Journal of Computer Assisted Learning

Using Different Approaches in Multimodal Learning Analytics for Estimating Success in Project-based Learning

Journal:	*Journal of Computer Assisted Learning*
Manuscript ID	JCAL-17-148.R1
Manuscript Type:	Special Issue Article
Technology and Tools:	Shared Workspaces, Intelligent systems, Information systems, Handheld/Mobile devices
Learning process / Pedagogy:	Teams (including dyads and triads), Experiential Learning, Collaborative Learning
Paradigm:	Constructionism
Level of education :	Undergraduate
Place of learning:	Formal Learning
Type of research:	Quantitative
Research technique:	Design Based Research, Experimental
Analysis/evaluation paradigm:	Video Analysis, PLS Regression
Issues:	Knowledge, Skills

Using Different Approaches in Multimodal Learning Analytics for Estimating Success in Project-based Learning

Daniel Spikol

Malmö University, Sweden

Emanuele Ruffuldi

Scuola Superiore Sant'Anna, Italy

Giacomo Dabisias

Scuola Superiore Sant'Anna, Italy

Mutlu Cukurova

University College London, United Kingdom

Author Note

Daniel Spikol, Faculty of Computer Science and Media Technology, Malmö University.

Correspondence concerning this article should be addressed to Daniel Spikol, MAH, Nordenskioldsgatan 1, Malmö, Sweden, 211-19 E-mail: daniel.spikol@mah.se

Abstract

Multimodal learning analytics provides researchers new tools and techniques to capture different types of data from complex learning activities in dynamic learning environments. This paper investigates high-fidelity synchronised multimodal recordings of small groups of learners interacting from diverse sensors that include computer vision, user generated content, and data from the learning objects (physical computing components). We processed and extracted different aspects of the students' interactions to answer the following question: which features of student group work are good predictors of team success in open-ended tasks with physical computing? To answer this question, we have explored different supervised machine learning approaches (traditional and deep learning techniques) to analyse the data coming from multiple sources. The results illustrate that state-of-the-art computational techniques can be used to generate insights into the "black box" of learning in students' project-based activities. The features identified from the analysis show that distance between learners' hands and faces is a strong predictor of students' artefact quality which can indicate the value of student collaboration. Our research shows that new and promising approaches such as neural networks as well as more traditional regression approaches can both be used to classify MMLA data, and both have advantages and disadvantages depending on the research questions and contexts being investigated. The work presented here is a significant contribution towards developing techniques to automatically identify the key aspects of students success in project-based learning environments, and ultimately help teachers provide appropriate and timely support to students in these fundamental aspects.

Keywords: Multimodal Learning Analytics, Project-based Learning, Machine Learning

Using Different Approaches in Multimodal Learning Analytics for Estimating Success in Project-based Learning

Over the last several years the field of learning analytics (LA) has grown rapidly in conjunction with Massive Open Online Courses (MOOCs) and other technology systems. These systems include but are not limited to virtual learning environments, mobile applications, and student-response systems which are rapidly becoming part of the everyday educational landscape. These systems collect and provide diverse types of data about learners' interactions that take place both with the systems and among learners, allowing, overall, new insights into education. Such systems often highlight the importance of big data in education that is of the interest of diverse actors for utilising learning analytics for educational management and policy making (Clow, 2013). However, from a learning sciences research perspective, the aim of learning analytics is to understand and optimise the learning process and most learning happens outside of these systems between people in face-to-face situations (Greller & Drachsler, 2012; Siemens & Baker, 2012). In this research paper, we investigate MultiModal Learning Analytics (MMLA) to make sense of students' learning process in project-based learning activities with the purpose of optimising it for students and teachers.

Project-based learning activities have the potential to help educators to achieve high tier institutional and policy goals such as developing 21st century skills in Science, Technology, Engineering, and Mathematics (STEM) subjects. More specifically for teaching Technology subjects, such as computer science and ICT, project-based learning is a commonly employed approach and its popularity is increasing particularly after the introduction of the "Makers Movement". Most of these project-based approaches involve learning activities that combine hands-on computing technologies to explore various topics in both secondary and post-secondary learning institutions (Halverson & Sheridan, 2014). However, these hands-on activities introduce many challenges due to their dynamic and multifaceted nature, specifically regarding their design, implementation, and evaluation.

Looking at the existing evidence, it becomes very clear that students do not become effective learners when they are left on their own within such "student-led" learning environments (Kirschner & van Merriënboer, 2013). Therefore, students' appropriate monitoring and guidance in these pedagogical approaches is an essential requirement for their success. Nevertheless, due to practical challenges of project-based learning including the fact that teachers lack the required time and resources to attend and support each student group (or each student within groups) during their engagement with the projects, these type of project-based learning approaches often struggle to satisfy their common learning outcomes.

However, MMLA offers researchers new tools to capture different types of data from complex learning activities including project-based learning. The ability to collect multimodal data from bodily movements, face tracking, affective sensors, hardware and software log files, user and research generated data, provide opportunities to obtain unique features which can be interpreted to understand and appropriately support project-based learning. The multimodal data from these sensors provides new opportunities for investigating learning activities in the real-world between small groups of learners working on tasks with physical objects (Blikstein & Worsley, 2016). The automated collection and presentation of insights from MMLA to support project-based learning approaches is an exciting emerging field with the learning analytics domain and it has the potential to provide the required support for students and teachers involved in project-based learning approaches to help them achieve their learning outcomes.

Starting from the initial assessment conducted by the authors (Spikol, Cukurova, & Ruffaldi, 2017), in this paper, we investigated how MMLA data can be used to support project-based learning from a specially designed worktable environment where small groups of students use new physical computing components to solve open-ended tasks. In order to achieve this, we built a multimodal learning analytics system that is part of the students' project worktable and collected diverse streams of data. We processed and extracted multimodal interactions to answer the following question: *which features of students' group work that can be automatically collected with our MMLA system, are good predictors of students' project outcomes in open-ended learning activities with physical computing?* In order to answer this question we have explored different supervised machine learning approaches employing traditional and deep learning (DL) techniques to analyse the data coming from multiple sources. Our work is a significant contribution towards providing ways to automatically identify the key aspects of students success in project-based learning environments and ultimately help teachers provide appropriate and timely support to students in these key aspects.

The paper is structured as an experimental design work: first, we present the background, then the system context, then the material and methods that included the design of the intervention, followed by results, discussion and conclusions.

Background

The roots of project-based learning extends back almost a century to John Dewey's approach that argues for "laboratory schools" in which students are engaged with the process of inquiry in their learning activities (Dewey, 1959). The history of this approach is rich, and a detailed literature review of the approach is outside the scope of this paper. However, it is important to define the concept and explain its main features. Project-based learning is a form of situated learning, in which students engage in real-world activities that are similar to the activities that professionals engage in (Krajcik & Blumenfeld, 2006). Project-based learning activities that support learners' participation in open-ended tasks are one of the most commonly used teaching approaches for improving 21st century skills (Bell, 2010) and they emphasise the engagement of learners in projects that are personally meaningful and they encompass driving questions, investigations, and collaboration (Krajcik, 2010). However, the hands-on and open-ended nature of project-based learning creates challenges for tracking the learning process. One of the key challenges faced in project-based work is the support of the group work and ensuring that students succeed in the planned learning outcomes (Blumenfeld et al., 1991; Krajcik & Blumenfeld, 2006).

