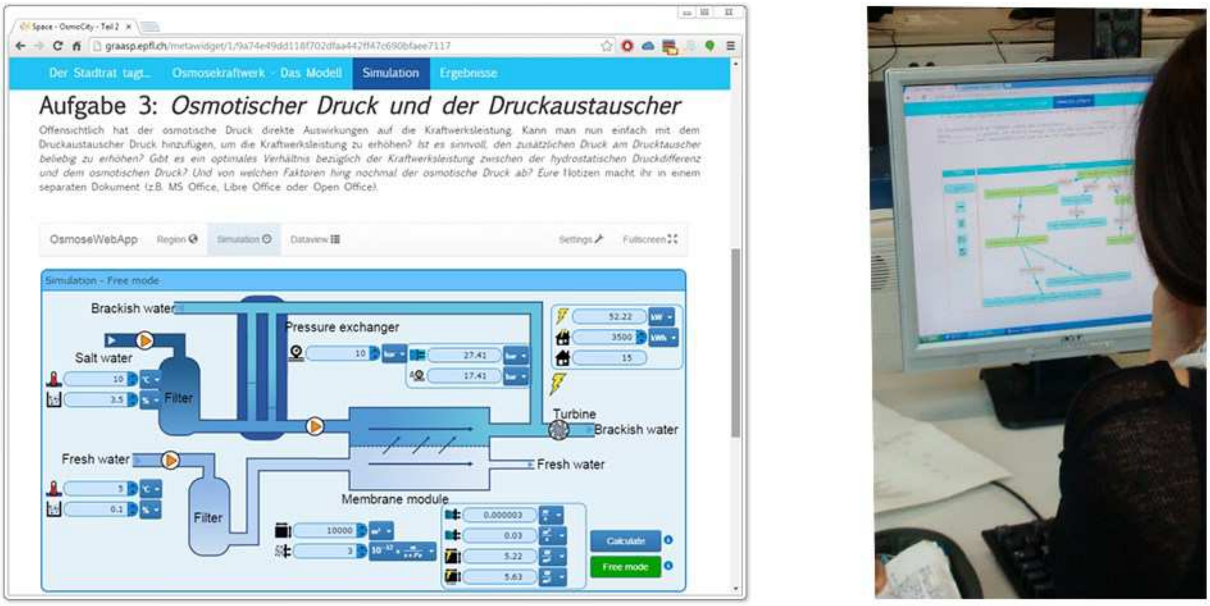
The Go-Lab portal is a web based learning environment for the authoring of inquiry-based learning scenarios and their implementation in classes (Govaerts, et al., 2013). The Go-Lab system follows an innovative approach to inquiry learning by providing a general infrastructure and acts as an access point to online labs. It aggregates learning resources and scaffolds, and provides guidance to learners (de Jong, Sotiriou, & Gillet, 2014). Figure 1 shows an example Inquiry Learning Space (ILS) of the environment, which has been used for this study, from the student perspective. Such learning activities consist of different inquiry phases, which are displayed as "tabs" in the navigation bar of the web environment and thus define a guided path through the inquiry process.

Learning goals

The main goal of our learning scenario is to understand the mechanism of osmotic power and how the location of an osmotic power plant influences the power generation. The learning scenario demands multidisciplinarity from the students in a way that knowledge from different subject domains such as biology, chemistry and physics is used. Also competencies from different fields such as text writing, metacognitive skills, concept mapping and inquiry skills are released during this experiment.

Critical thinking skills are demanded in the second phase of the study, where the students perform the group work task. At the beginning of the group phases, they get confronted with the "aggregated concept map" of all students (Manske, et al., 2014), which can be seen as a union of all concept maps represented as graphs. Such a structure contains useful and useless concepts and possibly wrong connections. This enforces a critical group discussion about the correctness of specific parts. In the following, students take this knowledge to create a new concept map capturing the parameter model of an osmotic power plant, while they are also confronted with some ecological factors of osmotic power and sustainability. Explicit assignments guide them through this scenario although they have to structure a final report by themselves.

Such a complex and multidisciplinary scenario, which incorporates different skills and competencies, possibly lead to a big diversity of the results. The students provide a non-standardized report as a final result, which does not allow for a simple and automated assessment. However, the benefits are in the qualitative evaluation of the reports and the group observations, which shows that it is possible to track different competencies and to have a detailed view on the students' performances.



