This is a pre-print version of the following article:
Zotou, M., Tambouris, E. & Tarabanis, K. Data-driven problem based learning: enhancing problem based learning with learning analytics. *Education Tech Research Dev* **68**, 3393–3424 (2020).
https://doi.org/10.1007/s11423-020-09828-8

Data-driven Problem Based Learning: Enhancing Problem Based Learning with Learning Analytics

Maria Zotou[1], Efthimios Tambouris[1], Konstantinos Tarabanis[1]
[1]University of Macedonia, Egnatia 156 Street, Thessaloniki Greece
[1]{mzotou, tambouris, kat}@uom.gr

Conflict of interest: The authors have no conflict of interest to declare.

Short biographies

Dr. Maria Zotou (female) is a Research Assistant at the Information Systems Laboratory at the University of Macedonia, Greece. She holds a PhD (2018) in Problem Based Learning, Learning Analytics and Learning Semantics from the same university, an MSc (2011) in the Use of Information and Communication Technologies in Education from Aristotle University of Thessaloniki, Greece and a BSc in Computer Science (2007) from the University of Macedonia, Greece. Her main research interests include Learning Analytics and data-driven Problem Based Learning

Prof. Efthimios Tambouris (male) is a Professor at the Applied Informatics Department, University of Macedonia, Thessaloniki, Greece. Before that, he served as a Researcher Grade D at the research center CERTH/ITI, at research center NCSR "Demokritos" and the IT industry. He holds a Diploma in Electrical Engineering from the National Technical University of Athens, Greece, and an MSc and PhD from Brunel University, UK. During the last fifteen years he has initiated, coordinated and participated in several international research projects and service contracts funded by the European Union and the Greek public sector. He has served as an expert in standardization activities at CEN and is a reviewer for the European Commission.

Prof. Konstantinos A. Tarabanis (male) is Professor of Information Systems at the Department of Business Administration of the University of Macedonia, Thessaloniki, Greece, where he has taught since 1994, and the Director of the Information Systems Laboratory at the same university conducting research in the fields of electronic government and electronic business. He is also the Head of CERTH/ITI's eGovernment Research Unit. He received an Engineering Diploma in Mechanical Engineering from the National Technical University of Athens (1983), an MS in both Mechanical Engineering and Computer Science (1984 and 1988, respectively), and a PhD in Computer Science (1991), at Columbia University, New York, NY. He was the recipient of the Best Paper Award at the 1991 IEEE International Conference on Robotics and Automation. In recognition of his work in the field of electronic government, he was the recipient of the Best Paper Award in the eGovernment track at the European Conference on Information Systems in 2000.

This is a pre-print version of the following article:
Zotou, M., Tambouris, E. & Tarabanis, K. Data-driven problem based learning: enhancing problem based learning with learning analytics. *Education Tech Research Dev* **68**, 3393–3424 (2020).
https://doi.org/10.1007/s11423-020-09828-8

Abstract

Problem based learning (PBL) supports the development of transversal skills and could underpin the training of a workforce competent to withstand the constant generation of new information. However, the application of PBL is still facing challenges, as educators are usually unsure how to structure student-centred courses, how to monitor students' progress and when to provide guidance. Recently, the analysis of educational data, namely learning analytics (LA), has brought forth new perspectives towards informative course monitoring and design.

However, existing research shows that limited studies have combined PBL with LA to explore their potential in offering data-driven, student-centred courses. This paper presents a framework, termed PBL_LA, that aims to address this gap by combining PBL with LA. The framework is populated from the literature and discussions with PBL and LA experts. The paper also presents results from redesigning, delivering and assessing ten courses in different disciplines and countries using the proposed framework. Results showed positive feedback on all different testing settings, exhibiting reliability of the framework and potential across countries, disciplines and sectors.

Keywords Problem based learning · Learning analytics · PBL model · Course design · Technology enhanced learning

Introduction

Problem based learning (PBL) is a well-established learning strategy that enables active participation of students who "learn by doing" and supports the development of transversal and lifelong learning skills (Sohmen 2020; Zhou and Zhu 2019). When PBL is reinforced with the utilization of collaborative Web technologies in blended settings, termed PBL2.0 (Tambouris et al. 2012), students can use diverse tools in order to more effectively perform the required tasks in solving their problems, which leads to the generation of large amounts of data (Ünal 2019; Zotou 2015). However, educators can rarely make sense of what this data entails for the progress of the course and what relevant decisions can be made. The application of PBL in courses faces other challenges as well, since educators usually feel it is not that easy to change their teaching style to the PBL format (Chen et al. 2020). During this process, they are usually unsure of each student's learning progress, contribution to the group work and need for assistance. This limits their ability to provide fair assessment, ongoing scaffolding and reduce the drop-out numbers (Chen et al. 2016).

