Using Data Pipelines to reflect on a flipped classroom, problem-based learning course

Tsoni Rozita[1], Zotou Maria[2], Tarabanis Konstantinos [2], Tambouris Efthimios[2], Verykios Vassilios S.[1]
rozita.tsoni@ac.eap.gr, mzotou@uom.edu.gr, kat@uom.edu.gr, tambouris@uom.edu.gr, verykios@eap.gr
[1] Hellenic Open University
[2] University of Macedonia

Abstract

The digital era includes the rapid generation of data, knowledge, and technologies. This demands changes in education and the way that people learn in general. Problem-Based Learning PBL enables active participation in the learning process and supports the development of transversal and lifelong learning skills. This paper presents a LA process to evaluate a PBL course by leveraging Data Pipelines. Data from a PBL course, were imported into a data pipeline that was designed to incorporate all the processes of a Learning Analytics (LA) cycle. The results of the analysis were visualized to provide actionable knowledge to educational stakeholders for evidence-based decision-making.

Key words: Problem based learning, Flipped classroom, Positivity indicator, Data pipelines, Correlation

Introduction

The digital era includes the rapid generation of data, knowledge, and technologies. The amount of knowledge in the world has doubled in the decade 2004-2014 and is doubling every 18 months (Siemens, 2014). This requires the development of skills that can support us to adapt to these changes, make sense of all the new knowledge and be competitive in the existing and newly emerging professional fields (Persico & Pozzi, 2015). A promising solution to address the above requires a shift in the way we learn new things. This requires that we stop passively receiving knowledge, and learn how to think critically, solve problems, make sense of new data, and collaborate with others (Hung, 2011). A well-established learning strategy that allows the development of such skills is Problem Based Learning (PBL), as it enables active participation in the learning process and supports the development of transversal and lifelong learning skills.

As a consequence, when learners participate actively in learning, they produce a lot of data that could help educators facilitate and scaffold learners' progress and improve their performances. This data can be utilized and exploited with the usage of technologies that can help store, analyze and visualize learners' progress through the application of Learning Analytics (LA) methodologies and tools (Carr, 2010; Zotou, 2015). These LA methods and tools can record and analyze data that is generated during learning and provide informative insights on the learning process (e.g., navigation paths, social networks formed, passive learners, most commonly made mistakes, drop-outs, etc). This can in turn help educators become more aware of their learners' possible low engagement and at-risk failures and in turn, provide them with additional learning materials, hints, adaptive practice assignments, etc.

This paper presents a LA process to evaluate a PBL course by leveraging Data Pipelines. The course's methodology is presented along with the data analysis of learners' responses to

Panagiotakopoulos, C., Karatrantou, A., & Armakolas, S. (Eds.), *Proceedings of 7th Panhellenic Scientific Conference "Integration and Use of ICT in the Educational Process"*, pp. 103-116, Dept. of Educational Sciences and Social Work, University of Patras, Greece, 16-18 September 2022. ISBN 978-618-83186-7-0.

questionnaires designed to retrieve information on how positive learners' felt about their weekly knowledge levels, and how they assessed each weekly lecture. The analysis of this data follows, using Data Pipelines and the paper draws conclusions based on the evaluation results.

The rest of the paper is structured as follows: Section 2 presents the background of our work. Section 3 presents the research methodology and the dataset to be analyzed, while Section 4 provides analytical information on the evaluation results and discussion. Section 5 presents the conclusions drawn from the research that was carried out.

Background and theory

Learning from problems is an instinctual human impulse that accompanies us in our everyday life. We spend a significant amount of effort trying to solve a variety of issues and accumulate relevant knowledge to resolve them accordingly and habitually (Barrows and Tamblyn, 1980). Thus, in theory, transferring this process to educational contexts should not be very difficult. That is not the case though; learning through solving problems appears to face significant challenges, mostly regarding the mentality change that is needed by all involved to abandon traditional teaching and make way for innovative student-centered learning styles (Northwood et al, 2003).

Traditional learning usually does not allow students to capitulate on already gained knowledge, since its retrieval requires the stimulus of the working memory with corresponding signals, a process that passive knowledge delivery does not entail. The learning strategy known as Problem Based Learning (PBL) allows such events to occur. The known Chinese proverb: "tell me and I will forget. Show me and I will remember. Involve me and I will understand. Step back and I will act." is often referenced as a representative definition of PBL (Enemark, 2002). Moreover, according to Barrows and Tamblyn (1980), PBL is considered to be "…the learning that results from the process of working toward the understanding or resolution of a problem. The problem is encountered first in the learning process."