Current research in MMLA focuses on better understanding the complexity of learning through the advances of high-frequency multimodal data capture, signal processing, and machine learning techniques (Ochoa & Worsley, 2016). MMLA offers an opportunity to capture different insights about learning in project-based learning tasks in which students have the opportunity to generate unique artifacts like computer programs, robots, and small-groups collaboration to solve open-ended tasks (Blikstein, 2013; Blikstein & Worsley, 2016). MMLA builds upon multimodal human interaction, educational data mining, and many other fields that include learning sciences and cognitive sciences to capture the complexity of learning through data intensive approaches (Siemens & Baker, 2012; Worsley, 2012).

In terms of the focus on purposes and context, there is an emerging body of work with in the field of MMLA to capture small group work on project-based learning that has grown mainly out of the work of Blikstein and Worsley investigating engineering students'design activities (Blikstein, 2013; Chen et al., 2014; Ochoa et al., 2013). Within this research domain, Blikstein (2011) explored multimodal techniques for capturing code snapshots to investigate students learning computer programming as well as video and gesture tracking for engineering tasks; Worsley (2014) presented different approaches for data classification that included points about how these techniques have a significant impact on the relation of research and learning theories. Both of these initial approaches provided the means for other researchers to begin to explore MMLA with small groups of students across different subjects. In addition, notable data sets from the MMLA grand challenges workshop Ochoa and colleagues (2013), presented the Math Data and Oral Presentation Quality Data Corpora that has enabled the community to analyse and discuss the different requirements and results within this field. Moreover, Ochoa and colleagues' work (2014) used existing multimedia processing technologies to produce a set of features for accurate predictions of experts in groups of students solving math problems which illustrated the benefits of MMLA to support students' learning in these contexts. Similarly, Chen and colleagues (2014) expanded from the Oral Presentation Quality Data corpus to further examine the feasibility of using multimodal technologies for the assessment of public speaking skills; and Grover and colleagues (2016) have explored how to develop computational models of social learning environments. In their work Grover and colleagues managed to classify the quality of collaboration from body movement and gestures of pair programmers working together with acceptable accuracy rates. Although most of the existing MMLA research approaches focus on learners' data, Prieto and colleagues (2016) and Martinez-Maldonado and colleagues (2016) have focused their research efforts on how MMLA can support teaching actions and orchestration in the classroom. On the other hand, regarding the technical focus, to make sense of complex data streams coming from multiple data sources, MMLA researchers employ various computational techniques. These approaches include logistic regressions (Ochoa et al., 2013), different feature reduction algorithms (Schneider & Blikstein, 2014; Worsley, 2014), and statistical models to investigate MMLA to identify features and predict student performances (Schneider & Blikstein, 2014). These approaches all have advantages and disadvantages depending on the main research question and the purposes of data analysis and have potential to provide insights how to proceed with a multimodal data-set. Regardless of which computational approach is taken, it is clear to us drawing from the literature that MMLA has a role to play to support education in project-based learning approaches, and it has the potential to provide new means for gathering insights for complex, open-ended learning activities (Blikstein & Worsley, 2016) which otherwise are extremely challenging to monitor and support with existing traditional standardised evaluation approaches.

System Context

The work discussed in this paper is based on the European project "Practice-based Experiential Learning Analytics Research And Support" (PELARS)[1]. The central goal of the project was to develop learning analytics tools for hands-on, open-ended STEM and STEAM project-based learning activities using physical computing. The learning contexts we have investigated are high schools, engineering, and design departments at universities. The current system includes customised furniture with an integrated Multimodal Learning Analytics System (LAS) such as tracking hands, faces and other objects and the Arduino platform with a visual web-based Integrated Development Environment (IDE) that captures interaction information of physical computing. The learners and observers use mobile devices to capture multimedia data (text, images, and video) to self-document the learning activities.

Overall, the PELARS project has developed an intelligent system for collecting activity data (LAS) for diverse learning analytics (with data-mining, reasoning, and visualisations) and active user-generated material and digital content (that include mobile tools and physical computing platform) for project-based learning activities (Cukurova, Avramides, Spikol, Luckin, & Mavrikis, 2016; Spikol, Ehrenberg, Cuartielles, & Zbick, 2015). See examples of the PELARS system in action with university engineering students in figure 1.

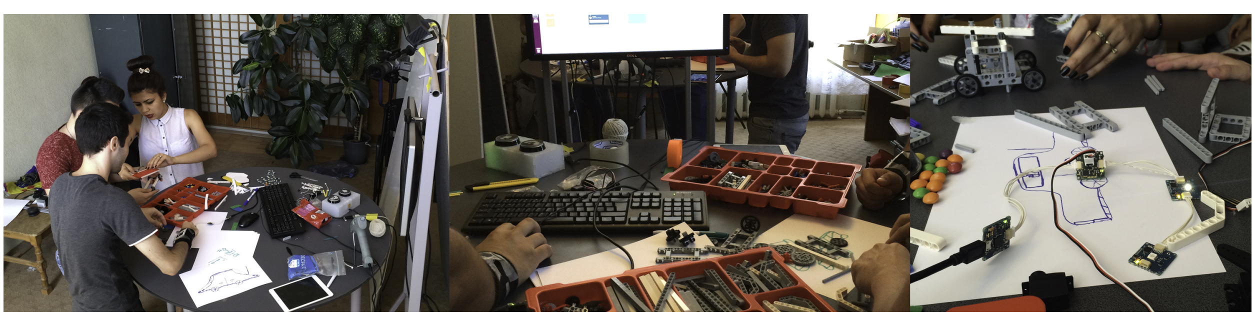

Figure 1. University engineering students working in the PELARS environment.

[1]http://www.pelars.eu

PELARS LAS

The LAS collects multimodal data from different sensors and input from the learners and researchers. The learning environment is designed to foster collaboration and includes an integrated screen and standing round table to allow learners to share and work together. The LAS collects data from both ambient (sensors) and live sources (human interaction). The ambient collection of data includes a computer vision system that uses color and depth cameras with audio for understanding how people interact around the workstation furniture. The LAS uses a Web-based architecture in which a classroom located data collector performs data acquisition and vision processing sending data to a remote server using WebSockets. The system has been designed to work in offline mode allowing to later synchronize the content on the remote server. The data on the server is further processed for extracting learning analytics and statistics. For details about the architecture please refer to Ruffaldi, Dabisias, Landolfi, and Spikol (2016).

Physical computing

A core part of the system are small Arduino based boards that play a fundamental role in the project-based activity of the students. These boards are using the TALKOO IDE. This IDE has been designed to allow users to start building electronic devices without having to build circuits neither on breadboards nor prototyping boards and without having to write complex lines of code (Katterfeldt, Cuartielles, Spikol, & Ehrenberg, 2016). The visual programming interface is a web tool (HTML5 based) to the standard Arduino IDE. This platform has been developed for the project with plug-and-play sensors and actuators together with a flow-based visual programming IDE that allows learners to prototype artefacts rapidly. A set of “sentiment / affective” buttons has also been developed with thundercloud and sunshine icons to allow the students to mark critical events in their activities.