(a) (b)

Figure 1. The web based learning environment with the interactive osmotic power plant simulation (a), and a student using the concept mapper in this environment during the study (b).

Conceptual framework

This section covers the conceptual framework for the group formation. First, it describes the process chain for the conceptual model of the group formation, which has been applied in our experimental setting. A key aspect is the composition of a feature set to describe different performance related characteristics of learning objects and motivation. Therefore, we provide an overview about all used measurements and their backgrounds.

Group formation processing chain

We define a heterogeneous learning group as a learning group, where each member has different performance characteristics. The learners produce artifacts during an inquiry-based learning scenario as described in the experimental setting. The artefacts, particularly learning objects and the assessment of motivational scores form the dataset for the group assignment. These characteristics span a feature space, while the vector containing the scores for a single student is called feature vector, which is an element of the feature space. To use simple Euclidean distance measurements in such a vector space, the feature vectors are normalized.

In total, we capture the performance characteristics through six artefact-related and three motivational scores, leading to a nine-dimensional feature space. In terms of classroom size, the dimension is too high to produce meaningful clusters. To tackle this curse of dimensionality, we perform a feature selection to minimize the dimension to a plausible number derived from the number of students and groups.

Performance characteristics and indicators

To decide whether a group is heterogeneous or homogeneous, we capture different performance characteristics which serve as a basis for the group formation. These incorporate not only artefact-based assessment on concept maps and small texts but also motivational assessments based on the SMTSL questionnaire (Tuan, Chin, & Shieh, 2005). The following section lines out different measurements of these performance characteristics.

Concept Maps

Concept mapping (Novak, 1984) is a technique for the externalization of knowledge structures in form of semantic networks. A learner creates a concept map by connecting concepts that are considered important for a given domain by labeled relations. Since concept maps reflect the structure of domain knowledge of individual learners, these artefacts are particularly suitable as a factor additionally to knowledge tests for characterizing students (Stoddart, Abrams, Gasper, & Canaday, 2000). In order to use the concept maps of students as parameters for group formation, quantification is needed. One approach is to compare a concept map to a reference map created by a tutor or expert (Conlon, 2004; McClure, Sonak, & Suen, 1999). This requires a matching of concepts between both maps comparing labels. This can be done automatically using computer linguistic methods (Conlon, 2004; Hoppe, Engler, & Weinbrenner, 2012). However, this is not trivial and can lead to wrong matching. Since the aim of this study is to measure the impact of group formation on student performance as accurately as possible, we decided to do the matching manually in order to avoid biases introduced by automatisms. Each concept map cm with a set of concepts N_S and a set of relations E_S is compared to an expert map with the concepts N_E and relations E_E. Five different measures were calculated:

- Node precision: Node precision measures the fraction of concepts in a student concept map that can be matched to concepts in the expert map, $np(cm) = \frac{|N_S| \cap |N_E|}{|N_S|}$.
- Node recall: This measure indicates to what extent the concepts in the expert map are covered by the student map, $nr(cm) = \frac{|N_S| \cap |N_E|}{|N_E|}$.
- Edge precision: The fraction of concept connections in the student concept map that can be also found in the expert concept map, $ep(cm) = \frac{|E_S| \cap |E_E|}{|E_S|}$.
- Edge recall: Edge recall is defined as the fraction of edges in the expert concept map that can be found in the student concept map, $er(cm) = \frac{|E_S| \cap |E_E|}{|E_E|}$.
- HEW-measure: Hoppe, Engler & Weinbrenner (2012) introduced a quality indicator for concept maps based on the comparison of a concept map to a given ontology. The measure was obtained based on empirical observations of structural properties that correlate with expert quality judgments.

$$hew(cm) = \frac{|N_S|}{1 + 3|N_S \cap N_E|} + \frac{7|E_S|}{1 + 6|E_S \cap E_E|} + \frac{|N_S||E_S \cap E_E|}{1 + 6|E_S||N_S \cap N_E|}.$$

We are aware that using these measures alone for a reliable assessment of the students' actual domain knowledge is limited. However, the concept map measures contribute to the creation of heterogeneous and homogeneous student groups by providing additional discriminating factors. In this sense the measures do not necessarily answer the question which students produce better concept maps but they give insights in which students produce different concept maps, and thus have different characteristics.