An interesting emerging field that could address these challenges is learning analytics (LA). LA methods and tools analyse data generated during learning and provide informative insights on the learning process (El Alfy et al. 2019). This can in turn

empower educators in becoming more aware of students' progress, assessing their contributions based on evidence-based criteria and identifying patterns of low engagement and at-risk failures (Foster and Siddle 2019; Wong 2019).

Although this combination of PBL with LA is potentially interesting, relevant studies are limited. As a result, the academic literature does not provide a clear reference framework for this field, i.e. a structure that underlines all basic elements and provides guidance for future application. Such a framework could contribute in understanding PBL's combination with LA and in redesigning courses with less risks and better chances of success by exploiting the insights of LA within PBL.

This paper aims to construct a framework that combines PBL with LA to assist educators in designing and delivering more adaptable, data-driven and student-centred courses. The framework aims to bridge the gap between promising pedagogical and technological solutions and to empower educators to reap the benefits of employing LA within PBL.

The rest of the paper is structured as follows. The research methodology is presented, followed by literature review results on PBL, on LA, and on their combination. We then proceed to describe the proposed PBL_LA framework layers and their content. The empirical evaluation of the framework follows, where course design decisions are made per layer, a web-based application developed to access the framework's content is presented, and the application of the framework in redesigning ten courses and the corresponding evaluation is described. Finally, the paper presents the conclusions drawn and future work.

Methodology

The methodology followed in this paper is divided into three main phases, as shown inFig. 1.

Each phase aims to contribute to the achievement of the paper's aim, i.e. in understanding how PBL can be combined with LA for the redesign of courses with less risks and better chances of success. Thus, the methodology steps cover both pedagogical and technological aspects that will allow educators to reap the benefits of employing LA within PBL.

Phase A: Literature review

The first phase involves conducting literature reviews on three domains: PBL models, LA and combining PBL with LA.

This is a pre-print version of the following article:
Zotou, M., Tambouris, E. & Tarabanis, K. Data-driven problem based learning: enhancing problem based learning with learning analytics. *Education Tech Research Dev* **68**, 3393–3424 (2020).
https://doi.org/10.1007/s11423-020-09828-8

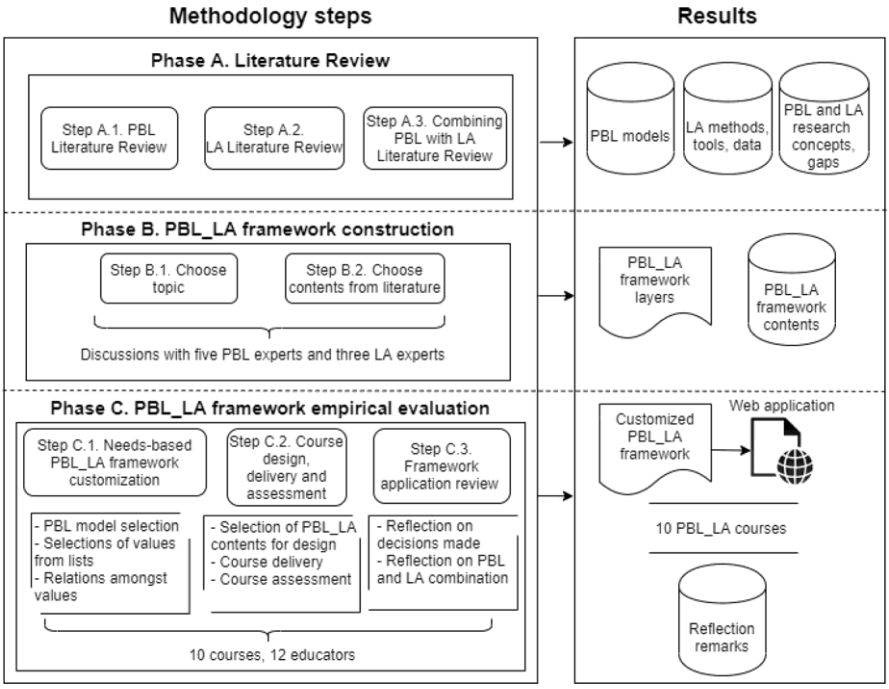

Fig. 1 Methodology for PBL_LA framework

Step A.1: PBL literature review

A literature review was carried out on existing PBL models. A PBL model is an instructional methodology that has been designed to provide guidelines on how to design, deliver and assess courses using the PBL method. As the PBL domain is mature and relevant research on PBL models has been carried out over the years, we based our research on existing reviews or articles where PBL models were reviewed as part of the relevant study. This culminated in considering the work by Wijnia et al. (2019) and Zotou and Tambouris (2014), where we examined the steps of each model, the learning processes followed and the different monitoring/assessment methods used.