Mobile tools

The set of mobile tools has been developed to provide the means for the learners to self-document the learning process across planning, building, and reflection phases on their projects with different content and multimedia data. Also, it allows researchers and teachers to mark critical incidents, and researchers to time stamp the different stages of the learners' project. The tool is developed based on modern web technologies which run across different platforms (Zbick, Vogel, Spikol, Jansen, & Milrad, 2016).

Collected Data

The automatically collected data includes the capture of objects, the positions of people, hand movements, faces and audio levels and video as well as interactions of plugged components from the Arduino-based physical computing platform and the interaction with the sentiment buttons. Instead the mobile based tool allows to gather self-documentation annotations from students, and progress annotations from researchers or teachers looking at students. In particular, in the experimental settings employed in this work, the researchers have annotated the activity cycle marking the phases of planning, building and reflecting. PELARS Arduino blocks, sentiment buttons, and students working in Figure 2.

Materials and Methods

The automatic approach discussed in this paper is performed over a data set acquired with engineering students with the PELARS platform. In this section the data acquisition processes are discussed, with the analysis performed based with machine learning classification.

Datataset Acquisition

The data analysed in this paper is from 3 sequential educational interventions with 18 engineering students at an European university (17 men and 1 woman, average age 20

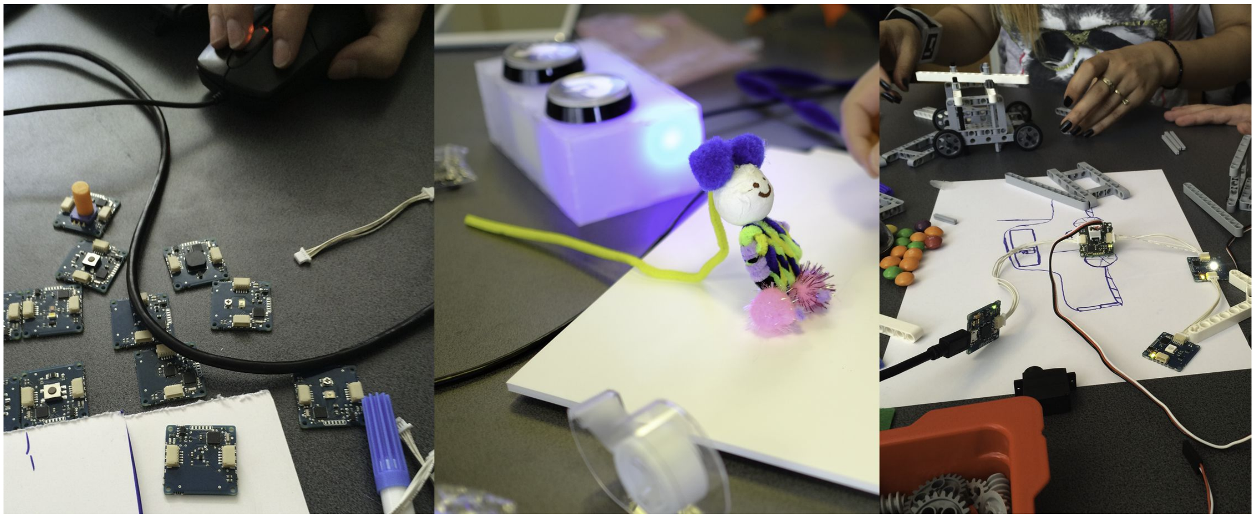

Figure 2. Details of the Arduino based boards, sentiment buttons, and diverse student projects.

years old). The students were divided into 6 groups made up of 3 students. Each student group used the system over 3 days completing one open-ended design tasks for each session. First, the students were introduced to the system with a workshop to familiarise them with it, and then their first task was to prototype an interactive toy. The second task was the prototyping a colour sorter machine, and in the third task the students have been asked to build an autonomous automobile. Each of these design sessions ranged from 33 to 73 minutes. As can be seen, each of the tasks introduced a more complex design concept to be solved with respect to the previous ones. Students were asked to perform an initial phase of planning, followed by execution/building and finally a documentation/reflection phase. During the activity the students had to document their planning, building, and reflecting phase through a mobile tablet. The tablet allowed the students to take photographs, record video, and report via a form and free text their plan, progress, and reflective thoughts. No specific instructions about the timing of these phases were given to students. Additionally, the research observers used the mobile tool to divide the students work flow into the planning, building, and reflecting phases.

Initial Classification of Students' Project Outcomes

To grade the students' design projects, a scoring scheme was developed that combined different approaches for collaborative problem solving (CPS) in small groups as well as bringing the design thinking principles. We started with the seminal work done with engineering students (Atman et al., 2007) that was initially adopted by (Worsley & Blikstein, 2014) for multimodal learning analytics. From these initial frameworks, we began to develop a framework for CPS (Cukurova et al., 2016) that we could apply for the PELARS context. We used a version of our CPS framework with the mobile system with an agreed set of codes for on-fly observations to initially grading of the students' projects. From the initial score of the students' work, the team of researchers reviewed the students work collected in the LAS which included snapshots of the students' plan, video of solutions, and learners text input. The 18 session were graded with these criteria, where 50% of the grade was the expert's opinion based on the documentation collected by students, 25% was how the students planned and delivered the artifact, and the remaining 25% was the students' own self-assessment of the quality of their projects. The resulting scores were categorised in three classes: **poor, ok, and good**. This classification of the sessions was used as the reference point for the previous machine learning based classification work (Spikol et al., 2017) in which the nature of this evaluation allowed only to reliably classify the works in two classes: good and poor.

Improved Classification of Students' Project Outcomes

Based on the on the binary approach present in the previous scoring each of sessions has been re-evaluated and re-scored by experts looking at videos, documentation (from the mobile tools) and final project outcome (the artefact). The aim was to generate a more rich scoring that reflected the learning practices for engineering courses. The new scoring has been based on 5 different aspects expressed in a scale from 1 to 5:

- Level of Clarity [Loc] (5=very clear, 3=legible, 1=not understandable)

- Independent Thinking [InTh] (5=independent, 3=based off instruction, 1=same as instruction)

- Corresponds with plan [CorPI] (5=Fully, 3=partially, 1=not at all)

- Does it Work? [DoWo] (5=fully, 3=partially, 1=not working)

- Quality of solution [QuaOS] (5=great, 3=mediocre, 1=poor)

Table 1 presents all the scores while Figure 3 shows the quality score for the 6 teams.

