ASCILITE
SHAPING THE FUTURE OF TERTIARY EDUCATION
Show Me The Learning
Adelaide, Nov 27-30, 2016

Visualising Individual Profiles and Grouping Conditions in Collaborative Learning Activities

Augusto Dias Pereira dos Santos
School of Information Technologies
The University of Sydney

Kalina Yacef
School of Information Technologies
The University of Sydney

Roberto Martinez-Maldonado
Connected Intelligence Centre
The University of Sydney

Collaborative learning has been shown to be conducive to better and deeper learning for particular tasks, but is dependent on a number of factors, including how students are grouped together. We are interested in finding out whether data captured from students working individually and/or collaboratively can reveal useful information about the impact of the grouping conditions on learning. We explore whether these findings can be detected early on (possibly, before students start working in groups). If such information can be reliably captured, then it could be used to drive group formation dynamically and at a large scale. This paper presents our initial visual exploration with two case studies: one from a first-year programming course (N = 372) where students alternately worked individually and in pairs; and another (N = 60) from a concept-mapping environment where students first worked individually and then in groups.

Keywords: Collaborative Learning, Group Formation, Visualisation, Clustering

Introduction

Collaborative learning activities require a strategy to determine how groups are formed. In very small classes, teachers generally know their students well and can use their pedagogical knowledge to assign students to groups in a way that is expected to maximise their learning benefits. However, such a strategy is not scalable to larger cohorts as the complexity of the task increases exponentially with the number of students (Sinha, 2014). As a result, teachers often rely on other techniques, such as self-arrangement (students decide who they work with and what roles they enact) or random assignment (e.g. using a systematic process such as alphabetical order to achieve random group allocation) (Cohen, Goodlad, Darling-Hammond, & Lotan, 2014). However, none of these methods directly attempt to maximise the learning benefits for the students. Some group formation algorithms have been proposed to facilitate group allocation. These algorithms automatically assign students to groups based on specific criteria selected by the teacher (Craig, Horton, & Pitt, 2010; Demetriadis & Karakostas, 2008; Konert, Burlak, & Steinmetz, 2014). However, the teacher may not always be aware of which are the most relevant criteria for the given task. If teachers were able to deepen their understanding of how certain aspects of students' activities have an impact on collaborative work, they would be able to better select the student's activity features that are most relevant for the desired collaborative task. Examples of features include past performance on individual assessments, demographic data, or previous group work strategies. Educational technologies are commonly present in teaching environments, especially as student cohort sizes increase. These tools, when supporting students in their learning tasks, capture unprecedented amounts of data about the students' learning behaviour and progression, sometimes at very finegrained levels (Verbert et al., 2014). Research communities, such as Learning Analytics and Knowledge (LAK) and Educational Data Mining (EDM), use these data as an opportunity to improve education by understanding its processes; plan and select interventions; and improve assessments (Siemens & d Baker, 2012). It is possible that these data can also reveal useful information about what makes students work better in groups.

In this paper, we explore two questions. First, can data captured from students' previous individual or group work reveal useful information about whether and how grouping conditions affect learning? Second, can this information be made easily accessible to teachers, using visualisations? We present a data-driven approach for understanding how the student's individual profile, inside a group, is related to group collaboration and performance. By datadriven, we mean relying on data to formulate our assumptions, instead of theory-driven, where a hypothesis is based on theories and data is used to validate them (Choi et al., 2016). We explore this by clustering the students, using different individual information, and plotting it against several measures of group performance. Our aim is to help teachers understand the individual profiles in the collaborative task, which will assist them in planning collaborative activities in future iterations. We illustrate our approach with two very different datasets: one collected in a classroom of 1st-year programming students working both individually and in pairs over a semester through an online programming tool (N=372), and another from a cohort who worked with concept maps, first individually and then face-to-face around an interactive tabletop (N = 60).

Background

Learning is commonly a social process and thus collaborative learning research focuses on unveiling the complex social mechanisms that are associated with learning. There are several important aspects to come to play in collaborative learning: the notion of what a group is, what learning means, and what collaboration is (Dillenbourg, 1999). The notion of group is already quite complex. Researchers investigating collaborative learning and psychological researchers studying groups have been trying to understand how groups behave for almost 60 years (Beal, Cohen, Burke, & McLendon, 2003). Some important dimensions that can strongly shape the collaborative learning process include the size of the group (e.g. 2, 10, 40 students or a community of learners); the length of time the group works together (e.g. 30 minutes, one day, the whole semester, life-long learning); the task involved (e.g. solve a puzzle, write a proposal, code a system, build a robot); and the subject matter (e.g. Science, Technology, Engineering, Maths (STEM), Humanities, Health and so on). Researchers all agree that there is no ultimate solution fitting all the different group configurations as the final outcome and the collaborative processes depends on the different combinations of all the possible dimensions of the collaborative activity (Stahl, 2006).