Step A.2: LA literature review

A literature review was carried out on LA, aiming to retrieve general information on the domain (e.g. LA terms, LA steps), and to identify the most representative examples of howLA is applied (e.g. LA methods, LA tools, data analysed). As the field has been of great interest in the research community for more than 10 years, a large number of relevant papers was identified in Web of Science and Scopus (more than 2000). The selected papers (78 in total) were isolated for further study based on publication date (later than 2007), research level (we preferred literature review papers over primary

research papers), relevance (papers focusing mainly on the search terms), volume of information provided (papers that analyse the search terms in depth) and language (English). Out of the 78 studied papers, 31 are cited in this paper, as they covered the whole field satisfactory by providing complementary information with minimum overlapping.

Step A.3: Combing PBL with LA literature review

The final literature review was carried out on research papers where LA has been employed within PBL educational and training settings. This review was conducted in Web of Science and Scopus using as keyword PBL AND "Learning Analytics". The search revealed four papers in Web of Science and 15 in Scopus, which were read for relevance to the scope of this study. In addition, we used the *forward* and *backward* reference searching to identify additional papers. This process resulted in retaining six papers. The remaining papers were not included in this study as their scope was not relevant to our aims, e.g. they focused on standalone LA tools' development, they presented mathematical models and algorithms for dataanalysis etc.

Phase B: PBL_LA framework construction

The second phase involves the construction of the framework, i.e. its layers and their contents. The research methodology employed in this phase is adapted from work by McGaghie et al. (2001) on the design of conceptual frameworks. The resulting framework was intendedto be abstract thus accommodating any PBL model or even any other collaborative learning strategy. The concepts identified in the literature were included in the proposed framework thus ensuring the framework is in line with established relevant theories and existing research. This step was conducted with the assistance of *experts*, i.e. academics with extensive experience in teaching PBL courses and performing research on data analytics. More specifically, the group of experts included five PBL experts (from Denmark and the Netherlands) and three LA experts (from Spain and the Netherlands), while one of the authors participated as facilitator and contributor. The steps that were followed to construct the framework were:

Step B.1: Choose topic: decision on which topic(s) the research will focus on

The work focuses on the combination of PBL with LA features.

Step B.2: Choose concepts from literature review

The layers of the framework were derived from the results of the literature review performedin the first phase. The contents of each layer were derived from the literature as well as brainstorming sessions with the PBL and LA experts. Each session focused on one framework layer and aimed to answer the questions "How can this layer be populated with instances for additional guidance to educators?" and "How are these instances related to the other layers?"

This is a pre-print version of the following article:
Zotou, M., Tambouris, E. & Tarabanis, K. Data-driven problem based learning: enhancing problem based learning with learning analytics. *Education Tech Research Dev* **68**, 3393–3424 (2020).
https://doi.org/10.1007/s11423-020-09828-8

Phase C: PBL_LA framework empirical evaluation

The third phase of the methodology includes the empirical evaluation of the framework's usage. In this paper, the framework was customized to meet the specific evaluation needs. It is noted however that different customizations are possible thus enabling researchers to use and evaluate the proposed framework under different conditions.

In our case, for evaluation purposes, twelve *educators* used the PBL_LA framework to redesign, deliver and evaluate ten different courses (presented in Table 1). The group of educators included six of the experts who participated in the previous phase, two educatorsfrom Greece and three from Austria. The use of multiple courses and different educators enabled us to derive more reliable results. Each step is now outlined.

Step C.1: Needs-based PBL_LA framework customization

In this step, the framework was customized to accommodate the needs of educators, learners and courses. An important decision was related to the choice of a specific PBL model. In this research, the Aalborg PBL model was selected, as some educators were already familiar with its use, while additional experts were also available for consultation, if needed. Consequently, the PBL_LA framework's activities, LA methods, data and ICT tools were also selected. In addition, the overall LMS was selected for each course, which included Moodle, JIRA, and yOUlearn. In those selections, the courses' context, the profile of students, and the courses' educational objectives were considered along with educators' familiarity with specific activities, LA methods and ICT tools. It is important to note that those selections should not be too overwhelming for both students, as their active participation is required, and educators, as they are required to observe, scaffold and adapt students' interactions throughout the course (Ørngreen et al. 2019). Finally, a web-based application was developed that allowed educators to browse the contents of the framework in order to make informative decisions when designing the courses.

Step C.2: Course design, delivery, and assessment

In this step, the educators redesigned their courses with the assistance of the framework, i.e. organized the PBL-oriented activities, launched all ICT tools and delivered the courses to the students. In this paper, due to space limitations, we report details on just one course, i.e. C1 from Table 1, which was delivered by one of the authors with the assistance of the others. The traditional structure of the course was transformed into the PBL model by mapping weekly lectures into PBL steps and identifying suitable PBL activities. In addition, wedelivered the course and assessed each PBL step by consulting the LA tools.