Text writing

While text analytics and approaches of text mining still have huge deficits especially for short texts written by students in STEM fields (Leeman-Munk, Wiebe, & Lester, 2014), we used a non-automatic measurement for the text quality characteristics. A teacher creates a model solution and scores the texts in respect to the model. This led to plausible scores without the possible downsides of text mining on short texts in sciences.

Motivation

As motivation is one of the key ingredients for successful group work, we incorporated three measures for motivation towards science. The SMTSL questionnaire has been used in a shortened version to assess scores in three different categories of motivation: (a) Self efficacy, (b) Science Learning Value, and (c) Learning Environment Stimulation.

Evaluation

In this study we aim to explore how group formation affects the practice of students and their performance in collaborative learning activities. To that end, we formed homogeneous and heterogeneous student groups using a multidimensional clustering schema based on artefact-related characteristics and motivational scores, as described above. In order to evaluate the practice of students we used both a qualitative (expert observations) and a quantitative approach (learning analytics). In order to assess the students' performance, we carried out pre-knowledge and post-knowledge tests. In the following paragraphs, we present the results of the analysis and discuss the findings of the study.

Quantitative analysis

The interaction of students with the learning platform was recorded in log files. We used the log files to extract metrics of students' activity and further explore any possible relation with qualitative characteristics and the overall knowledge gain. The scores of the knowledge tests ranged from 0 to 35 points and we used them to assess the learning outcome. Additionally we defined the activity metrics portrayed in Table 1 in order to evaluate the interaction of students with the learning platform.

Table 1: Activity metrics extracted from user log files

Metrics of students activity on the concept map		
	name	description
learning platform related activity	#actions	number of actions
	duration (min)	overall duration
	avgtimegap (sec)	time gap between consecutive actions on average
concept map related activity	#concepts	number of created concepts
	#relations	number f drawn relations
	#add	number of added objects
	#update	number of updates
	#delete	number of deleted objects

In Table 2, we present the results of the knowledge tests per group. According to the results, the heterogeneous groups appeared to have a higher knowledge gain than the homogeneous groups. The heterogeneous groups improve their score in the post knowledge test on an average of 33% while the homogeneous groups improved their score about 20%. In the current study, group homogeneity does not ensure that the members of a group share similar knowledge background. For example, the members of group G2 that is considered heterogeneous, scored similarly in the pre-knowledge test (pre-STDEV = 0.5). On the other hand, the pre-test scores of the members of group G6 that is considered homogeneous, portray a big deviation (pre-STDEV = 6.50).

Table 2: Results of the pre and post knowledge tests for heterogeneous and homogeneous groups

	Heterogeneous groups				Homogeneous groups			
	G1	G2	G3	avg(G1_3)	G4	G5	G6	avg(G4_6)
Pre-test score	16.00	15.33	15.50	15.61	12.67	18.67	12.50	14.61
Pre-STDEV	1.41	4.78	0.50	2.23	2.05	1.70	6.50	3.42
Post-test score	23.33	23.00	23.50	23.28	16.17	24.00	15.00	18.39
Post-STDEV	3.40	4.90	0.50	2.93	3.32	1.47	5.00	3.26
Avg gain	7.33	7.67	8.00	7.67	3.50	5.33	2.50	3.78

The results of the knowledge tests were studied in comparison with the metrics of user activity. However, we were not able to draw any plausible conclusion for possible relations. The groups' activity, as portrayed in the log files of the learning platform, was similar for all groups (Table 3). A common hypothesis made in similar studies is that collaboration quality and knowledge gain are usually depicted in activity metrics, i.e. intense activity will lead to a solution of better quality (Kahrimanis, Chounta, & Avouris, 2010). This hypothesis however was not confirmed in this study.