One important question in collaborative learning research is knowing which students are going to work together more effectively– in other words, how to arrange the groups in a classroom in order to maximise their opportunities for learning. Some research has addressed the problem of group formation in learning contexts by satisfying constraints defined by the teacher. In these scenarios, it is assumed that the teacher knows which selection criteria are best. Konert et al. (2014) compared many other alternative solutions to form learning groups, including assisting the teacher in forming groups through algorithms that maximise the opportunity of collaborative learning higher achievement (teacher-driven approaches). Group formation algorithms can be classified by local and global evaluation methods (fitness function); the number of criteria; criteria weighting; and homogeneous, heterogeneous and mixed-group options. An example of these algorithms is FROG (Craig et al., 2010) which allows the teacher to choose different types of attributes, such as numeric, categorical, and timetable attributes. For each attribute, it is possible to choose an evaluation method, such as homo/heterogeneity, average, at least one, or at most one. The evaluation method is defined for groups and/or overall for the whole class. Another example is GroupAL (Konert et al., 2014) which redefined the evaluation method, introducing performance indices for pairs of students, groups, and the entire cohort, as well as matching approaches that are group-centric and participant-centric. However, these teacher-driven approaches require the teacher to know exactly what parameters and weights to use, which often is not the case. Our work aims at addressing this important step: we propose a visual approach to assist teachers exploring how various criteria influence their students' group performance.

In regards to previous research on visualisations or dashboards to enhance a teacher's awareness, very few of them targeted visualising different aspects of collaborative learning. Most researchers have focused on visualising what occurs during the collaborative process. For example, (Martinez, Kay, & Yacef, 2011), created a dashboard that shows the teacher, in real time, three aspects of the collaboration: students' verbal and physical participation in the group; interactions between participants; and overall collaboration level as assessed by a machine learning algorithm. At Class-on, Rojas and García (2012) created a map of the classroom using colours and numbers to present information of the groups, such as time taken in a task, progress, and information that will help manage student assistance. A more extensive survey on learning dashboards counted only four studies targeting collaborative learning (Verbert et al., 2014). Our research aim is not to assist collaboration during the activity, but after it, and to support the teacher in gaining understanding about the data that can be useful to tune their group formation strategies.

Approach

Our general approach is to keep the teacher in the loop and support them to make informed, data-driven decisions. We explore in this paper whether or not the data captured from students' previous activities – either individual or collaborative – can reveal relationships between the combination of certain student profiles and learning or collaboration outcomes. For this, we extract these student profiles and then provide a visualisation that can be used to explore these relationships. The proposed visualisations are initial building blocks towards assisting teachers in making data-driven decisions for forming student groups. Our approach is two-stepped:

1. Generate student profiles through a data-mining technique that clusters students according their individual data (e.g. behaviour, performance).

2. Create a visual learning analytics interface allowing teachers to rapidly examine these profiles according to specific criteria of group performance.

Clustering Method

Given that the social and epistemic aspects of the groups can strongly shape both the collaborative activity and the learning task itself (Carvalho & Goodyear, 2014), we illustrate the potential of our approach with two datasets captured in two very different collaborative learning contexts. Understanding the context where the data comes from is a crucial step, as it needs to be pre-processed before applying statistical or data-mining algorithms to extract meaningful information.

We decided to use a clustering algorithm to extract common student profiles, as it provides good results for profiling students according to their behavioural or performance traces (Bovo, Sanchez, Héguy, & Duthen, 2013; Kardan, Roll, & Conati, 2014; McBroom, Jeffries, Koprinska, & Yacef, 2016). The clusters provide a high-level description of the different types of profiles found among students. The number of clusters was decided based on a voting system using several indexes[1]. Those indexes compute the optimal number of clusters based on metrics, such as cohesion inside the cluster member and separation between different clusters. After, we experimented with different clustering algorithms and chose the ones that extracted profiles with the most meaningful characteristics. As a result, we used the K-means and EM clustering algorithms and visually selected the resulting clusters that provided a better representation of different student profiles. The next step was to design a simple visualisation tool for helping teachers investigate the contrasting characteristics occurring in the different profiles and associate the co-occurrence of these profiles with collaborative performance.

Visualizations

Two types of visualizations are proposed in this paper. Together, they may help the teacher understand their collaborative design more deeply. The tile chart, shown in Figure 1, presents all the students in the classroom distributed in their respective groups. This visualisation may assist the teacher in identifying patterns relating the students' individual profiles and group performance. Each student is represented as a tile coloured by their cluster profile. Each group is organized in one column, where every line of the column includes the students in that group. Inside each group column, the students are ordered by their cluster number. This helps the teacher identify patterns among groups with the same or similar profiles arrangement. A gradient bar is presented below the first part of the chart, presenting the performance of each group, with the intensity of colour ranging from white to dark green, where white indicates low performance and dark green indicates high performance. The gradient bar is also divided by tiles, where each tile represents one group. For instance, in Figure 1, the first column shows that the group that achieved the lowest performance consisted of three students, who were respectively in cluster 3 (green), 2 (blue) and 1 (red), whereas the highest performance group was composed of students in cluster 4 (yellow), 1 (red) and 1 (red) respectively. Looking at the overall distribution, students from cluster 4 (yellow) tend to have good results, while students in cluster 2 (blue) have poor results. More specifically, cluster 1 (red) had good results when working with cluster 4 (yellow), but poorer results when working with cluster 2 (blue).