This allowed us to intervene when necessary, adapt the course as required (e.g. provide more content, encourage more active participation, change questions in complex quizzes etc.) and change our design decisions if needed (e.g. choose different activities if some

were not preferred, choose other LA tools if some did not provide helpful information etc.).

Step C.3: Framework application review

In this step, the framework's application in real world settings was reviewed. The review followed the paradigm of reflective writing from the study of Vigentini et al. (2016). Based on this methodology, each educator that designed a course reflected on the process, analysed what happened during the course execution, and scheduled action plans for future courses. All educators followed the same review process. A qualitative evaluation approach was selected as a means to allow deeper understanding of educators' opinions and experiences and thus draw more meaningful conclusions (Chalhoub-Deville and Deville 2008).

Table 1 Courses redesigned with the PBL_LA framework

Course	Course description	Country	Course level	Number of students	Duration
C1	Information Systems Analysis and Design	Greece	Postgraduate	32	13 weeks
C2	Project management course	Greece	Undergraduate	40	13 weeks
C3	Information Systems Analysis and Design	Greece	Undergraduate	29	13 weeks
C4	Audio-Visual Experiments	Denmark	Undergraduate	16	16 weeks
C5	Human–Computer Interaction	Denmark	Undergraduate	94	16 weeks
C6	Advanced Software Engineering at Computer Science	Spain	Undergraduate	42	15 weeks
C7	Modelling Training	Austria	Private training	12	2.5 days
C8	Task in Online study course	The Netherlands	Undergraduate	20	2 weeks
C9	Master of Science in Educational Science	The Netherlands	Postgraduate	41	6 weeks
C10	Creative Play—Applied Technology	Denmark	Undergraduate	69	8 weeks

Additionally, the research's aim to understand how the combination of PBL with LA is perceived and practiced calls for a qualitative approach, where participants' reflections can help us understand which features of the framework were successful and which require improvement (Leung 2012).

The review from the educators focuses on examining how they combined PBL with LA using the PBL_LA framework. The aim of this review was to determine how the educators used the PBL_LA framework and applied PBL and LA features in their classrooms, to report possible benefits and challenges, as well as to identify possible future plans and recommendations for improvement. It should be noted that all educators have medium to advanced technical background and are somewhat or highly

experienced in the PBL strategy, albeit not necessarily using the Aalborg model. The results were analysed manually (i.e. no content analysis tool was used) due to the moderate size of relevant transcripts. Selecting self-reflection suggests that if some elements of the PBL_LA approach did not work as expected, valuable knowledge could still be gained. This also provides useful guidance for other educators who want to adapt their courses by combining PBL with LA.

Study limitations

We acknowledge that this study has several limitations. First, the proposed framework is based on the analysis of studies written in English. Second, searches for studies were conducted in only two scientific databases, namely Web of Science and Scopus. Third, we acknowledge that subjectivity in the framework's design constitutes an additional imitationof the study. Although we relied on literature review and held discussions with several PBL and LA experts, we had to make subjective choices regarding e.g. the number and names ofthe framework's layers and the number and names of the concepts within each layer. Additionally, the medium to advanced technical knowledge and PBL experience of the participating educators could have contributed positively to their evaluation remarks. Finally, the lack of a quantitative evaluation of the framework by the educators and learners may have limited our research findings on how and to what degree was the framework able to facilitate them in their PBL_LA course transformation. These limitations do not impact directly on the proposed framework as an appropriate guide of PBL_LA course design. However, they do indicate scope for further research.

Literature review

PBL

PBL is a student-centred learning strategy which aims to educate students through solving problems (Neville 2009). This strategy has been applied over the last 50 years in multiple educational institutions and different domains. A series of models have been proposed by universities across the world, which structure the PBL method into specific steps. These models aim to help educators design, deliver and asses their courses. The results of the conducted review regarding PBL models are presented in Table 2.

This table suggests all models include steps regarding the analysis of the problem and the formulation of ideas related to specific objectives. Additionally, in all PBL models students are required to form a solution and, apart from Maastricht and Samford, in all other models students discuss findings and share their research amongst the group. Finally, apart from Manchester, all other models require students to evaluate, report and/or defend their work.