Table 3: Group activity metrics for heterogeneous and homogeneous groups

	Heterogeneous groups			Homogeneous groups		
	G1	G2	G3	G4	G5	G6
#actions	30	57	56	38	56	60
duration (min)	23.82	48.42	21.67	16.77	25.67	29.75
avgtimegap (sec)	49.28	52.94	23.64	27.19	28.00	30.25
#concepts	13.00	28.00	24.00	19.00	23.00	29.00
#relations	10.00	25.00	28.00	12.00	23.00	27.00
#add	13.00	24.00	24.00	17.00	24.00	26.00
#update	12.00	24.00	28.00	14.00	23.00	27.00
#delete	0.00	6.00	1.00	2.00	2.00	4.00

Qualitative analysis

During the group phase of the study, the users had to create a concept map based on what they learned and to write a report. A teacher rated both the concept maps and the reports of the groups. This way, we wanted to ensure the findings from the pre and post knowledge tests. The concept maps were rated in a [0, 8] range and the reports were rated within the range [0, 12]. The ratings of the teachers for the concept maps and the work reports are presented in Table 4. The results of the ratings with respect to group homogeneity confirm the findings of the knowledge tests. The heterogeneous groups are higher graded than the homogeneous ones for both the concept maps (21.4%) and the final work reports (29.6%).

Table 4: Teacher ratings of the concept maps and work reports per group

	Heterogeneous groups				Homogeneous groups			
	G1	G2	G3	avg(G1_3)	G4	G5	G6	avg(G4_6)
Concept Map Scores	4	7	3	4.67	1	2	2	1.67
Report Scores	10	11	6	9.00	2	6	4	4.00

The practice of students was recorded in transcripts by two experts who attended the study. In addition, a third expert took general notes of the activity (e.g. notes about the timeline and events that might affect the activity). From the analysis of the transcripts, we identified three main group types: (a) Type A: One student operates the computer, the rest comment or guide him verbally, (b) Type B: Group members change roles frequently regarding typing and directing and (c) Type C: One student is actively involved in the task, the others watch silently or do not do pay attention.

Two out of three heterogeneous groups (G1 and G3) were identified as of type B. The experts stated that even though they started out shyly, they managed to create a common ground and share responsibilities and tasks. They were enthusiastic about the activity until the end and seemed to enjoy it. For the group G3 in particular, the experts noted that they did not communicated openly (talking or arguing, etc.) and sometimes they were hesitant to act. Towards the end of the activity they did not interact between them but they carried on working separately even though they shared the use of the computer. This group score the highest score in the post-knowledge test and the maximum knowledge gain. The third heterogeneous group (G2) was identified as type C. According to the experts' observations, one particular student took over the activity but continuously tried to involve the other members by giving detailed explanations on every step of the process.

For the homogeneous groups, two were identified as type A (G4 and G5). According to the transcripts, the students of both groups were hesitant in the beginning. For group G4, it took them considerable effort to start communicating and one student took action in order to move forward with the activity. In group G5 one student appeared to be more aggressive and active and dominated the activity from the start. Gradually all students began to participate within their groups. For group G4, however, it was too late to catch up while group G5 lost motivation towards the end. We think that the time the individuals needed in order to coordinate with the rest of the rest of the group members was critical and in the end this is depicted in the learning outcome. The third homogeneous group (G6) was identified by experts as type C. According to the experts' observations, one group member carried out the whole task while the other one was silently watching. Despite the fact that the active student tried to involve the other member in the activity, there was no collaboration or argumentation. Group G6 had the lowest score on average in the post-knowledge tests and also the minimum knowledge gain.

It is worth mentioning that the groups which were identified as type C had the maximum group deviation in the pre-knowledge tests (Pre-STDEV, Table 2). This practically means that in both groups there was a “strong” student who eventually dominated the activity. However, in the case of the heterogeneous group the knowledge gain on group average was higher than in the case of the homogeneous group.

Discussion

In this paper, we discuss the effect of group formation strategies on students’ collaboration and the learning outcome. Analysis showed that the heterogeneous groups increased their performance and appeared to have a higher knowledge gain than the homogeneous groups. On an individual level, the students who were members of heterogeneous groups had a knowledge gain of 33% on average while the students who formed homogeneous groups improved their individual performance for about 20% with respect to the pre-tests. This finding was also confirmed by the teacher ratings of the concept maps and the group reports. Heterogeneous groups were graded higher than homogeneous groups for the quality of the concept maps they provided through the learning platform and for the quality of the written reports.