[1] https://cran.r-project.org/web/packages/NbClust/NbClust.pdf

The second visualisation, shown in Figure 2, is a boxplot chart (Tukey, 1977) of the student profiles extracted from the clustering algorithm. This chart may assist the teacher in understanding how one profile is different from the others and gain understanding about the data that was used to define each cluster. Each boxplot represents the population distribution of one cluster. The thick line in the middle of the bar is the median of the population. The bar represents the interquartile range (IQR) of the population. The upper limit of the bar is the upper quartile (Q3), the lower limit is the lower quartile (Q1), and the upper (Q3 + 1.5 * IQR) and lower (Q1 - 1.5 * IQR) whiskers represent the maximum and minimum value of the population, respectively. Values beyond the whiskers are consider outliers and represented by dots. In Figure 2, we also present the same data that is presented in the boxplot chart, plotted in a histogram chart to visually explain how the boxplot represents the data. To explore this problem, in the next section, two versions of this chart are provided, each one containing different dimensions from the student's individual profile and also different measures of group performance. This would allow the teacher to choose which perspective most represents their intention when designing the new collaborative activity.

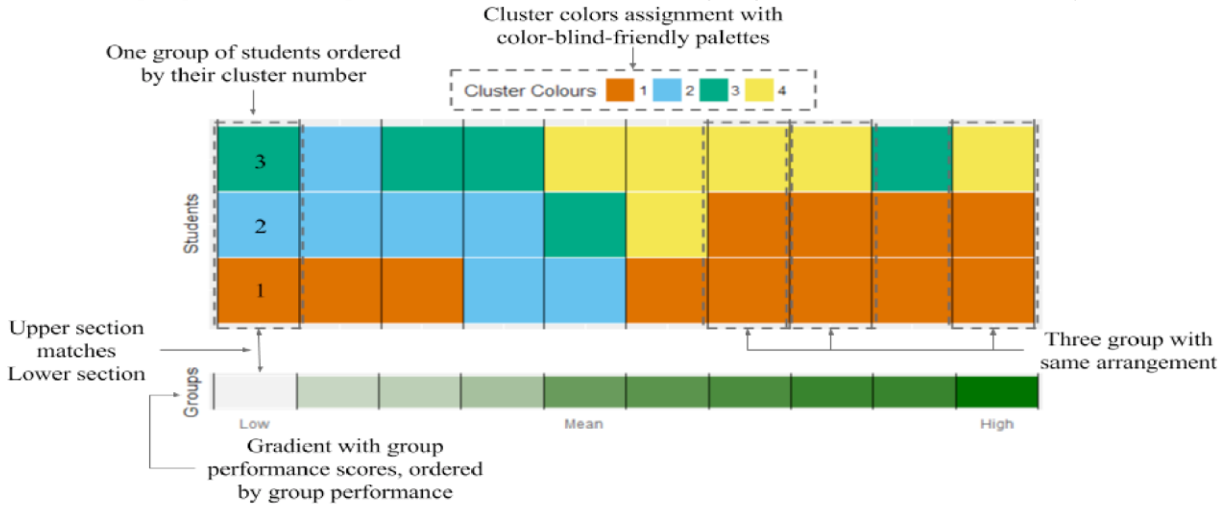

Figure 1: Example of Tile Chart

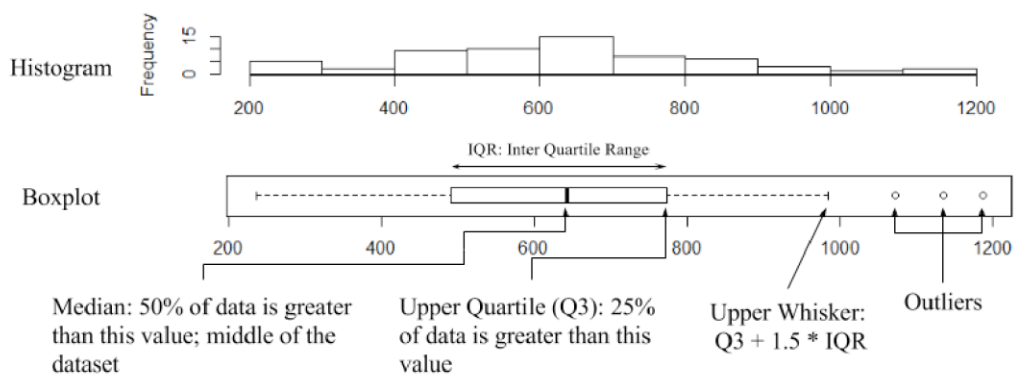

Figure 2: Example of Boxplot Chart compared with Histogram

Case Studies

To illustrate our clustering and visualisation approach for showing the relationship between students' individual profiles in a collaborative learning context, we used data from two different case studies. The datasets from these two studies (1 and 2) present very different characteristics, which enriches the illustration of our approach, and justifies the need for a flexible data-driven methodology as well as a tool to enhance teachers' decision-making by including the teacher into the analysis loop. The learning situations vary in terms of scale (n = 372 for Study 1 and n = 60 for Study 2), group sizes (dyads and triads respectively), learning modalities (blended collaboration and full face-to-face group work), time-scale (1 university subject and only 1 group session), learning tasks (pair programming and concept mapping) and domains (engineering and nutrition). The following subsections provide more details about the studies, present the clustering analysis, show the resulting visualizations, and provide a discussion of the results for each case.