This is a pre-print version of the following article:
Zotou, M., Tambouris, E. & Tarabanis, K. Data-driven problem based learning: enhancing problem based learning with learning analytics. *Education Tech Research Dev* **68**, 3393–3424 (2020).
https://doi.org/10.1007/s11423-020-09828-8

Table 2 Review of PBL models

Higher education institute	**Course(s)**	**Steps**	**Assessment**	**Skills**
Aalborg (Kolmos et al., 2004)	Project Management	**8 steps** (Group formation, Problem formulation, Task formulation, Problem delimitation, Solution, Discussion, Implementation, Evaluation)	Group-based assessment with individual grading	Knowledge processing, Analytical thinking, Argumentation, Communication of ideas, Group-work
McMaster (Saarinen-Rahiika and Binkley, 1998)	Medicine	**7 steps** (Objectives identification, Interaction with the scenario, Identification of self-study questions, Self-directed study, Discussion, Review and synthesis, Evaluation)	Self-assessment, Peer assessment, Tutor assessment	Problem solving, Group-work, Self-directedness, Communication
Maastricht (Schultz and Christensen, 2004)	Science, Healthcare, Business etc	**7 steps** (Setting clarification, Problem definition, Case investigation, Problem re-structure, Learning goals formulation, Individual learning, Report)	Performance in the problem-solving process	Presenting viewpoints, Debating, writing texts, Working together
University of Newcastle, Australia (Neame, 1989)	Medicine	**8 steps** Cue recognition, Initial formulation, Hypothesis generation, Hypothesis organization (possible mechanisms), Inquiry strategy with recursive cycles, Problem reformulation, Final formulation, Diagnostic Decision	Individual and group assessment	Reasoning skills, critical thinking, problem solving

Southern Illinois University (Koschmann et al., 1994)	Medicine	**5 steps** Problem formulation Self-directed study Problem reexamination Abstraction Reflection	Individual, peer and group assessment	Communication of ideas, presentation, teamwork, synthesis of information, reflection
Manchester (David et al., 1999)	Medicine, Engineering,	**8 steps** (Terms clarification, Problem definition, Hypotheses brainstorming, Arrangement of ideas, Learning objectives definition, Information gathering, Results sharing, Discussion experience)	Individual, peer and group assessment	Problem-solving, Teamwork, Communication
Samford (Mauffette and Poliquin, 2001)	Business, Education, and Pharmacy	**7 steps** (Problem analysis, Conceptualization, Prioritization of hypotheses, Plan identification, Data collection, Hypotheses verification, Defence)	Reflection and peer assessment	Critical thinking, Problem solving, Decision making
Delaware (Allen et al., 2003)	Biochemistry, Biology, Science etc	**5 steps** (Problem analysis, Information identification, Sharing research findings, Formulate solution, Evaluate)	Evaluation forms	Problem-solving, Research, Social skills, Critical thinking, Writing

Learning analytics

Learning analytics (LA) is defined as "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" (Long and Siemens 2011). The field of LA has emerged from and is closely connected to multiple and different research fields and areas related to analysis, such as business intelligence, statistics, web analytics, academic analytics, data mining, Social Network Analysis (SNA), as well as research interest in the field of learning sciences such as pedagogies, Technology Enhanced Learning, cognitive sciences etc. LA is strongly related to learning technologies ranging from cognitive toolsto more sophisticated and complex environments, such as Learning Management Systems (LMSs), Virtual Learning

Environments and the recent Massive Open Online Courses (MOOC), that generate large amounts of educational data.

The LA domain can thus reinforce education and training through providing feedback based on generated data and allowing an in-depth understanding of the learning experience (Wong 2019). This can be done by accumulating as much educational data as possible and enabling students and educators/trainers to comprehend the information provided and make decisions in regards to the learning process, learners' knowledge and skills as well as more easily identify students' weaknesses and misconceptions, the assessment's efficiency etc. All these insights can then underpin successful personalized and adaptive learning that improve all aspects of education and training (Gong and Liu 2019).

Table 3 shows the results of the literature review carried out regarding how LA can be applied, i.e. research on LA methods, LA tools and educational data that can be analysed. Table 3 reveals that LA research can be structured around analysis methods used, ICT tools employed and underlying data processed for analysis purposes. The decision of which methods, data and tools are relevant in each case depends on the context of the course, the availability of online learning technologies as well as the educational objectives set by the educators (Picciano 2012).

Combining PBL with LA

The literature review revealed six empirical studies where LA was used in the context of PBL. The study by Saqr et al. (2020) aims to investigate which interactivity factors can improve monitoring and student support in online PBL and whether these factors can predict student performance. Towards this goal, the authors gather Moodle data and analyse it using SNA. The study concluded that SNA analysis of interactions and participation enables predicting performance in groups and supports students with limited participation and interactions.

The study by Spikol et al. (2018) focuses on applying machine learning and LA methods on educational data deriving from diverse sensors (computer vision, user-generated content and data from the learning objects) during PBL. The authors present an LA dashboard that was developed to visualize the results and help educators determine whether groups are performing well. The study concludes that the analysis of diverse data can provide interesting insights for educators and help them make more informed decisions on how to assist their students.