This approach shifts the focus from understanding common knowledge to the ability to develop new knowledge through "learning by doing" activities. Additionally, the gained knowledge tends to be stored in memory patterns that facilitate later recall. Hence, consistent practice in PBL courses gradually enhances students' performance and underpins in-depth knowledge comprehension.

As the transformation from traditional learning to student-based courses can be challenging for both educators and learners, different models have been proposed to facilitate and guide the design of PBL-based courses. One of the most known and well-established models is the Aalborg PBL model. Aalborg University holds more than 30 years of experience in PBL, and more specifically in project–organized courses, thus focusing on the application of the new knowledge instead of solely its production (Perrenet et al., 2000).

The steps this model proposes are as follows:

Group forming. Breaking down into groups, where each group will address a different problem.

Problem formulation. Definition of the problem that needs solving. Each problem must be consistent with the course curricula and approved by the teacher in order to ensure the availability of a solution.

Task formulation. Problem objectives specification and subsequent task distribution.

Problem analysis. Group investigation of the problem using specific parameters and dimensions. Decisions such as the problem's scope and the needed resources are made in this step. During this step each group must define the limitations of the problem and re-configure the task and responsibilities allocation depending on the group's findings.
Solution (Data gathering, Analysis, Design). Determination of solutions by each group through constant discussions and teacher guidance.
Implementation. Realization of the solution with corresponding resources.
Evaluation and Reporting. Quantitative and qualitative review and evaluation of the project by the class.

Related work

Evaluating the learning experience is not a simple process. It is related to objective factors like design and organizational issues and the available teaching resources for the course. However, it is related to subjective factors as well, like students' learning behavioral patterns and sentiment, the level of collaboration, and the structure of the learning community that reflects the social aspect of learning. Several studies are aiming to assess these factors to add knowledge to the educational domain. LA provides information that educational stakeholders can leverage by diving into students' data and transforming them into actionable knowledge. One of the five areas of LA that Bienkowski, Feng, and Means (2012) describe is student profiling. Several research papers are focusing on students' profiling in higher education. Tsoni, Sakkopoulos, and Verykios (2021) combined Social Network Analysis with Principal Component Analysis with Clustering to identify groups of students with special features that could help tutors in their task to support them. Additionally, this type of analysis was also used to compare two successive courses of a postgraduate program, revealing the maturation of the learning community concerning communication and interaction. Cluster analysis based on e-tutorial trace data was used by Tempelaar et al. (2018) allowing student profiling into different at-risk groups. Khalil & Ebner (2017) classified students into appropriate categories based on their level of engagement in Massive Open Online Courses to improve the course's levels of completion. An additional factor that affects learning is the educational material and the methodology. Paxinou et al. (2017; 2020) used advanced methods of analysis to evaluate students' performance in virtual laboratories.

When log data are combined with sentiment features, rich information is revealed allowing us to draw conclusions about learning (Tsoni et al., 2019; 2020). For example, by visualizing the forum interaction of students where each post was annotated based on its polarity (positive, negative, and neutral posts) typical roles like the "Most Knowledgeable Other" of the theory of Vygotsky (Tsoni & Verykios, 2019). Kagklis et al., (2015) analyzed students' posts polarity in the discussion fora and concluded that their sentiment was marginally related to their academic performance, however, this knowledge can help tutors improve their support. In the study of Nkomo et al. (2020) Social Network Analysis along with sentiment analysis help researchers investigate students' attitudes towards online learning material. Moreover, sentiment analysis combined with additional data can contribute to evaluating students' progress and produce results that are easier to interpret (Watkins, 2020).

Especially in Higher Education where students hand in assignments there is often the concern of plagiarism and cheating (Tsoni & Lionarakis, 2014). While there is relevant software to evaluate cases of plagiarism, there is a possibility of students taking external help

or presenting foreign unpublished work as their own. Gkontzis et al. (2018) presented an LA approach to spot outliers who probably cheated in their assignments. Jaramillo-Morillo et al. (2020) developed a learning analytics algorithm to detect dishonest students based on submission time and exam responses providing as output a number of indicators that can be easily used to identify students. Trezise (2019) suggested that keystroke and clickstream analysis may be able to distinguish between a student writing an authentic piece of work and one transcribing a completed work.