Study 1: PASTA - Automated Programming Assignment Assessments

Learning Situation and the Dataset

The first dataset was collected through an in-house automatic marking and instant feedback system named PASTA (Koprinska, Stretton, & Yacef, 2015). This system was used during the second semester of 2015 for the Data Structures unit of the Computer Science at the University of Sydney. There were 372 students enrolled to this unit. The assessment comprised of individual weekly programming tasks (from weeks 2 to 12) and two group programming assignments (in weeks 8 and 11). For each student we therefore had individual activity data interspersed with group assignments. We tested our methodology on the outcome of the first group assignment (in week 8) by visually exploring the students' profiles, which we clustered using the data from individual tasks from week 2-5 (groups were formed in week 6 so, for authenticity, we only used the individual data across these weeks). Each of these weekly activities consisted of a series of programming tasks that students submitted to PASTA. The system uses unit tests to provide feedback. The students could re-submit their solutions as many times as they needed before the deadline. A progress bar showed the number of tests they passed on each attempt, with the final percentage being used to calculate their mark for their submission. More technical details about PASTA can be consulted in Gramoli et al. (2016).

Out of the 372 students, 162 submitted the first group assignment in pairs, while the remainder worked alone or did not submit at all. For each assessment, the PASTA logs contain: 1) information about the student's behaviour (e.g. number and timing of submissions) and 2) submission quality as assessed by the system. Table 1 summarizes the attributes from the data used in this study, where each attribute is replicated in each week, from week 2 to week 5.

Table 1: Nine individual attributes description for each weekly individual task

Attribute	**Description**
percent_early	Percentage of attempts made three days or more before the due date
percent_normal	Percentage of attempts made that were neither early nor late
percent_late	Percentage of attempts made on the due date
num_compile_errors	Number of attempts involving compilation errors
first_mark (0-100)	Percentage of tests passed on first attempt
last_mark (0-100)	Percentage of tests passed on last
num_submissions	Number of attempts not involving compilation errors
time_taken (seconds)	Time taken from the first to the last submission
avg_improvement	Average improvement from the first to the last submission

Students Profile Generation

A number of student profile clusters were generated aggregating the weeks from 2 to 5 for each feature. For instance, the feature first_mark from week 2 to week 5 was summarised using mean and standard deviation. So, first_mark_mean represents the average first mark of the student when doing the submission, and first_mark_sd describes how regular/irregular their first submissions were. The resulting summarised features, for each of the students, were then used as an input for the cluster methods (K-means). Our intent is to extract the student's behaviour using the platform over the weeks that anticipate the group formation for the first group assignment. The charts are ordered based on the group's first assignment result. On the left side of the chart, are the groups with the lower scores, and on the right side of the chart are the groups with higher scores.

Students Clustered by First and Last Marks

For this clustering task, we only selected the first and last marks feature to cluster students. The resulting clusters are shown in Figure 3. Cluster 1 (red, 38 students) is comprised of students with a first mark average of 66 and an improvement to 73 in the last mark. These students did not have a regular first and last mark over the four-week assignment, with a high standard deviation of 42 in both the first and last mark. Cluster 2 (blue, 29 students), aggregates students with low marks and no improvement from the first to the last attempt. Their average mark was 38.5 and a high variability of 51 standard deviations between assignments. Cluster 3 (green, 74 students) had the best students with high marks, low variability and a mild improvement from 96 to 99. Cluster 4 (yellow, 21 students) contains students that started well and finished with excellent marks. They have a high improvement rate, from 79 to 96, and a decrease on the variability from 29 to 5, in the last mark.

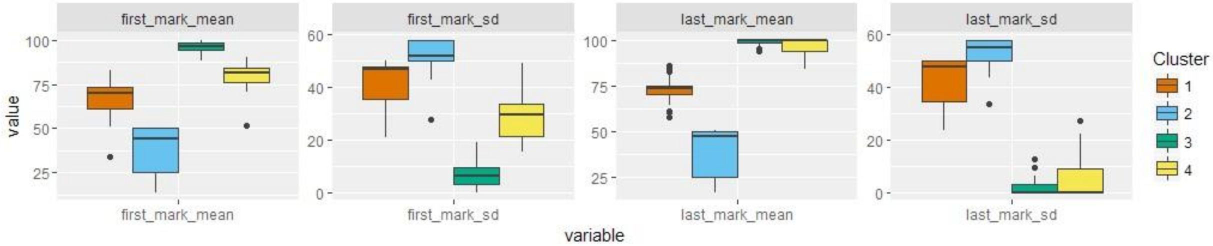

Figure 3: Students clustered by first and last mark

Figure 4 depicts all the dyads that performed the first group assignment. Each column represents a dyad and the coloured squares represent the cluster that each student belongs to. The dyads are ordered by score on the first group assignment (from left to right). Interestingly, the visualisation shows how most of the groups had a really good performance in the first group assignment (e.g. the mean indicator is leaning to the left side and most of the gradient bar show a high value, dark green colour). The average is 5.8 from a maximum score of 8, and the standard deviation is ±1.7. In this figure, we can observe that:

1. Students that individually have good marks tended to have good marks when in groups, especially when working together with another student from the same cluster. Cluster 3 (green) is more prevalent in the right part of the chart.
2. Students with low individual performance tended to also have low marks when working with groups, especially when working together. Cluster 2 (blue) has a tendency to the left of the chart.
3. Clusters 1 (red) and 4 (yellow) are spread throughout the chart showing no trend when working in groups.

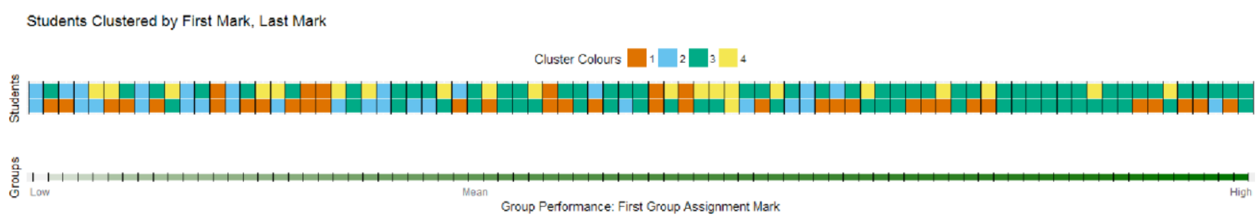

Figure 4: Students in their groups, coloured by their profiles and ordered by group performance

Students Clustered by Percentage Early, Normal, Late

The second analysis in this dataset consisted of exploring group behaviour based on other individual features. The next group of features that resulted in meaningful profiles was related to the submission times – that is, the percentage of early, normal, and late submissions.

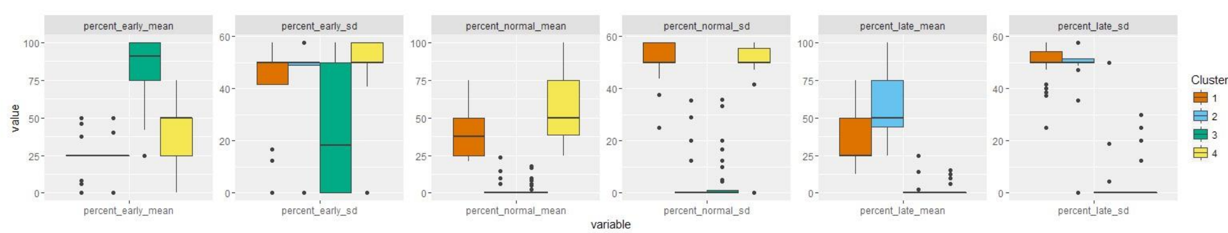

Figure 5: Students clustered by percentage early, normal, late

In Figure 5, it is possible to compare the four profiles of students with regards to the timing of their submissions. Cluster 1 (red, 49 students) contains students that presented irregular behaviour, submitting early, normal and late, with high variability along the four-week assignments. In cluster 2 (blue, 29 students) students tended to submit late.

Cluster 3 (green, 44 students) contains students who were generally consistent in making early submissions. Cluster 4 (yellow, 40 students) includes students who made submissions early and normal, but never late. In this analysis, the results are not as clear as in the analysis described in the previous subsection. For example, there is no evident trend regarding the profile distributions in Figure 6, such as the ones we saw in Figure 4. However, it is possible to observe some group behaviours, shown in Figure 6:

1. Some students that consistently performed late submission were associated with low group performance. Cluster 2 (blue) gathers at the left end of Figure 6.
2. Some groups with both students from students that made early submissions, cluster 3 (green), did not perform well together.
3. Students with irregular submissions are gathered around the mean group score, especially when working together.