In order to assess the group activity with respect to collaboration quality, we used activity transcripts where experts recorded their observations. The experts stated that the students of heterogeneous groups adjusted their practice easier than the students of homogeneous groups. They undertook roles and responsibilities faster and without conflicts. Even in the case when they didn’t seem to communicate on a satisfactory level, they managed to carry out the task efficiently. On the other hand, homogeneous groups needed more time in order to create a common ground and to collaborate effectively. In some cases this was proven critical since some of the students lost interest and others were unable to carry out the task in time.

Additionally, we used the log files of the learning platform to define metrics of user activity. To that end, we followed popular approaches where activity metrics were introduced as indicators of good collaboration quality or efficient group practice (Kahrimanis, Chounta, & Avouris, 2010). However, we were not able to prove any relation between activity metrics and the learning outcome. The group practice was similar for most cases with respect to activity metrics and group homogeneity. We should however keep in mind that due to the study setup one could argue that the activity metrics do not reflect group work or collaborative practice and therefore we should not expect a correlation with the overall group picture.

Conclusion

To tackle the issue of how to engage students in sciences and capture their interest, we propose the usage of rich inquiry-based learning scenarios, as demonstrated in the Go-Lab project. Incorporate online learning with classroom presence leads to blended learning scenarios. This gives the opportunity to take the collaborative parts of the learning into the classroom, with all its benefits and challenges for the teacher. We propose a way to support the group orchestration through the application of learning analytics, particularly the analysis of learning objects and assessed motivation. Finally, we conducted an experiment and applied methods of sequence and log file analysis to validate our hypotheses through multi-level analysis.

The analysis of our results indicated that heterogeneous groups outperform the homogeneous ones and achieve higher knowledge gain. Thus, there is no benefit of choosing homogeneous groups in terms of performance. Even though when having a group with only good performing students, they still do not perform significantly better than the heterogeneous groups, but they don’t compensate the weaker performance in the other homogeneous groups with weaker characteristics. For the class average, heterogeneous groups are better in sum, while it also covers basic principles of fairness, which is reflected by a lower diversity between the groups’ performances. Fairness is both a principle that can influence the motivation of students in a further way but also underpins pedagogical decisions and thus is one of the important steps towards successful internal differentiation of learner groups. The results of this study cannot be generalized due to the small number of participants; however they can serve as indications for group formation. In future work, we aim to conduct large-scale studies in order to confirm the outcome in a statistically significant way.

References

Balmaceda, J. M., Schiaffino, S. N., & Pace, J. A. (2014). Using constraint satisfaction to aid group formation in CSCL. *Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial, 17*(53), 35-45.

Chounta, I.-A., Giemza, A., & Hoppe, H. U. (2014). Multilevel Analysis of Collaborative Activities Based on a Mobile Learning Scenario for Real Classrooms. In *Collab. Techn. and Soc. Computing*,127-142.

Conlon, T. (2004). 'Please Argue, I Could Be Wrong': a Reasonable Fallible Analyser for Student Concept Maps. *World Conf. on Educational Multimedia, Hypermedia and Telecomm., 2004*, 1299-1306.

de Jong, T., Sotiriou, S., & Gillet, D. (2014). Innovations in STEM education: the Go-Lab federation of online labs. *Smart Learning Environments, 1*(1), 1-16.

Dillenbourg, P. (2013). Design for Classroom Orchestration. *Comput.& Educ., 69*, 485-492.

Dimitriadis, Y., Prieto, L. P., & Asensio-Pérez, J. I. (2013). The role of design and enactment patterns in orchestration: Helping to integrate technology in blended classroom ecosystems. *Comput.& Educ, 69*, 496-499.

Gago, J., Ziman, J., Caro, P., Constantinou, C., & Davies, G. (2004). Europe needs more scientists. *Brussels: European Community Conference Increasing Human Resources for Science and Technology.*

Gijlers, H., & De Jong, T. (2005). The relation between prior knowledge and students' collaborative discovery learning processes. *Journal of Research in Science Teaching, 42*(3), 264-282.

Global Science Forum. (May 2006). *Evolution of Student Interest in Science and Technology Studies.* Organisation for Economic Co-operation and Development.