An LA cycle aims to inform stakeholders to take action based on the evidence that data revealed. That means that reporting the results is of high importance. Interactive dashboards can be a solution that offers large amounts of information in a readable and understandable way even for non-experts (Tsoni et al., 2022a). A data pipeline that begins with data importing from a data warehouse and ends up in an interactive dashboard can be beneficial for educators due to the up-to-date knowledge and the adaptability that can provide (Tsoni et al., 2022b). In the next section the methodology to build such a pipeline to evaluate a PBL course is presented.

Methodology and dataset

The design, delivery, and assessment of a course using the PBL model have been carried out within the context of a postgraduate course named "Systems Analysis and Design" at the [removed for peer review purposes]. For technological support, we used the Moodle Learning Management System (LMS) to simulate an online PBL environment, due to the COVID-19 restrictions and quarantine. Each step of the PBL model was mapped to specific weeks of the course, based on the problem-solving phase students were in, as shown in Table 1.

Table 1: PBL model mapping to SAD course

PBL step	Course session
Group forming	Week 2
Problem formulation	Week 2
Task formulation	Week 2
Data gathering	Week 3
Analysis	Weeks 4,5,6,7,8
Design	Weeks 9, 10, 11
Implementation	Week 12
Evaluation	Week 13
Reporting	Week 13

Each week, students were asked to voluntarily fill out the "optimism indicator" (scale of 0 to 10) using their actual names to indicate how optimistic they were feeling that week about their progress in the course. Students were also asked to grade each weekly lecture anonymously, by adding a score (scale of 0 to 10) in four parameters, namely:

1. Interesting
2. Easy to understand
3. Student participation
4. Teaching methodology and supporting material

Research questions

RQ1: How does the positivity indicator vary in time and during different steps of the PBL process?
RQ2: What were the weekly differences in students' positivity and evaluation of the course?
RQ3: Are there any statistically significant correlations between the positivity indicator and students' grades?
RQ4: Are there any statistically significant correlations between the positivity indicator and the evaluation metrics indicating the important factors that affect students' disposition towards their learning experience?

A Pipeline for pre-processing and analysis

To address the abovementioned questions, we designed and implemented a data pipeline to support the sequence of processes needed in a Learning Analytic (LA) cycle. Data pipelines offer benefits that are very important for educational research. Firstly, they minimize human intervention and allow automated periodic data uploading to produce updated results. These results can be visualized and deployed so that educational stakeholders can be timely informed and make evidence-based decisions regarding teaching and support. Additionally, it is easier to protect students' privacy by implementing customized anonymization and privacy-preserving processes. Their flexibility is mainly due to the ability to make interventions in every step of the process enabling the cyclical process that is necessary for LA. Taking into account that education is a vivid and evolving procedure, LA should be able to adjust to these changes through testing and evaluating the results to trigger new cycles of analysis that are made possible by re-using the workflows that are already stored by resetting and executing them from scratch. Data pipelines also offer extensibility and scalability in big data allowing methods that have been tested in pilot studies to be extended in real-life big educational data. Finally, the validity of the results can be easily confirmed due to the repeatability that data pipelines provide.

The pipeline that was used for our analysis starts with importing data retrieved from the LMS. Although all data files are originated from the same source, they come in multiple data files. Thus, the pre-processing part of the analytical process is one of the most challenging (Figure 1). Next, there are the steps of analysis, evaluation, and reporting.

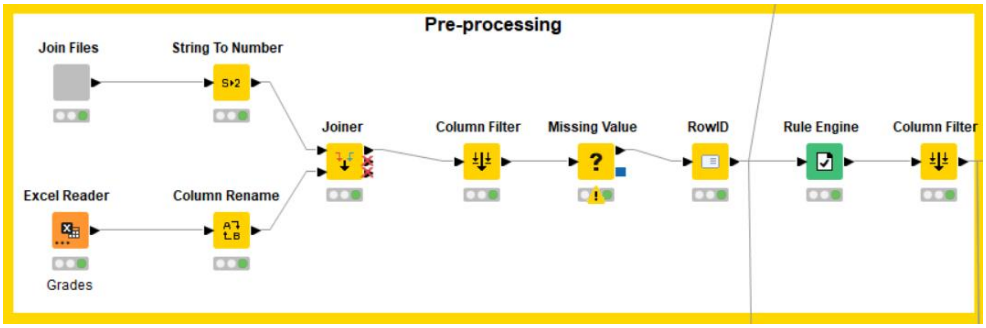

Figure 1: The pre-processing part of the workflow

The implementation of the pipeline was carried out using the KNIME platform. The KNIME is an open-source, freely available online software for data analysis and reporting. Its graphical interface of visual programming minimizes the requested programming skill level

to basic. Users can create workflows consisting of nodes that is, entities that deliver a certain task. Complex sequences of nodes or sub-workflows can be wrapped-up in components. Nodes can be modified and adjusted by configuration windows that provide a range of options depending on the node's type. Once a node is executed the output ports are activated providing results and graphs. The workflows can be stored combined or exported offering flexibility and repeatability. Additionally, they can be fed with updated data for longitudinal research.