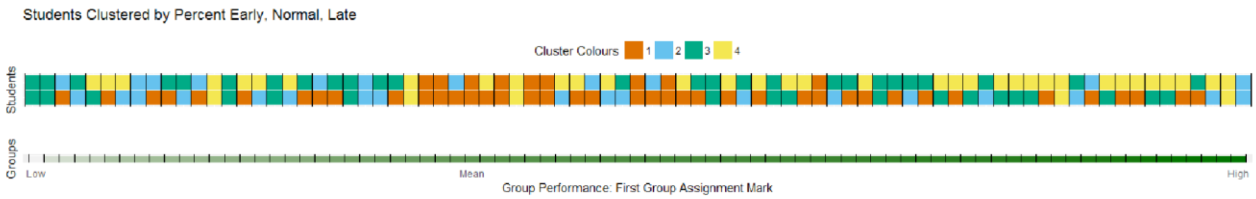

Figure 6: Students in their groups, coloured by their profiles and ordered by group performance

The previous charts show how some simple visualisations may help to make visible certain trends happening as the semester unfolds. It also may give the teacher the ability to understand how the arrangement of students in groups can influence group performance. We changed the variables used to cluster students to explore different profiles of students. Even though the results were simple, they provide insights for further investigations.

Study 2: CMATE - Building Concept Maps Using Tabletop

Learning Situation and the Dataset

In an experiment involving 60 students from science courses, participants were asked to build a concept map after reading the Australian Dietary Guidelines 2011(Martinez-Maldonado, Dimitriadis, Martinez-Monés, Kay, & Yacef, 2013). The experiment had three phases: first, the students built the concept map individually; second, grouped in triads, the students built concept maps together; and third, students were asked to build the concept map again individually. During the individual phases, the students were asked to build the concept map using CmapTools that recorded each student's steps when building the map. At the group phase, the students built the concept map in an interactive tabletop system called CMATE, where all the touches were recorded, together with audio and video. A method that uses nine qualitative dimensions was used to assess the group's collaboration quality level. We aggregated the scores to come up with a single indicator of quality of collaboration that we will refer in this paper as a Spada score. More technical details about the learning situation and CMATE can be found in Martinez-Maldonado et al. (2013).

In this analysis, we mainly focused on how the individual profiles of the students, which were extracted from individual concept maps (hence data collected prior collaborative task) influenced the group collaboration. We used two measures of group performance: the Spada score and a comparison score with a Master Map created by an expert of the subject matter (Martinez-Maldonado et al. 2013).

The individual features that were extracted from the individual concept maps are shown in Table 2.

Table 2: Features extracted from the concept maps

Type	Description	Examples
Graph Analysis	Reading the concept map as a graph, producing features.	Avg Links per Concept, Avg Words per Concept, Taxonomy Score, Proposition Count
Time Analysis	Sum of how much time the student spends before executing each action.	Add Concept Time, Delete Connection Time, Move Concept Time, Resize Concept Time
Process Analysis	Coding each action as a letter and producing triples and quadruples of sequences, and analysing the most frequent sequences.	CBB: Add Linking Phrase → Add Connection → Add Connection JJJJ: Move Concept → Move Concept → Move Concept → Move Concept

Student Profile Generation

We also explored different combinations of features to cluster students. The one that produced the most meaningful profiles was based on Time Analysis features. This cluster schema was designed to extract peculiarities of each student when building their concept map. In contrast to the previous dataset, and to show the potential of the proposed visualisation, we plotted the same students' profile clustering over different measures of group performance. The cluster profiles are shown in Figure 7. Cluster 1 (red, 5 students) is comprised of students that did not spend much time adding new concepts but a lot of time moving them around. Cluster 2 (blue, 13 students) gathers students that spent more time adding and moving linking phrases. Cluster 3 (green, 11 students) has students that spent more time adding concepts and linking phrases. Cluster 4 (yellow, 31 students) contains students that spent much less time working on their individual concept map compared to the other profiles.