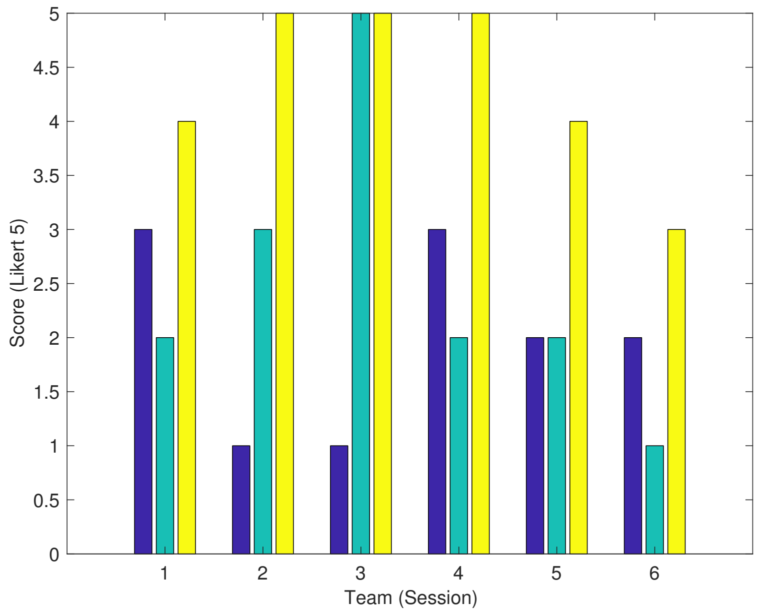

Figure 3. Quality of solution scores (QuaOS) of each team during the three sessions. The blue bars are session 1, green session 2, and yellow session 3.

Acquired Data: The Investigated Features of the LAS

For each sessions recorded, the LAS system collected data from the students comprising activity performed, user generated content (text and multimedia) and actions on the Arduino visual IDE. In particular, the following data was acquired.

Face Tracking. By means of a frontal camera (Logitech C920, 960x540 resolution at 30Hz) and the Viola-Jones algorithm (Viola & Jones, 2004) all the visible faces of students were tracked, and, through camera calibration and assumptions about face size, it was possible to estimate the 3D position of students's head with respect to the camera. Thanks to per-session calibration between the cameras and a fixed point on the table it is possible to express the pose of student faces in a 3D reference frame of the table. From the face tracking data two metrics have been identified: the the count of Faces Looking at the Screen (FLS) and the distance between the faces which provides an indicator of Distance between Learners (DBL). We hypothesise that the measure DBL may be a proxy of collaboration, since students' physical proxy is a required but not sufficient condition for students collaboration. We expect that when the DBL is small it is more likely that the collaboration would occur among students and there is enough evidence that collaboration has potential to improve students' learning outcomes.

The adopted algorithm is quite robust to facial differences and illumination conditions although it is primarily designed for frontal faces. Additional detectors are available for lateral faces. For compensating sudden motions we interpolated pose information when the face was not detected for a short period of time.

Hand Tracking. The top down color with depth camera (Microsoft Kinect One, 1920x1080 resolution at 30Hz) monitored the motion of the hands of the students that were wearing fiducial markers (Munoz-Salinas, 2012) that disambiguate each primary hand. The pose is estimated by combining the image based marker tracking with the depth information. Again, thanks to the calibration of the camera and the size of the markers the 3D position of the hands was obtained with respect to the Table. Based on the 3D position of the hands we were able to calculate two metrics: the Distance between Hands (DBH) and the Hand Motion Speed (HMS).

In terms of tracking capabilities wristbands with fiducial markers provide precise information when the marker is visible and with a non-lateral orientation. In comparison to markerless trackers this solution is robust to object handling, although research is progressing well in this direction thanks to Deep Learning (Sridhar et al., 2016).

Arduino IDE. The interface between the Visual Arduino IDE and the data collection system provided information about the types of physical and software blocks used in the project and their connections. In particular we counted the number of Active Blocks (IDEC), the Variety of Hardware (IDEVHW) and Software Blocks used (IDEVSW) and the number of interconnections between blocks as a Measure of Complexity (IDEX) in students' programming during their project-based activities.

Audio Level. By means of the microphone included in one of the cameras and Fast Fourier Transformation (FFT) we computed the sound level during the sessions. The resulting feature was a value sampled at 2Hz called Audio Level (AUD).

Pre-processing

From all these MMLA data points the data was collected at variable data rates (around 2Hz), yet it was not synchronised. For this reason, we needed a processing stage that aggregates indicators from the different variables in windows of same duration. The aggregation was performed based on counting for most of the variables. However, only for the distance/proximity features we employed averaging. Considering the fact that, students sessions were different in terms of their lengths due to the open-ended nature of the project-based learning activities, we employed zero padding for sessions that were too short.

Machine Learning Approach

A supervised machine learning approach has been employed for associating the measured student actions with the resulting scores by the experts. In particular we have performed a two stage approach with different techniques. One assessment is based on large data quantities and uses Deep Learning for regressing the 6 scores by the experts. The second, based on traditional machine learning, deals with the simpler 3-levels assessment of the sessions and tries to address the problem of explaining the causes of the outcome depending on measured features and phases. Table 2 shows a synthetic view of the two tasks together with the inputs, outputs and details about the algorithms as discussed in the rest of this section.

Deep Learning Regression of Outcome. Deep learning has been tested to check the feasibility of non-linear regression on the input data gathered from the sensors. A deep neural network (DNN) is composed by a graph of linear matrix multiplications which are followed after each stage by a non linear function called *activation function*. The general behavior can be synthesized as follows: given an input vector x, a series of matrices A_i composed of weights $w_(k, j)$, a bias vector b, an activation function F and an output Y_i, it is possible to write stage i as:

$$Y_i = F(A_i x + b).$$

The output Y_i will then be the input of the next stage of the pipeline, until reaching the end of it, where a classifier or regressor computes the final output. DNNs can be used for classification or regression: in the first case the network is trained to obtain a label indicating the category to which the input belongs, while in the latter case the network learns to fit an unknown function using the input and output data in order to estimate points which are not present in the input set. For the purpose of this experiment regression has been used since the output values can be a set of continuous values.

The input data is a set of timeseries that have different data rates and partial synchronization. In this work we decided to use a windowing approach with dense network for compensating such difference leaving the use of recurrent neural network techniques for future work. Given a session of duration T seconds we split it into non-overlapping windows of length L seconds (120,240 and 360) obtaining $\lceil T/L \rceil$ windows. For a given input we compute an aggregated statistics for each window (averaging or summation). Each window is sent separately as input to the NN. The following aggregated statistics (18 values in total) have been employed:

- Total number of faces looking toward the screen

- Total number of connected Arduino components
- Mean distance between faces (DBF)
- Mean distance between hands (DBH)
- Mean hand movement speed (HMS)
- Mean audio level (AUD)
- Mean hand positions (HP)
- Mean Arduino components activity

Given this, the network has been trained to fit a function which has an 18 dimensional domain and a 6 dimensional co-domain. Several additional network parameters have been tuned to obtain the best possible solution along with the window size for the input data creation. These parameters include: (1) Dropout, (2) Regularization method, (3) Epochs, (4) Layers.

Input data is randomly split, as usual, in training and test data, with a minor split of the training data again into training and validation. In these experiments 20% of the sessions are removed as test sessions leaving 80% for training. Of this 80% another 20% has been used for validation set during the training phase. It is important to notice that complete sessions have been left out for testing and not just random inputs (windows) since they are usually correlated and could alter the final results if used. The results of the network are evaluated using a mean squared error distance between the predicted value vector and the true value vector obtained in the test data set. A mean squared error has also been computed for each of the six output values along with the variance in order to understand if any of the output values had a different behavior. Six different DNN architectures have been tested, growing from one to six fully connected layers starting with a size of 1024 and decreasing at each layer by half. The best obtained network was created using the following parameters: Dropout 0.5, No regularization, 100 Epochs, 3 Dense Layers (1024,512,256) and 240 seconds window size. The network structure can be see in Figure 4.