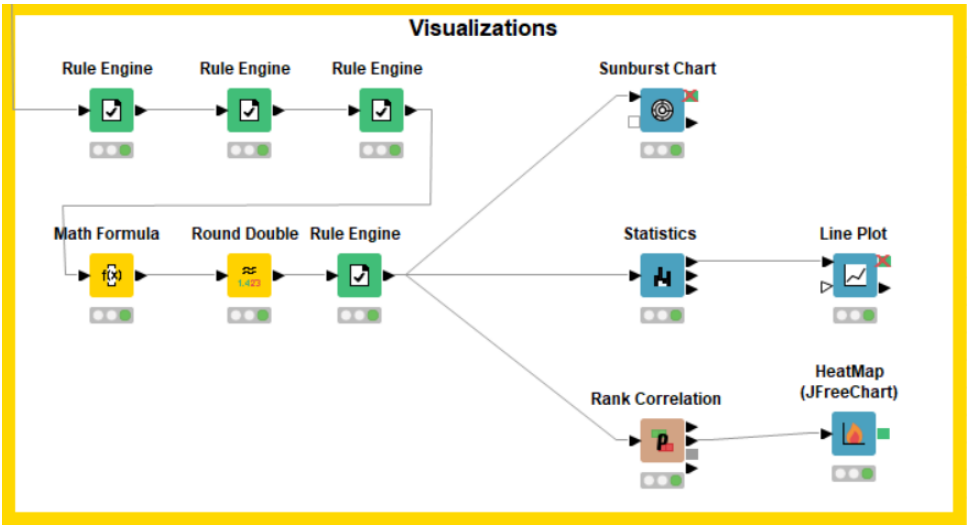

Figure 2: Visualization of students' course evaluation

Results and discussion

In this section, the main results of the experimental assessment of the data pipeline are presented and discussed. In the first part of the section, the results that concern students individually are presented, followed by the course's evaluation and the PI means per week. Finally, the results of the correlation analysis are visualized.

The positivity indicator during the steps of the PBL course

The PI captures students' overall attitude toward their learning experience in this certain moment. It is expected to reflect their perceived progress, and thus be affected as the teaching process unfolds. Figure 3 shows the PI for each student per week. Although each student follows a different path concerning his/her attitude as it is expressed through the positivity indicator, there are some students that, in some weeks, show changes in their positivity levels reflecting their attitude towards a new part of the PBL process.

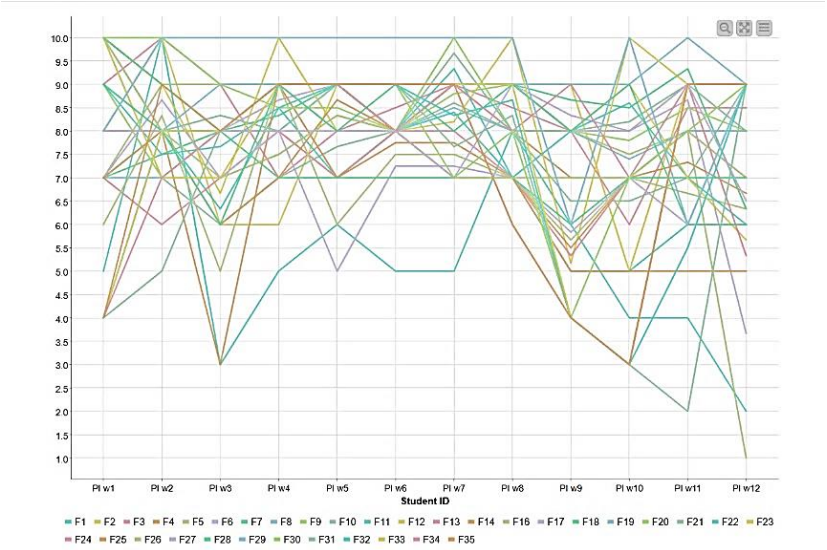

Figure 3: PI for each student per week

Student F1 is an indicative case of a student who experiences mood transitions according to the progress of the learning process. He/she starts with a mediocre optimism towards the course (5 out 10 in the first week) that becomes very high the next week, just to drop almost to the minimum level a week later. Although he/she manage to successfully complete the course the last week he/she presented descending positivity levels most probably reflecting difficulty in stress management. It is worth noticing that this student started with a target grade of 8/10 and manage to score 7/10 underachieving his/her goal only at 10%. However, his/her average positivity was below 6 on a scale from 0 to 10.