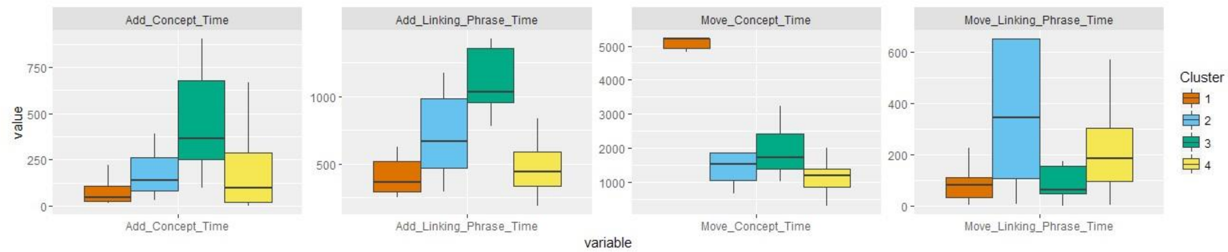

Figure 7: Students clustered by time analysis features

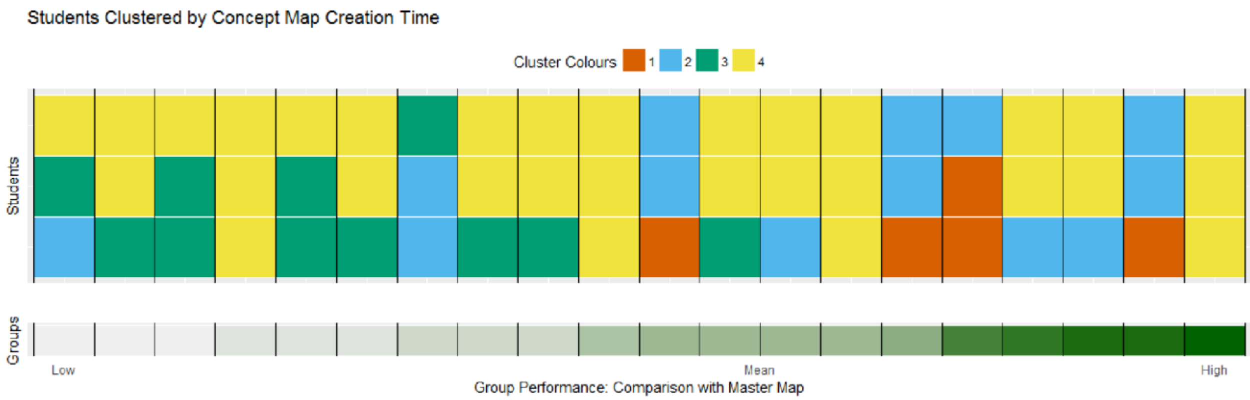

Figure 8: Students in their groups, coloured by their profiles and ordered by comparison with master map

We can observe the following from Figure 8 and Figure 9:

1. Students who spent more time adding concepts, Cluster 3 (green), had performed poorly when compared with Master Map, probably because they did not initially use the concepts available, and created concepts with names different from the master map. At the group phase, they may convince others to use the new names.
2. Students who spent more time moving elements, Concepts for Clusters 1 (red) and Linking Phrases for Cluster 2 (blue), had a good interaction working together regarding the Spada score and when compared with the Master Map.
3. Cluster 4 (yellow), when analysed against the Spada score are in both the left and right ends of the chart.

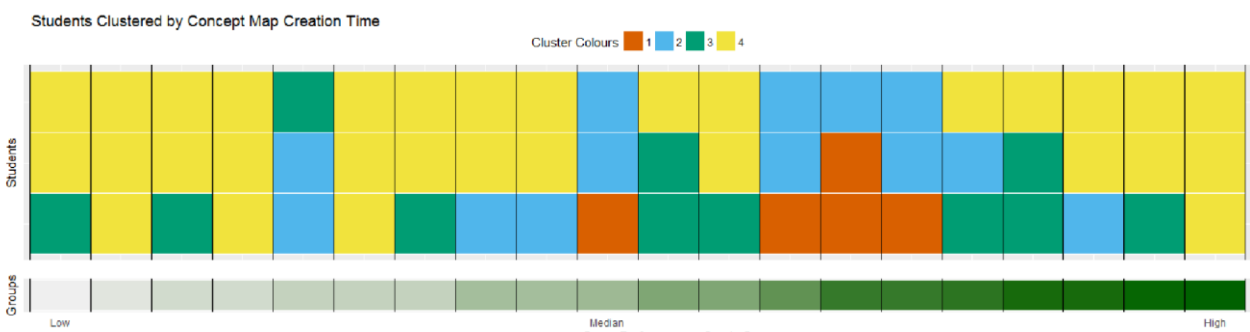

Figure 9: Students in their groups, coloured by their profiles and ordered by Spada score

Discussion and Conclusion

Collaborative learning raises multiple challenges for learning sciences and related fields. An important aspect of a successful collaborative learning experience is to get students grouped in a way that may foster, hopefully maximise, this learning experience. The way groups are formed in classrooms, blended or online environments requires a careful learning design process. Although there are some tools that automate the group formation phase, it is up to the teacher to choose the criteria for arranging which students should work together. This choice requires a deep understanding of the task, the students, and the desired outcome, which is not always the case because of lack of time, number of students and resources available.

Because of the above, understanding collaborative learning is a complex problem constrained by multiple variables, such as the number of students in the groups, time of collaboration, the nature of the task, and the environment in which the task is being executed. So far, the issue of group formation has been done either following social theories, or random or systematic processes. Our work aims to harness the data collected by collaborative educational technologies and empower teachers to explore all aspects of these collaborative learning processes so that they can make informed, data-driven decisions to support collaboration. This paper is a first step for understanding how different student profiles interact together when doing a collaborative work. Profiles were extracted from learning systems data and using clustering algorithms. Student interactions were measured by different group performance metrics. The charts presented are a first attempt to equip teachers with tools to explore what different profiles of students exist in their cohorts and how to link them to group performance.

The student profile clustering is an unsupervised attempt to find patterns within students' profiles in terms of their behaviour using the learning systems, their performance or demographics profiles. The boxplot chart intends to contrast the difference between these patterns so it can be easily perceived by the teacher. Similarly, the tile chart has the same purpose but targets the group patterns. The tile chart presents all the groups and their arrangements so it may give insights about why some groups performed better than others did. It also provides the ability to perform a novelty visual groupwise comparison, expanding more traditional comparisons made just individually between students (e.g. Verbert et al. (2014)).