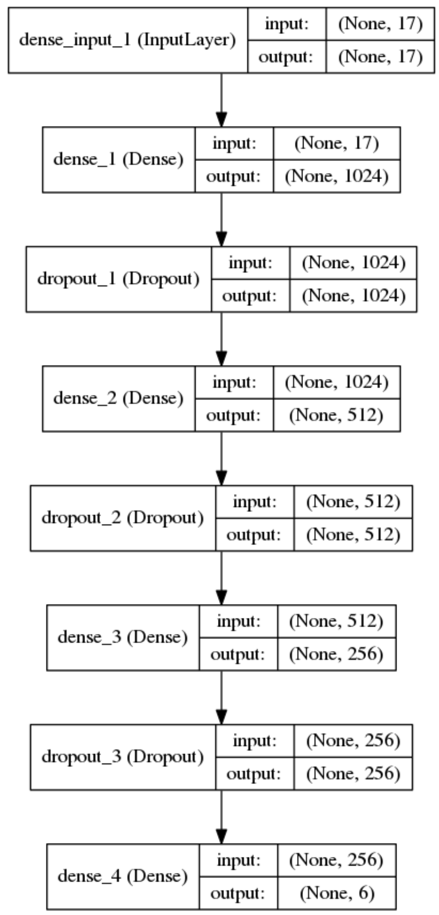

Figure 4. Neural Network structure of the model which obtained the best results

The network has been implemented using a Python library for deep neural networks called *Keras* (Chollet, n.d.). This high-level library allows to abstract the use of the GPU optimized processing libraries *Tensorflow* (Abadi et al., 2016) and *Theano* (Bergstra et al., 2010).

Outcome classification. At this stage we performed a supervised classification task that matches the observers' scores. The purpose of this approach is to identify the data features that can support different score classifications that have been evaluated by human observers (experienced teachers) as poor, ok, and good. Among the different families of classifiers available, we tested various parametric ones namely Naive Bayesian (NB), Logistic Regression (LR) and Support Vector Machines with linear (SVML) and Gaussian kernel (SVMR). We avoided the non-parametric ones (Nearest neighbours) or decision trees with the purpose of reducing the overfitting effect. In particular the Naive Bayesian is a simple classifier that employs a strong assumption about features, a condition that holds valid for most of the variables we employed in our investigation except for the ones related to the Arduino IDE. We decided not to use the ensemble of classifiers (Kotsiantis, Patriarcheas, & Xenos, 2010), as we would like to study the model behind these classifications as much as performing the classification itself. For this classification task time has been considered by using larger windows of size (10, 20, 30 minutes and whole session) aggregating the data similarly with the previous approach but considering the values from all the windows together and padding the sessions with a smaller number of windows.

We used cross-validation (k=4) for understanding the effect of different parameters such as window size and the inclusion of different phases. Due to the small sample size (18 sessions from 18 Engineering students working in 6 groups of 3 students), we avoided the leave-one-out scheme. The data acquired from the PELARS LAS was exported and then processed in Python using the sklearn (Pedregosa et al., 2011) toolkit that provides state-of-the-art machine learning techniques integrated with a common interface. The test of the classifiers was performed by varying the window size, the score (binary or original 3-level), the inclusion of the different phases (planning, building, and reflecting) and, most importantly, the effect of features identified and described above (FLS, DBL, DBH, HMS, IDEC, IDEVHW, IDEVSW, IDEX, AUD).

Results

The aim of this paper has been to investigate different machine learning approaches to estimate success in small group work through multimodal learning analytics. We compare the results of Deep Learning techniques against traditional Supervised Learning techniques to explore the performance and give insights on these techniques to support research in learning analytics.

Deep Learning Regression of Outcome Results

The overall results for the different network structures are illustrated in table 6.Table 3, 4 and 5 show the mean and variance for the error between expected output and predicted value. We then compared 120s, 240s and 360s window sizes and the 240s network achieves a mean squared error of 0.13 as shown in table 4 across the improved classification of the students' outcomes. We then investigated the different features by removing them individually. In general, the results get worse as expected, however in the case of distance between faces DBF, see table 7. This result illustrated that this feature of distance between faces is a substantial input for project-based work in the PELARS context. Additionally, the results show that the smallest window performs worse than the others, see table 3. The network achieving the best results is shown in Figure. 4 and is using a window size of 240s.

Outcome Classification Results

Phases. Although, we had a small sample size of 18 sessions, the total amount of data generated from these sessions was rich and large due to the multimodal nature of our investigation. The project-based learning activities lasted within the range of 33 minutes to 75 minutes (median 63 min$\pm$13) with a total activity time of 17 hours and 10 minutes. Each project-based learning activity's project outcome was graded based on the criteria described earlier and different patterns along the three sessions were observed.

The design phases annotated by the observer (planning, building, and reflecting) varied broadly among the sessions as well as among the groups. The mean scores for the time spent on these phases among the sessions are planning (11min$\pm$10min), building (41min$\pm$16min) and reflection (4min$\pm$7min). Figure 5 shows the duration of each session and the timing of the phases for different groups of students. Below the results are being presented in Figure 5.

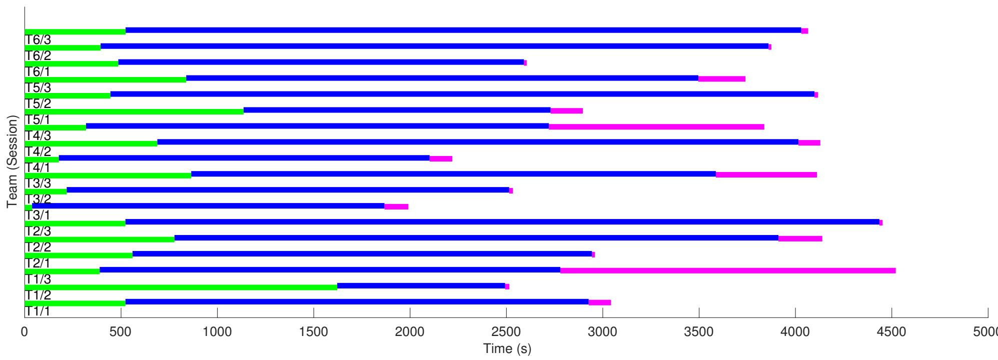

Figure 5. Distribution of phases among session of the 6 teams. Each session is split in the three phases, first plan (green), then build (blue) and finally reflect (purple).

Scoring. The three-level scoring we initially identified using human observation (poor, ok, good) posed difficulties to the classification activity and we needed to move to a binary version in which we aggregated ok graded groups with good graded groups. For example, NB and SVM classifiers score 0.8 and 0.75, respectively with a window of 30min and binary classification, however this value decreases to 0.5 for both of the classifiers when we use a three-way classification. This situation is clearly not ideal, however in order to achieve adequate results we took this binary approach which still has great value to be able to identify project-based learning groups who perform poorly from others. Alternatively, it can be used to identify those group performances that are considered as good from the rest in a binary fashion. We see this as the first step towards further more detailed classifications.