Table 3 LA domain review

LA concept	Author	Examples
LA methods	(Pistilli et al., 2014), (Dyckhoff et al., 2013), (MacNeill et al., 2014) (Ferguson, 2012), (Papamitsiou and	Learner modelling (learner profile, behaviour modelling, natural language processing), Interventions in learning

	Economides, 2014), (Siemens, 2013), (Dimitracopoulou, 2015), (Elias, 2011), (Pardo, 2014), (Buckingham, 2011), (Chatti et al., 2012), (Verbert et al., 2013), (Baker and Inventado, 2014), (Herskovitz et al., 2013), (Dyckhoff et al., 2012), (Dimopoulos et al, 2013), (Prinsloo et al., 2012), (Long and Siemens, 2011), (Otte and Rousseau, 2002)	(adaptation, predictions, mentoring, personalization), Relationship mining (sentiment analysis, discourse analysis), association rule mining, adaptive content to learners, recommendations on content, activities and interactions changes in behaviour, knowledge domain modelling, SNA, semantic analysis, clustering, information flow analysis, early risk identification, Assessment (monitoring, guiding, scaffolding, feedback, reflection)
LA tools	(Picciano, 2012), (West, 2012), (Leony et al., 2012), (Mazza and Dimitrova, 2007), (Fortenbacher et al., 2013), (Santos et al., 2012), (Ali et al., 2012), (Dyckhoff et al., 2012), (Mazza et al., 2012), (Dimopoulos et al, 2013)	GLASS, SNAPP, LeMo application, StepUp, LOCO-Analyst, NetlyticeLAT, Gismo, MOCLog, Learning Analytics Enhanced Rubric, SmartKlass, Engagement analytics, Analytics and recommendations, Configurable reports, Adaptive quiz
Educational data	(Pistilli et al., 2014), (Dyckhoff et al., 2013), (MacNeill et al., 2014,) (Ferguson, 2012), (Papamitsiou and Economides, 2014), (Van Harmelen and Workman, 2012), (Elias, 2011), (Pardo, 2014) , (Siemens, 2013), (Dimitracopoulou, 2015), (Romero and Ventura, 2013)	Activities accessed / used, Posts on forums, Number of participants, Clusters of students who made mistakes, Contributions to shared documents , Social media posts and interactions (replies, shares, tags), Time spent, Performance in assignments / activities / quizzes, Grades, Frequency of interactions

The study by Tempelaar et al. (2015) investigates teaching and learning of mathematics and statistics in a blended learning environment. The study used the Maastricht PBL model, where students formed groups and were coached by an expert. Students' engagement with online technologies was optional as this is more in line with the Maastricht model. Regarding LA, the data collected included frequency of using the practice tests, time spent on practice tests, number of attempts to solve a problem etc. The case study concluded that the usage of the online environments as a complimentary tool to PBL proved to help students, as tools supported self-direction, reflection and decision making. As students of PBL are usually new at this learning model, where they hold the majority of responsibility to gain knowledge, it seems that visual feedbacks on how self-directed learning is proceeding has made them more confident in this control shift.

The study by Göhnert et al. (2014) reports research carried out on a workbench that was developed for analysing and visualising data. The study describes three cases where the workbench was tested in educational settings, in which trainees generated

cases, collaborated, shared and discussed the cases in small groups, while the trainers guided the process. No specific information is provided in the study regarding the effectiveness of the solution and actual LA outputs from the doctors' PBL practices.

Kotsiantis et al. (2013) used a PBL approach where students were required to submit problem-based assignments. The employed blended learning approach did not rely on a specific PBL method and included LA features such as visualizations, decision trees, class association rules and clustering. No actual conclusions were formed regarding the combination of PBL with LA; the case study however pointed out that students' performance seemed to be highly connected to their negative or positive perceptions of the LMS used during the course.

The last study examined how PBL is carried out in lab sessions with large number of students (Rojas and Garcia 2012). Groups of students worked together to solve specific problem statements and a supporting LA tool was developed to provide visualizations from the data generated. The data gathered for analysis included observational data (e.g. timestamps or questions, answers, time devoted to a problem etc.) and questionnaire data (e.g. number of questions asked in a session, fairness of the time devoted to each group by teacher etc.). This data was then analysed using statistical analysis. The study's conclusions focus on the developed tool's functionalities and visualizations provided, while also stating the promising improvement of PBL through diverse visualizations of educational data.

Table 4 presents an overview of the main concepts covered in each study, which provide us with substantial material for constructing the PBL_LA framework. The Table also presents limitations of each study regarding combining PBL with LA and how the PBL_LA framework should address these limitations and become a more structured and holistic reference tool for educators.