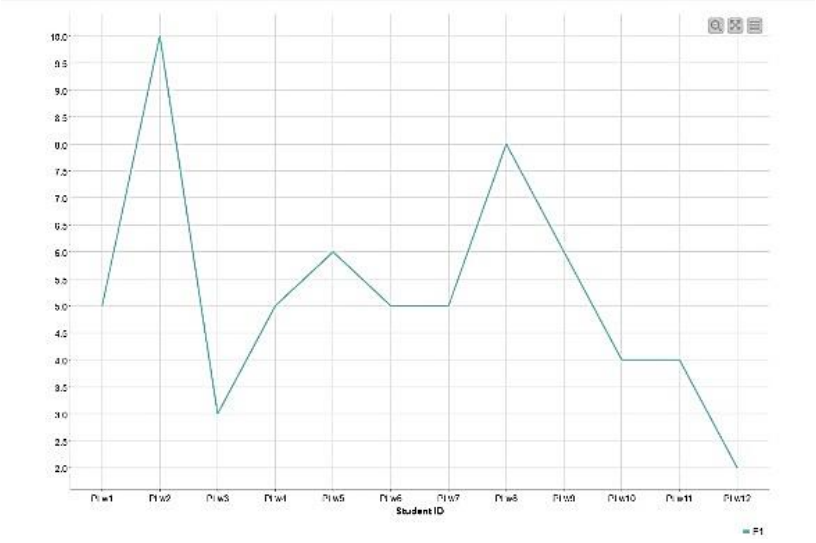

Figure 4: The descending trajectory of the PI for a (probably stressed out) student.

The positivity indicator of some students might be indicative of the difficulties that they faced during a certain part of the course. In Figure 5 three students with differences in their positivity levels follow a common pattern with a straight drop in weeks 9 to 11 that matches the design step of the course.

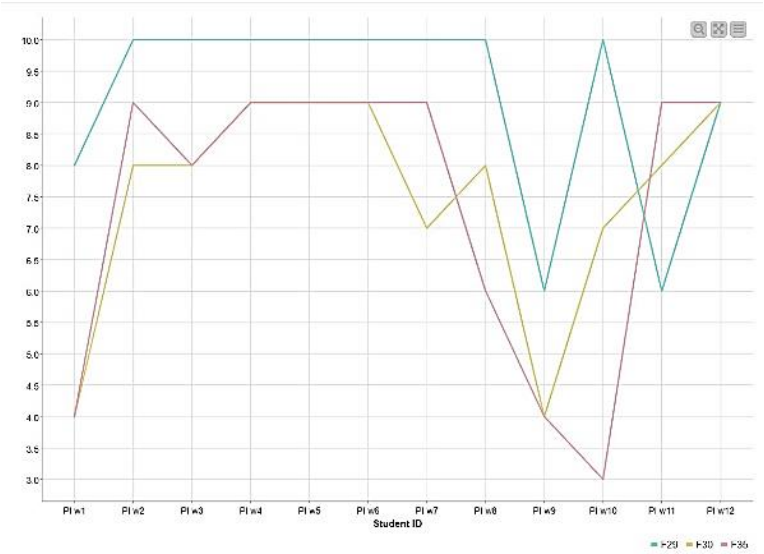

Figure 5: Similar students' PI patterns

Evaluation of the lectures

As was mentioned in the previous section, students could evaluate each lecture anonymously on a scale from 0 to 10, using four factors (Table 1):

1. Interest (i.e., the degree to which the lecture was interesting)
2. Easy to comprehend (i.e., the degree to which the lecture was easy to comprehend)
3. Students' participation (i.e., the degree to which students were motivated to participate)
4. Education method and supporting material

Table 2: Evaluation metrics

PI	Positivity Indicator
IN	Interest
EC	Easy to comprehend
CP	Students' Participation
EM	Educational Methodology and Materials

The data from the semester's evaluation were imported into the data pipeline for analysis and visualization. In the following graph (Figure 6) the variation of these metrics through time is shown. Additionally, the average PI for each week was calculated and it was compared with the evaluation of the course. This step was necessary since different weeks represent different steps of the course's methodology.