Effect of Phase. Across the different conditions, the selection of the phases used to train a strong effect of the capacity to recognise the classifiers. For example, with a 30min window and binary classification, the exclusion of reflection (PW) phase in student activities, provided stronger results across the different classifiers, while the exclusion of both planning and reflection reduced the classification power. Please note that the decision to omit reflection phase from data was taken due to statistical arguments. This decision does not reflect our lack of interest in the reflection stage. We think that reflection is an important phase of learning and would like to improve our algorithms in the future with further data collection to be able to generate meaningful results with all significant phases. See Table 8 for the details.

In order to provide the most reliable results and use the strongest classification power, we focus our results on data collected from the planning and working stages of the student activities and excluding the reflecting stages.

Type of Classifiers. As can be seen in table 1, across the different tests of the classifiers, those behaved the most consistently were Naive Bayesian (NB) and the Support Vector Machines with linear kernel (SVML).

Effect of Features. Having established the window size as 30 mins, grade classifications as poor vs. ok plus good scored projects, learning activity stages as planning and building phases, and the statistical methods we will use as NB and SVML, we now present the results of our analysis on the effects of the multimodal learning analytics features. We start from the full set of features with a given selection of the other parameters mentioned above and we proceed removing features, as a form of model selection.

Regarding the effects of the multimodal learning analytics features on predicting students group performances in open-ended project-based learning, below results are found:

- IDEC (Arduino IDE) removal does not effect the results of the classifiers,
- Removal of all face and hand duration has very little effect on the classifiers,
- Distance measures DHB and DBL **alone** are capable of predicting the results with a high accuracy (0.75) across classifiers,
- The audio level feature AUD **alone** is currently a **strong** feature for classification (1.0 with Naive Bayes) with time windows 5min,10min and 30min and binary scoring.

Interestingly the logistic regression is capable of an optimal result (1.0) when considering IDEX, IDEVHW, IDEVSW, and DBL, which are the main IDE features, except component counts and the distance between learners (DBL). One of the main limitations of our approach is on the scoring of the sessions that is limited to a binary classification with respect to a richer 3-level human scoring.

Discussion

In this article, we started from the hypothesis that specific features in MMLA can provide useful information about the quality of groups' interactions, therefore to the artifacts produced in as part of students' project-based learning. From the high-frequency multimodal data collected, we compared different machine learning approaches (that employed deep learning techniques and traditional) for their accuracy to predict human grading of the groups' artifact quality. In our first approach, using these classifiers, we identified the most effective features of MMLA to predict the students' group performances in project-based learning activities. More specifically, we used various machine learning classifiers to predict the poor student performances in terms of the groups' artifact quality based on multimodal data. We were not satisfied with the binary grading system or the large time window. These issues let us to the second approach where first we improved the classification of the student's project outcomes into 5 categories. Then we used deep learning neural networks to further explore this research to evaluate student performances in project-based learning using multimodal data.

Traditional Approach

In the linear regression approach, we focused on identifying the different phases of work in relation to accuracy in predicting the groups' artifact quality. We found that the planning and building stages of students learning activities are better predictors of their artifact quality than the reflection stage (in the intervention the reflection phase signalled the end of making artefacts and coding to documenting with a mobile device the work).

After looking at the different phases, we investigated the certain features of the MMLA, in order to determine which features can predict the students' artefact quality with higher accuracy. Our results show that the Distance between Hands (DBH) and Distance between Learners (DBL) are key features to predict students' performances in project-based learning activities. In our case, they highly correlate with the quality of the students' artefacts in project-based learning. These results are aligned with existing research on PBL activities that show the value of nonverbal indexes of student interaction in estimating their success at learning processes (Cukurova, Luckin, Millan, & Mavrikis, 2018) as well as MMLA research findings that show the potential of hand motion and speed, and the location of the learners to predict student success at various learning outcomes (Blikstein, 2011; Grover et al., 2016; Ochoa et al., 2013). As mentioned in the background section there are three main aspects of Project Based Learning (PBL): students are asking driving questions, doing investigations to answer these questions, and collaborate together to solve these questions (Krajcik, 2010). It is important that MMLA research aims to support these three main aspects of PBL. The results presented here that show the value of the distance between students' hands and distance between students to predict students' success at PBL, are well aligned with the argument that closer students may fruitfully collaboration which is an important aspect of PBL.

The other features of MMLA such as Hand Motion Speed (HMS), Faces Looking at the Screen (FLS), did not perform very well to predict students' artefact quality across the classifiers. While the Arduino IDE the Number of Active Blocks (IDEC), the Variety of Hardware (IDEVHW) and Software Blocks used (IDEVSW) and the number of interconnections between blocks as a Measure of Complexity (IDEX) were able to predict students' outcomes, they were only marginal across the classifiers. Furthermore, the audio signal level(AUD) appears to be a promising feature to predict performance, however more investigation is needed for using this feature in combination with others.

Deep learning Approach

The DNN results are more promising and show the feasibility of this method as an efficient approach for MMLA. In our investigation with this approach, we obtained net achieves a mean squared error of 0.13 with a window of size 240s as shown in table 4. One important result emerged from our results that is worth to notice is how the smallest window performs worse than the others, see Table 3. This is possibly due to the low information amount in that time window. The 240s interval performs the best, while the 360s interval gives no performance gain as can be seen in Table 5. This suggests that the information gain from 240s to 360s is negligible for our purposes.

It is possible to see that (see Table 7) by removing a single feature, in general results get worse except partially in the case of the distance between faces. This shows that this is a very strong input feature. It is also important to notice that the network learned some higher level features which do not consist of a single input, given that by removing any single input we can not achieve the optimal results which we achieved using them all.

All results show a reasonably low variance evidencing the stability of the results, which is a positive sign in terms of the learned features. The fact that strong features have been trained is possibly due to the 0.5 dropout value which "encourages" the network to find high level, strong features discarding the low level, weak features. Regularization gave no significant boost of the results, but this is probably due to the relatively "small" amount of training data, avoiding partially the problem of over-fitting. This parameter should become more relevant when more data will be added to the training set. A future step could consist in removing pairs or triplets of features to understand the relationship and importance of the input features further and make the factors on the learning process more visible. We aim to further investigate these in our immediate future work.

Conclusion

Recently, there is a growing interest in project-based learning globally. This is, at least in part, due to an increased demand for the "21st century skills" and the potential of project-based learning to improve student skills to better prepare them for the future. The evidence set out in recent influential reports (see for instance Luckin, Baines, Cukurova, and Holmes (2017) confirms that these skills, look set to be increasingly relevant not just to many of the jobs that will survive new waves of automation, but also to our ability to cope in everyday life. However, project-based learning requires appropriate support of students while they are engaging with physical materials and with each other (Cukurova, Bennett, & Abrahams, 2017; Kirschner & van Merriënboer, 2013).

In this paper, we show that MMLA and the state-of-the-art computational techniques can be used to generate insights into the "black box" of learning in students' project-based activities. These insights generated from multimodal data can be used to inform teachers about the key features of project-based learning and help them support students appropriately in similar pedagogical approaches. Towards achieving this ultimate aim, this paper has three main contributions to the field. First, we show that the distances between students' hands and faces while they are working on projects is a strong predictor of students' artefact quality which indicates the value of student collaboration in these pedagogical approaches. Second, we show that both, new and promising approaches such as neural networks and more traditional regression approaches, can be used to classify MMLA data and both have advantages and disadvantages depending on the research questions and contexts being investigated. At last but not least, although, it is traditionally notoriously challenging to provide evidence about the robust and objective evaluations of project-based learning activities, techniques and types of data we presented here can be the first step towards effective implementation and evaluation of project-based learning at a scale.