Table 4 Studies that have combined PBL with LA

Study reviewed	Study concepts covered	Study limitations	PBL_LA proposed contribution
(Saqr et al., 2020)	Tools for students LA methods Data	No specific PBL model used. Only SNA as LA method. No information on LA tools.	All PBL models must be easily accommodated. Educators must be informed on a variety of methods and tools.
(Spikol et al., 2019)	LA methods Data LA tool	No specific PBL model used. No PBL tools used.	All PBL models must be easily accommodated. A variety of PBL tools must be included.
(Tempelaar et al., 2015)	PBL model (Maastricht steps) Tools for students (optional) Data gathered	Optional use of technologies / limited data gathering. Data gathered restricted to quizzes.	LA features and visualizations must be incorporated in each PBL step. A variety of PBL tools must

	LA tools for progress feedback	Limited LA visualizations for overall performance across each phase.	be included for students to use.
(Göhnert et al., 2014)	LA tools Data	Independent platform, requires gathering of data from external resources, no definite success results in academia.	A variety of PBL and LA tools must be included that are or can be integrated in existing e-learning platforms.
(Kotsiantis et al., 2013)	LA methods LA tools LMS technology Data gathered	No specific PBL approach to guide educators. Analytics performed used complex software and not easy to use tools.	Specific PBL steps for course design must be accommodated. A variety of easy to use LA tools must be included.
(Rojas et al., 2012)	LA tool for visualizations Data gathered	No step-by-step PBL approach to guide educators. Analytics performed used observation and simple statistical analysis of questionnaires.	Specific PBL steps must be included so that even large numbers of students can follow a specific step-by-step learning process. A variety of easy to use LA tools must be included.

The above studies conclude that LA insights during PBL proved to support students and educators, as they helped in self-direction, reflection and decision making. Students of PBL are usually new at this learning model, where they hold most of the responsibility to gain knowledge, and it seems that visual feedbacks on how the self-directed learning is proceeding has made them more confident in this control shift.

However, these case studies mostly follow the basic principles of PBL (e.g. students form groups and collaborate in any way to solve a problem) without necessarily adopting an established PBL model and do not relate the LA results with specific PBL steps. Consequently, educators face difficulties in identifying what part of the course needs adapting and have no guidance in what kind of adaptations they can make (e.g. choosing PBL activities, tools, LA visualizations etc.).

Finally, all papers report on findings from specific case studies, hence they do not propose a structured guideline for other educators to use in order to successfully combine PBL and LA.

This is a pre-print version of the following article:
Zotou, M., Tambouris, E. & Tarabanis, K. Data-driven problem based learning: enhancing problem based learning with learning analytics. *Education Tech Research Dev* **68**, 3393–3424 (2020).
https://doi.org/10.1007/s11423-020-09828-8

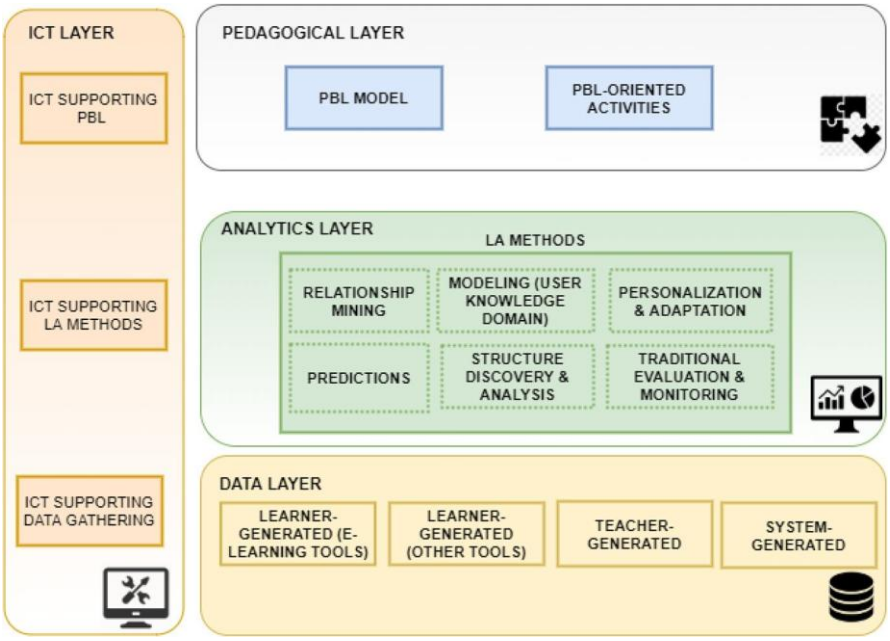

Fig. 2 PBL_LA framework

PBL_LA framework

PBL_LA framework overview

The proposed PBL_LA framework includes three main layers and one horizontal supporting layer, as shown in Fig. 2.

The following sub-sections outline the contents of each layer and their relationships.