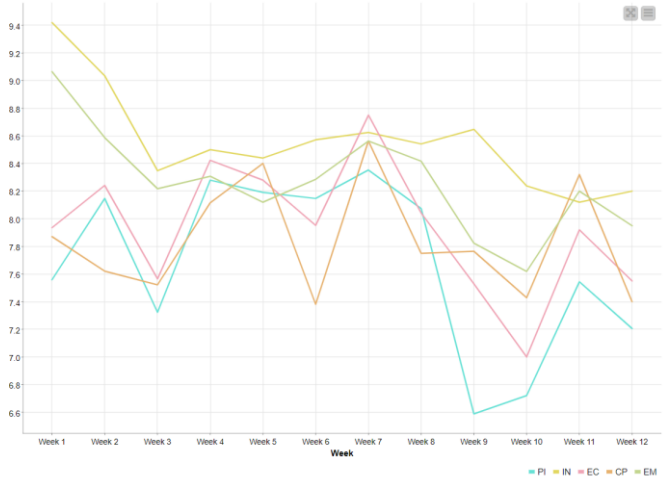

Figure 6: PI and evaluation metrics per week

Correlations

To test the covariation hypothesis that was supported by the graphs presented in the previous sub-section a correlation analysis was incorporated into the pipeline. First, the PI of each week and the grades were tested to reveal plausible relationships. The results are summarized in Figure 7. As was expected there is a high correlation between the grades of each student as they all express their academic performance. The fact that there is no significant positive correlation between the PIs of additional weeks indicates that the PI is not about a general attitude dependent mainly form each student's personality, but it is rather a fluid metric that indicates how learning is evolving.

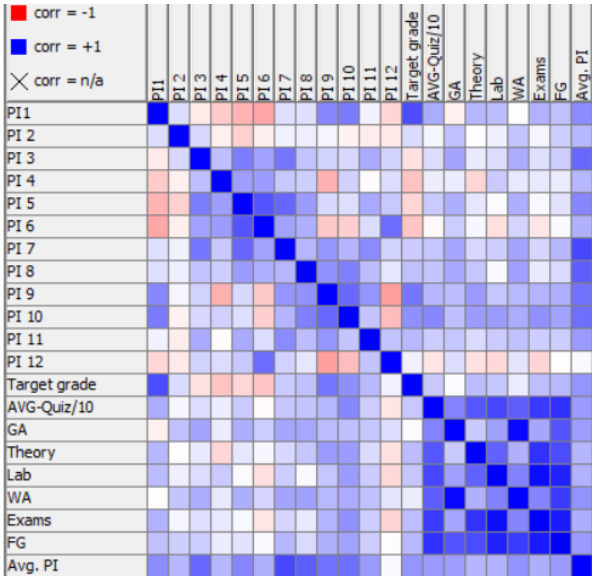

Figure 7: Correlation plot between PIs and grades

A relatively strong positive correlation was found between the Avg. PI and grade of the WA, and also, between the Avg. PI and the final grade. A moderately positive correlation between the Avg. PI and the grade of the GA. The PI of the 10th week has a moderately positive correlation with the grade in the quizzes, the grade in the exams, the grade in the theory, the grade in the lab, and the final grade. This indicates that this week is an important milestone in students' progress, strongly related to the level of awareness concerning their learning. Additionally, the PI of the 9th week is moderately positively correlated with the grade in theory, while the PI of the 7th and the 8th week is moderately positively correlated with the WA grade and the GA grade. The correlation analysis for the evaluation metrics and the Avg. PI per week is displayed in Figure 8 using a heatmap visualization.