Contributions

This work is the result of the collaborative effort between the institutions participating to the PELARS FP7 project. DS and MC designed the protocol with students; DS conducted the evaluation with students; ER and GD designed the software and analyzed data. All contributed to writing the manuscript.

Acknowledgments

The PELARS project received funding from the European Union's Seventh Framework Programme for research, technological development and demonstrations, under grant agreement 619738.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., … Isard, M., et al. (2016). Tensorflow: A system for large-scale machine learning. In *Proceedings of the 12th usenix symposium on operating systems design and implementation (osdi). savannah, georgia, usa.*

Atman, C. J., Adams, R. S., Cardella, M. E., Turns, J., Mosborg, S., & Saleem, J. (2007). Engineering design processes: A comparison of students and expert practitioners. *J. Eng. Educ. 96*(4), 359–379. doi:10.1002/j.2168-9830.2007.tb00945.x

Bell, S. (2010). Project-based learning for the 21st century: Skills for the future. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas.*

Bergstra, J., Breuleux, O., Bastien, F., Lamblin, P., Pascanu, R., Desjardins, G., … Bengio, Y. (2010). Theano: A cpu and gpu math compiler in python. In *Proc. 9th python in science conf* (pp. 1–7).

Blikstein, P. (2011). Using learning analytics to assess students' behavior in open-ended programming tasks. In *Proceedings of the 1st international conference on learning analytics and knowledge - lak '11* (p. 110). New York, New York, USA: ACM Press. doi:10.1145/2090116.2090132

Blikstein, P. (2013). Multimodal learning analytics. In D. Suthers & K. Verbert (Eds.), *Proceedings of the third international conference on learning analytics and knowledge - lak '13* (p. 102). New York, New York, USA: ACM Press. doi:10.1145/2460296.2460316

Blikstein, P. & Worsley, M. (2016). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*. Retrieved October 12, 2016, from http://epress.lib.uts.edu.au/journals/index.php/JLA/article/view/4383/5596

Blumenfeld, P. C., Soloway, E., Marx, R. W., Krajcik, J. S., Guzdial, M., & Palincsar, A. (1991). Motivating project-based learning: Sustaining the doing, supporting the learning. *Educ Psychol*, *26*(3-4), 369–398. doi:10.1080/00461520.1991.9653139

Chen, L., Feng, G., Joe, J., Leong, C. W., Kitchen, C., & Lee, C. M. (2014). Towards automated assessment of public speaking skills using multimodal cues. In *Proceedings of the 16th international conference on multimodal interaction - icmi '14* (pp. 200–203). New York, New York, USA: ACM Press. doi:10.1145/2663204.2663265

Chollet, F. (n.d.). Keras: Deep learning library for theano and tensorflow (2015).

Clow, D. (2013). An overview of learning analytics. *Teaching in Higher Education*, *18*(6), 683–695. doi:10.1080/13562517.2013.827653

Cukurova, M., Avramides, K., Spikol, D., Luckin, R., & Mavrikis, M. (2016). An analysis framework for collaborative problem solving in practice-based learning activities: A mixed-method approach. In *Proceedings of the sixth international conference on learning analytics & knowledge* (pp. 84–88). ACM.

Cukurova, M., Bennett, J., & Abrahams, I. (2017). Students' knowledge acquisition and ability to apply knowledge into different science contexts in two different independent learning settings. *Research in Science & Technological Education*, *35*(4).

Cukurova, M., Luckin, R., Millan, E., & Mavrikis, M. (2018). The nispi framework: Analysing collaborative problem-solving from students' physical interactions. *Computers & Education*, *116*, 93–109. doi:10.1016/j.compedu.2017.08.007

Dewey, J. (1959), In *Dewey on education*. New York: Teachers College Press.

Greller, W. & Drachsler, H. (2012). Translating Learning into Numbers : A Generic Framework for Learning Analytics Author contact details : *Educational Technology & Society*, *15*(3), 42–57. doi:http://hdl.handle.net/1820/4506

Grover, S., Bienkowski, M., Tamrakar, A., Siddiquie, B., Salter, D., & Divakaran, A. (2016). Multimodal analytics to study collaborative problem solving in pair programming. In *Proceedings of the sixth international conference on learning* *analytics & knowledge - lak '16* (pp. 516–517). New York, New York, USA: ACM Press. doi:10.1145/2883851.2883877

Halverson, E. R. & Sheridan, K. (2014). The maker movement in education. *Harv Educ Rev*, *84*(4), 495–504. doi:10.17763/haer.84.4.34j1g68140382063

Katterfeldt, E.-S., Cuartielles, D., Spikol, D., & Ehrenberg, N. (2016). Talkoo: A new paradigm for physical computing at school. In *Proceedings of the the 15th international conference on interaction design and children - idc '16* (pp. 512–517). New York, New York, USA: ACM Press. doi:10.1145/2930674.2935990

Kirschner, P. A. & van Merriënboer, J. J. (2013). Do learners really know best? urban legends in education. *Educational psychologist*, *48*(3), 169–183.

Kotsiantis, S., Patriarcheas, K., & Xenos, M. (2010). A combinational incremental ensemble of classifiers as a technique for predicting students performance in distance education. *Knowledge-Based Systems*, *23*(6), 529–535.

Krajcik, J. (2010). Project-based science: Engaging students in three-dimensional learning. *The Science Teacher*, *82*(1), 1–25.

Krajcik, J. & Blumenfeld, P. (2006). Project-based learning. In R. Sawyer (Ed.), *The cambridge handbook of the learning sciences* (pp. 317–334). New York, Cambridge, USA: Cambridge University Press.

Luckin, R., Baines, E., Cukurova, M., & Holmes, W. (2017). *Solved! Making the case for collaborative problem-solving*. NESTA. London, UK. Retrieved from http://www.nesta.org.uk/sites/default/files/solved-making-case-collaborative-problem-solving.pdf

Martinez-Maldonado, R., Schneider, B., Charleer, S., Shum, S. B., Klerkx, J., & Duval, E. (2016). Interactive surfaces and learning analytics: Data, orchestration aspects, pedagogical uses and challenges. In *Proceedings of the sixth international conference on learning analytics & knowledge - lak '16* (pp. 124–133). New York, New York, USA: ACM Press. doi:10.1145/2883851.2883873

Munoz-Salinas, R. (2012). Aruco: A minimal library for augmented reality applications based on opencv. *Universidad de Córdoba*.