Govaerts, S., Cao, Y., Vozniuk, A., Holzer, A., Zutin, D. G., Ruiz, E. S., . . . others. (2013). Towards an online lab portal for inquiry-based stem learning at school. In *Advances in Web-Based Learning*, 244-253.

Hannafin, M., Hill, J., Land, S., & Lee, E. (2014). Student-Centered, Open Learning Environments: Research, Theory, and Practice. In *Handbook of Research on Educ. Communications and Technology*, 641-651.

Hoppe, H., Engler, J., & Weinbrenner, S. (2012). The impact of structural characteristics of concept maps on automatic quality measurement. *Proceedings of the 10th ICLS. Sydney, Australia.*

Jermann, P., Soller, A., Muehlenbrock, M. (2001). From mirroring to guiding: A review of the state of art technology for supporting collaborative learning. *Proceedings of the EuroCSCL-2001*.

Kahrimanis, G., Chounta, I. A., & Avouris, N. (2010). Study of correlations between logfile-based metrics of interaction and the quality of synchronous collaboration. *Int. Reports on Socio-Informatics, 7*(1).

Kizilcec, R. F. (2013). Collaborative Learning in Geographically Distributed and In-person Groups. *AIED 2013 Workshops Proceedings Volume*, 67.

Konert, J., Burlak, D., & Steinmetz, R. (2014). The Group Formation Problem: An Algorithmic Approach to Learning Group Formation. In *Open Learning and Teaching in Educ. Comm.*, 8719, 221-234.

Leeman-Munk, S. P., Wiebe, E. N., & Lester, J. C. (2014). Assessing Elementary Students' Science Competency with Text Analytics. *Proceed. of the 4th Int. Conf. on Learning Analytics And Knowledge*.

Lou, Y., Abrami, P. C., & d'Apollonia, S. (2001). Small group and individual learning with technology: A meta-analysis. *Review of educational research, 71*(3), 449-521.

Manske, S., Hecking, T., Bollen, L., Göhnert, T., Ramos, A., & Hoppe, H. U. (2014). A Flexible Framework for the Authoring of Reusable and Portable Learning Analytics Gadgets. *ICALT, 2014*, 254-258.

McClure, J. R., Sonak, B., & Suen, H. K. (1999). Concept map assessment of classroom learning: Reliability, validity, and logistical practicality. *Journal of research in science teaching, 36*(4), 475-492.

Novak, J. D. (1984). *Learning how to learn.* Cambridge University Press.

Ounnas, A., Davis, H. C., & Millard, D. E. (2009). A Framework for Semantic Group Formation in Education. *Educational Technology \& Society, 12*(4), 43-55.

Rocard, M., Csermely, P., Jorde, D., Lenzen, D., Walberg-Henriksson, H., & Hemmo, V. (2007). *Science education now: A renewed pedagogy for the future of Europe.* European Commission.

Roschelle, J., Dimitriadis, Y., & Hoppe, U. (2013). Classroom Orchestration: Synthesis. *Comput.& Educ., 69*, 523-526.

Sadeghi, H., & Kardan, A. A. (2014). Toward effective group formation in computer-supported collaborative learning. *Interactive Learning Environments, ahead-of-print*, 1-14.

Stahl, G., Koschmann, T., & Suthers, D. (2006). Computer-supported collaborative learning: An historical perspective. *Cambridge Handbook of the Learning Sciences, 2006*, 409-426.

Stoddart, T., Abrams, R., Gasper, E., & Canaday, D. (2000). Concept maps as assessment in science inquiry learning-a report of methodology. *International Journal of Science Education, 22*(12), 1221-1246.

Tuan, H.-L., Chin, C.-C., & Shieh, S.-H. (2005). The development of a questionnaire to measure students' motivation towards science learning. *International Journal of Science Education, 27*(6), 639-654.

Webb, N. M., Nemer, K. M., & Zuniga, S. (2002). Short circuits or superconductors? Effects of group composition on high-achieving students' science assessment performance. *American Educational Research Journal, 39*(4), 943-989.

Acknowledgments

The Go-Lab project is partially funded by the European Union (Grant Agreement no. 317601) under the Information and Communication Technologies (ICT) theme of the 7th Framework Programme for R&D (FP7). This document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of its content.