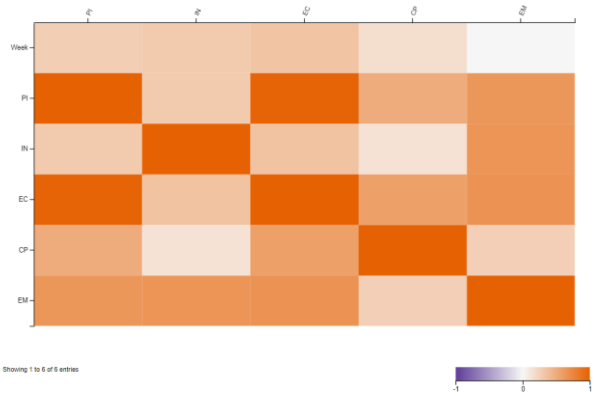

Figure 8: Heatmap of the correlation coefficients for PI and evaluation metrics

The results revealed five statistically significant correlations (Table 2). A very strong positive correlation was found between the Avg. PI and the EC, indicate that the main factor that affects students' attitudes is whether they understand the course or have difficulties. Additionally, a strong positive correlation between EM and EC proves that when students understand the lecture, they approve of the methodology. The INT is also strongly correlated with the EM which led us to the conclusion that successfully designed lectures improve the levels of students' participation. The EM is strongly correlated with the Avg. PI as well. Finally, the EC and the CP have a moderate positive correlation, as was expected, since students tend to be more engaged in learning processes that are within the range of their cognitive abilities.

Table 3: Correlation coefficients for the evaluation metrics

	PI	**IN**	**EC**	**CP**	**EM**
PI	1	NS	0.98	NS	NS
IN		1	NS	NS	0.65
EC			1	0.58	0.67
CP				1	NS
EM					1

Conclusion

This work presented a LA process to evaluate a PBL course by leveraging Data Pipelines. The pipeline that was designed and implemented in students' data provided simple visualization to capture the track of the positivity indicator, at first individually for each student. It was shown that differences in the learning behavior were imprinted in the different patterns of the graph (Figure 3). It appears that some indicative cases can be spotted, allowing tutors to make decisions about their actions that would support students who are under stress or in danger to fail the course. The evaluation metrics in combination with the PI provided us with information to reflect on concerning the educational design of the course. The 10th week seems to be an important week for students' self-regulation skills. The strong correlation of the 10th week's PI with their grades indicated that students have gained a strong awareness of their progress and what to expect as their final grade. The 9th week's PI, which is during the design phase of the PBL course, is correlated with their theory grade. Weeks 7 and 8 are important for the assignments' grades (both individual and group assignments). During these weeks students are involved with the tasks of analyzing the problem, finding the limitations, and sharing tasks and responsibilities. Thus, they already have a sense of how their assignments would go since there is a strong correlation between those weeks' PI and the assignments' grades. The Avg. PI as was expected is correlated with students' grades in the final exams and their written assignments, which is also an argument, supporting that the students demonstrate self-awareness skills.

Concerning the evaluation of the course, interestingly, students' interest is only correlated with the methodology and supporting materials. On the other hand, the ease of comprehension was strongly related to the PI, which is an expected outcome, but also to students' participation and the methodology, and the supporting materials. These results could drive decisions to improve the educational design and boost the learning outcomes. Hopefully, the knowledge deriving from the proposed LA method would raise learners' satisfaction.

References

Barrows, H. S., Tamblyn, R. M. (1980). *Problem-Based Learning: An Approach to Medical Education*. Springer Publishing Company.

Bienkowski, M., Feng, M., & Means, B. (2012*). Enhancing teaching and learning through Education*. Retrieved from http://www.ed.gov/edblogs/technology/files/2012/03/edm-labrief.

Carr, N. (2010). The shallows: *How the internet is changing the way we think, read and remember*. Atlantic Books Ltd. Washington, DC: U.S. Department of educational data mining and learning analytics.

Enemark, S. (2002). Innovation in surveying education. *Global Journal of Engineering Education*., vol 6, no. 2, pp. 153-159.

Gkontzis, A., Kotsiantis, S., Tsoni, R., & Verykios, V. S. (2018). *An Effective LA Approach to Predict Student Achievement*. In Proceedings of the 22nd Pan-Hellenic Conference on Informatics. ACM

Hung, W. (2011). Theory to reality: a few issues in implementing problem-based learning. *Educational Technology Research and Development*. Vol. 59, no. 4. DOI: 10.1007/s11423-011-9198-1.

Jaramillo-Morillo, D., Ruipérez-Valiente, J., Sarasty, M. F., & Ramírez-Gonzalez, G. (2020). Identifying and characterizing students suspected of academic dishonesty in SPOCs for

credit through learning analytics. *International Journal of Educational Technology in Higher Education*, 17(1), 1-18.

Kagklis, V., Karatrantou, A., Tantoula, M., Panagiotakopoulos, C. T., & Verykios, V. S. (2015). A learning analytics methodology for detecting sentiment in student fora: A Case Study in Distance Education. *European Journal of Open, Distance and E-learning*, 18(2), 74-94.