As further investigations, our focus will be to evaluate the visualisation tools presented in this paper and to collect feedback from teachers. We will also consider improvements to the design of the interface, especially to represent larger cohorts of students. As a wider contribution, we will publicly release the code that generates the charts with its pertinent documentation, together with sample datasets to allow other researchers to propose different methods of analysis and visualisations for collaborative scenarios.

References

Beal, D. J., Cohen, R. R., Burke, M. J., & McLendon, C. L. (2003). Cohesion and performance in groups: a metaanalytic clarification of construct relations. *Journal of applied psychology, 88*(6), 989.

Bovo, A., Sanchez, S., Héguy, O., & Duthen, Y. (2013). *Analysis of students clustering results based on Moodle log data.* Paper presented at the 6th International Conference on Educational Data Mining-EDM 2013.

Carvalho, L., & Goodyear, P. (2014). *The architecture of productive learning networks*: Routledge.

Choi, H., Lee, J. E., Hong, W.-j., Lee, K., Recker, M., & Walker, A. (2016). *Exploring Learning Management System Interaction Data: Combining Data-driven and Theory-driven Approaches.* Paper presented at the Proceedings of the 9th International Conference on Educational Data Mining.

Cohen, E. G., Goodlad, J. I., Darling-Hammond, L., & Lotan, R. A. (2014). *Designing Groupwork : Strategies for the Heterogeneous Classroom, a New Edition* (Vol. Third edition). New York, NY: Teachers College Press.

Craig, M., Horton, D., & Pitt, F. (2010). *Forming reasonably optimal groups:(FROG).* Paper presented at the Proceedings of the 16th ACM international conference on Supporting group work.

Demetriadis, S., & Karakostas, A. (2008). *Adaptive collaboration scripting: A conceptual framework and a design case study.* Paper presented at the Complex, Intelligent and Software Intensive Systems, 2008. CISIS 2008. International Conference on.

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

Gramoli, V., Charleston, M., Jeffries, B., Koprinska, I., McGrane, M., Radu, A., . . . Yacef, K. (2016). *Mining autograding data in computer science education.* Paper presented at the Proceedings of the Australasian Computer Science Week Multiconference.

Kardan, S., Roll, I., & Conati, C. (2014). *The usefulness of log based clustering in a complex simulation environment.* Paper presented at the International Conference on Intelligent Tutoring Systems.
Konert, J., Burlak, D., & Steinmetz, R. (2014). The Group Formation Problem: An Algorithmic Approach to Learning Group Formation *Open Learning and Teaching in Educational Communities* (pp. 221-234): Springer.
Koprinska, I., Stretton, J., & Yacef, K. (2015). *Students at risk: Detection and remediation.* Paper presented at the Proceedings of the 8th International Conference on Educational Data Mining.
Martinez-Maldonado, R., Dimitriadis, Y., Martinez-Monés, A., Kay, J., & Yacef, K. (2013). Capturing and analyzing verbal and physical collaborative learning interactions at an enriched interactive tabletop. *International Journal of Computer-Supported Collaborative Learning, 8*(4), 455-485.
Martinez, R., Kay, J., & Yacef, K. (2011). *Visualisations for longitudinal participation, contribution and progress of a collaborative task at the tabletop.* Paper presented at the Computer supported collaborative learning, CSCL.
McBroom, J., Jeffries, B., Koprinska, I., & Yacef, K. (2016). *Mining behaviours of students in autograding submission system logs.* Paper presented at the Proceedings of the 9th International Conference on Educational Data Mining.
Rojas, I. G., & García, R. M. C. (2012, 4-6 July 2012). *Towards Efficient Provision of Feedback Supported by Learning Analytics.* Paper presented at the 2012 IEEE 12th International Conference on Advanced Learning Technologies.
Siemens, G., & d Baker, R. S. (2012). *Learning analytics and educational data mining: towards communication and collaboration.* Paper presented at the Proceedings of the 2nd international conference on learning analytics and knowledge.
Sinha, T. (2014). *Together we stand, Together we fall, Together we win: Dynamic team formation in massive open online courses.* Paper presented at the Applications of Digital Information and Web Technologies (ICADIWT), 2014 Fifth International Conference on the.
Stahl, G. (2006). *Group Cognition: Computer Support for Building Collaborative Knowledge (Acting with Technology)*: The MIT Press.
Tukey, J. W. (1977). Exploratory data analysis.
Verbert, K., Govaerts, S., Duval, E., Santos, J. L., Van Assche, F., Parra, G., & Klerkx, J. (2014). Learning dashboards: an overview and future research opportunities. *Personal and Ubiquitous Computing, 18*(6), 1499-1514. doi:10.1007/s00779-013-0751-2

Please cite as: dos Santos, A.D.P., Yacef, K. & Martinez-Maldonado, R. (2016). Visualizing individual profiles and grouping conditions in collaborative learning activities. In S. Barker, S. Dawson, A. Pardo, & C. Colvin (Eds.), *Show Me The Learning*. *Proceedings ASCILITE 2016 Adelaide* (pp. 545-554).

Note: All published papers are refereed, having undergone a double-blind peer-review process.