Ochoa, X., Chiluiza, K., Méndez, G., Luzardo, G., Guamán, B., & Castells, J. (2013). Expertise estimation based on simple multimodal features. *Proceedings of the 15th ACM International Conference on Multimodal Iteraction (ICMI '13)*, 583–590. doi:10.1145/2522848.2533789

Ochoa, X., Worsley, M., Chiluiza, K., & Luz, S. (2014). Mla'14: Third multimodal learning analytics workshop and grand challenges. *Proceedings of the 16th*. Retrieved October 15, 2016, from http://dl.acm.org/citation.cfm?id=2668318

Ochoa, X. & Worsley, M. (2016). Editorial: Augmenting learning analytics with multimodal sensory data. *Journal of Learning Analytics*.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Prieto, L. P., Sharma, K., Dillenbourg, P., & Jesús, M. (2016). Teaching analytics: Towards automatic extraction of orchestration graphs using wearable sensors. In *Proceedings of the sixth international conference on learning analytics & knowledge - lak '16* (pp. 148–157). New York, New York, USA: ACM Press. doi:10.1145/2883851.2883927

Ruffaldi, E., Dabisias, G., Landolfi, L., & Spikol, D. (2016). Data collection and processing for a multimodal learning analytic system. In *2016 sai computing conference (sai)* (pp. 858–863). IEEE. doi:10.1109/SAI.2016.7556079

Schneider, B. & Blikstein, P. (2014). Unraveling Students ' Interaction Around a Tangible Interface Using Gesture Recognition. *Journal of Educational Data Mining*, *7*(2), 320–323. Retrieved from http://blog.bertrandschneider.com/wp-content/uploads/2012/01/23%7B%5C_%7D102-844-3-CE.pdf

Siemens, G. & Baker, R. S. J. d. (2012). Learning analytics and educational data mining: Towards communication and collaboration. In *Proceedings of the 2nd international* *conference on learning analytics and knowledge - lak '12* (p. 252). New York, New York, USA: ACM Press. doi:10.1145/2330601.2330661

Spikol, D., Cukurova, M., & Ruffaldi, E. (2017). Estimation of success in collaborative learning based on multimodal learning analytics features. In *Proceedings of the 17th ieee international conference on advanced learning technologies*. Romania: IEEE.

Spikol, D., Ehrenberg, N., Cuartielles, D., & Zbick, J. (2015). Design strategies for developing a visual platform for physical computing with mobile tools for project documentation and reflection. *Proceedings of the Workshops at the th International Conference on Artificial Intelligence in Education AIED.*

Sridhar, S., Mueller, F., Zollhöfer, M., Casas, D., Oulasvirta, A., & Theobalt, C. (2016). Real-time joint tracking of a hand manipulating an object from rgb-d input. In *European conference on computer vision* (pp. 294–310). Springer.

Viola, P. & Jones, M. J. (2004). Robust real-time face detection. *International journal of computer vision*, *57*(2), 137–154.

Worsley, M. (2012). Multimodal learning analytics. In *Proceedings of the 14th acm international conference on multimodal interaction - icmi '12* (p. 353). New York, New York, USA: ACM Press. doi:10.1145/2388676.2388755

Worsley, M. (2014). Multimodal learning analytics as a tool for bridging learning theory and complex learning behaviors. *3rd Multimodal Learning Analytics Workshop and Grand Challenges, MLA 2014*, 1–4. doi:10.1145/2666633.2666634

Worsley, M. & Blikstein, P. (2014). Analyzing engineering design through the lens of computation. *Journal of Learning Analytics*, *1*(2), 151–186.

Zbick, J., Vogel, B., Spikol, D., Jansen, M., & Milrad, M. (2016). Toward an adaptive and adaptable architecture to support ubiquitous learning activities. In A. Pena-Ayala (Ed.), *Mobile, ubiquitous, and pervasive learning* (Vol. 406, pp. 193–222). Advances in Intelligent Systems and Computing. Cham: Springer International Publishing. doi:10.1007/978-3-319-26518-6_8

Table 1

Table of the 18 session scores organized by team. The five scores expressed in a 5 level Likert-type are reported

Team	Session	Clarity	Indep Thinking	Plan	Solution Working	Quality
A	1	5	2	5	4	3
A	2	1	1	5	5	5
A	3	5	3	5	4	5
B	1	2	3	3	3	2
B	2	1	3	3	1	1
B	3	2	4	1	3	2
C	1	1	4	3	5	4
C	2	2	1	5	5	5
C	3	5	3	2	2	2
D	1	4	5	1	1	1
D	2	5	3	4	4	4
D	3	5	4	3	3	3
E	1	4	4	4	3	3
E	2	2	1	3	3	3
E	3	2	2	3	4	2
F	1	2	5	3	5	5
F	2	3	5	2	1	2
F	3	1	3	2	1	1

Table 2

Machine Learning Tasks performed over Data

Method	Deep Learning	Traditional
Task	Regression	Classification
Input	18 variables	9 variables per-window
Output	6 scores over 5 levels	1 score with 3 levels
Metrics	Regression Score	Classifier Accuracy
Windowing	120,240 and 360 seconds	10,20,30,90 minutes
Phase Exclusion	Reflection	Reflection
Method	Multiple layers	NB, LR, SVML, SVMR

Table 3

Results for the 120s window, 0.242 overall accuracy

120s Window	Loc	InTh	CorPi	DoWo	QuaOS	OG
Mean	0.182	0.238	0.166	0.197	0.155	0.228
Var	0.074	0.112	0.069	0.076	0.061	0.099

Table 4

Results for the 240s window, 0.129 overall accuracy

240s Window	Loc	InTh	CorPi	DoWo	QuaOS	OG
Mean	0.086	0.175	0.150	0.175	0.154	0.084
Var	0.074	0.056	0.084	0.092	0.062	0.048

Table 5

Results for the 360s window, 0.193 overall accuracy

360s Window	Loc	InTh	CorPi	DoWo	QuaOS	OG
Mean	0.213	0.077	0.237	0.147	0.196	0.181
Var	0.097	0.006	0.083	0.063	0.071	0.057

Table 6

Best network results for the different network configurations

Layers	Error	Window (s)
1024	0.186	360
1024, 512	0.174	360
1024, 512, 256	0.129	240

Table 7

Best error scores after removing isolated features

Removed Feature	Best Result
No features removed	0.129
All faces data	0.21
All Arduino data	0.21
DBF	0.15
DBH	0.21
HMS	0.19
AUD	0.18
Hand pos	0.21
Arduino comp	0.19

Table 8

Effect of phases in the inclusion of the classifier. P=plan, W=work, R=reflect

	PWR	**PW**	**W**	**WR**
NB	0.8	0.8	0.6	0.75
SVML	0.6	0.75	0.75	0.8
SVMR	0.75	0.75	0.75	0.75
LR	0.6	0.75	0.5	0.6

Practitioners Notes

What is currently known about Multimodal learning analytics?

- Multimodal learning analytics (MMLA) provides new tools and techniques to capture different types of data from complex learning activities in dynamic educational environments where learners interact in groups and with materials.

What does the paper add to the subject matter?

- MMMLA supports how teachers and learners can gain insights and support through the analysis of data (via computer machine learning) about small group work.
- These insights help educators design better learning situations, and students reflect on group work. The paper adds to the subject matter by comparing different techniques to analyse data.

The implications of the study findings for practitioners?

- MMLA and the state-of-the-art computational techniques can be used to generate insights into the "black box" of learning in students' project-based activities. These insights generated from multimodal data can be used to inform teachers about the key features of project-based learning and help them support students appropriately in similar pedagogical approaches.
- The study provides evidence that MMLA techniques and different types of can be the first step towards effective implementation and evaluation of project-based learning at scale.