Khalil, M., & Ebner, M. (2017). Clustering patterns of engagement in Massive Open Online Courses (MOOCs): the use of learning analytics to reveal student categories. *Journal of computing in higher education*, 29(1), 114-132.

Nkomo, L. M., Ndukwe, I. G., & Daniel, B. K. (2020). *Social network and sentiment analysis: Investigation of students' perspectives on lecture recording*. IEEE Access, 8, 228693-228701.

Northwood, M. D., Northwood, D. O., & Northwood, M. G. (2003). Problem-based learning: From the health sciences to engineering to value-added in the workplace. *Global Journal of Engineering Education,* 7, 157-164.

Paxinou, E., Kalles, D., Panagiotakopoulos, C. T., & Verykios, V. S. (2021). Analyzing Sequence Data with Markov Chain Models in Scientific Experiments. *SN Computer Science*, 2(5), 1-14.

Paxinou, E., Sgourou, A., Panagiotakopoulos, C., & Verykios, V. (2017). The item response theory for the assessment of users' performance in a biology virtual laboratory. *Journal for Open and Distance Education and Educational Technology*, 13(2), 107-123.

Perrenet, J.C., Bouhuijs, P.A.J. & Smits, J.G.M.M. (2000). The suitability of problem-based learning for engineering education: theory and practice. *Teaching in higher education*, 5, 3, 345-358.

Persico, D., & Pozzi, F. (2015). Informing learning design with learning analytics to improve teacher inquiry. *British Journal of Educational Technology*, 46(2), 230-248.

Siemens, G. (2014). Connectivism: A learning theory for the digital age.

Tempelaar, D., Rienties, B., Mittelmeier, J., & Nguyen, Q. (2018). Student profiling in a dispositional learning analytics application using formative assessment. *Computers in Human Behavior,* 78, 408-420.

Trezise, K., Ryan, T., de Barba, P., & Kennedy, G. (2019). Detecting academic misconduct using learning analytics. Journal of Learning Analytics, 6(3), 90-104.

Tsoni R., & Verykios V.S. (2019). *Looking for the "More Knowledgeable Other" through Learning Analytics*. ICODL 2019. 10(3A), 239-251.

Tsoni R., Garani G. & Verikios V. (2022). *Incorporating Data Warehouses into Data Pipelines for Deploying Learning Analytics Dashboards*. In proceeding of IISA 2022. IEEE (to appear)

Tsoni R., Kalles D. & Verykios V. (2022). *A Data Pipeline Approach for Building Learning Analytics Dashboards.* In proceeding of SETN 2022. ACM (to appear)

Tsoni R., Sakkopoulos E., & Verykios S. V. (2021). Revealing Latent Student Traits in Distance Learning through SNA and PCA. *Handbook on Intelligence Techniques in the Educational Process*. Springer (to appear)

Tsoni, R., & Lionarakis, A. (2014). *Plagiarism in higher education: The academics' perceptions.* In 2014 International Conference on Interactive Mobile Communication Technologies and Learning (IMCL2014) (pp. 296-300). IEEE.

Tsoni, R., Panagiotakopoulos, C., & Verykios, V. (2021). Revealing latent traits in the social behavior of distance learning students. *Education and Information Technologies*, 12(2), 25-38. https://doi.org/10.1007/s10639-021-10742-6

Tsoni, R., Paxinou, E., Stavropoulos, E., Panagiotakopoulos, C. T., & Verykios S. V. (2020). *Looking under the hood of students' collaboration networks in distance learning*. The Envisioning Report for Empowering Universities, 39.

Tsoni, R., Samaras, C., Paxinou, E., Panagiotakopoulos, C., & Verykios, V. S. (2019*). From Analytics to Cognition: Expanding the Reach of Data in Learning*. In CSEDU (2) (pp. 458-465).

Tsoni, R., Zorkadis V. & Verykios, V. S. (2021). *A data pipeline to preserve privacy in educational settings*. In Proceedings of the 25nd Pan-Hellenic Conference on Informatics. ACM.

Watkins, J., Fabielli, M., & Mahmud, M. (2020, July). *Sense: a student performance quantifier using sentiment analysis.* In 2020 International Joint Conference on Neural Networks (IJCNN) (pp. 1-6). IEEE.

Zotou, M. (2015). *Enhancing students' skills and capabilities to exploit Open Government Data.* Innovation and the Public Sector, 327.