Pedagogical layer

The *Pedagogical layer* consists of the *PBL model* and the *PBL-oriented activities*. The *PBL-model* refers to the specific PBL model to be employed when using the framework. The proposed PBL_LA framework does not prescribe the use of a specific PBL model; instead, it can be used with any PBL model. Actually, when using this framework, educators need to first select a specific PBL model that best suits their courses' educational needs. This will determine the PBL steps to be followed, which are subsequently related to the rest of the framework's contents. The *PBL-oriented activities* refer to specific educational activities that can be applied to support the

application of the PBL model. Annex 1 includes a list of activities identified in the literature and discussions with PBL experts.

Analytics layer

The *Analytics layer* consists of six main *LA method groups* available for gathering, processing, analysing and interpreting data into meaningful information. These LA method groups generate insightful visualizations for both educators and learners enabling them to exploit the analytics results. Each group is supplemented with a list of specific *LA methods*. For example, the LA method group "Structure discovery and analysis" includes specific LS methods, such as Social network analysis, Information flow analysis, Semantic analysis, and Clustering. The list of LA method groups and specific methods is shown in Annex 1. This list was derived from the literature and discussions with LA experts. The formulation of LA methods groups was adopted from the work reported by the co-founder and president of the Society for Learning Analytics Research (SOLAR), who is considered to be the instigator in the LA frameworks research (Siemens 2013).

Data layer

The *Data layer* consists of the four main types of data that can be generated during a blended PBL course. The specific data that can be gathered greatly affects the method of processing, analysis and visualisation to be employed within a course. This data is usually generated by learners when using an e-learning platform, by learners when using other ICT tools that are not specifically used for learning, by educators, and by the e-learning environment itself based on the interactions recorded. Possible contents of the Data layer are shown in Annex 1. The list is an extension of the educational data identified in the LA literature review, after discussions with all educators on what kind of data is usually generated during their courses as well as what data they believe is important to gather for students' progress monitoring.

ICT layer

The *ICT layer* is horizontal across all three main layers, as it aims to provide technological support for each layer's components. More specifically, this layer includes tools that are usually employed during a PBL course to scaffold students' engagement, tools that can be used for gathering and storing data generated during learning and tools that can be used to support LA methods and visualizations. The utilization of the PBL_LA framework should take into consideration ethical issues when gathering, processing and analysing data (e.g. anonymisation, cleaning, consent, privacy, transparency etc.). Thus, this layer includes tools that can support the PBL-oriented activities (e.g. social bookmarking, mind maps, forum, wiki etc.), LA tools that can be applied for analysing the generated data (e.g. GISMO, Analytics graphs, SNAPP etc.) and tools that support data gathering (e.g. Moodle, Blackboard, Trello, Twitter, MOOC etc.) as shown in Annex 1.

Relationships between layers

Important relationships exist amongst the framework's layers. These relationships can provide important recommendations to educators during course design. For example, a specific PBL model's step can involve a sub-set of the listed PBL-oriented activities, such as brainstorming, literature storing, argumentation, writing etc. (Pedagogical layer). Each of these activities can in turn be executed through various PBL tools, such as forums, social bookmarking, mind maps, wikis etc. (ICT layer). Similarly, educators can detect which LA methods (Analytics layer) and LA tools (ICT layer) are more suitable for providing helpfulinsights to themselves and to learners.

PBL_LA framework empirical evaluation

This section presents the framework's use and review using the ten courses presented in the "Methodology" section. We start by presenting the needs-based framework customization (Methodology Step C.3) for all ten courses. We proceed by illustrating the design, delivery and assessment (Methodology Step C.2) of just one course, namely course C1 from Table 1, due to space limitation. Finally, the review by reflection of the framework's usage (Methodology Step C.3) includes remarks by all educators of all courses, providing a more comprehensive overview of how multiple educators combined PBL with LA. The work conducted in each step along with relevant results are now outlined.

Step I: Needs-based customization

A summarized overview of the decisions made per layer follows.

Pedagogical layer

PBL model As mentioned in the methodology, amongst the PBL models reviewed, the Aalborg PBL model was selected for all ten courses.

Model steps The nine steps of the Aalborg PBL are outlined in Annex 2.

PBL-oriented activities The educators decided that the most relevant activities for the ten courses are group creation, literature searching, writing, presenting, report writing, report submitting, tasks allocation, conducting surveys, and reflecting.

Analytics layer

Educators of all ten courses consulted the list of available LA methods and identified thosethat allowed them to adapt their course based on students' engagement levels and to identify students in risk of failing so they can intervene promptly. As a result, the list of selected LA methods included adaptive content to learners, recommendations on content, activities and interactions, early risk identification, interventions, exam grades, assignments grades, presentations performance, and performance discussions.

Data layer

Three different e-learning environments were chosen for the courses. More specifically, six courses were delivered in the Moodle LMS, three courses in one of the

university's local e-learning platform (namely yOUlearn) and one course in JIRA. The data gathered include: *Student-generated (e-learning environments)* Assignments submitted, times of assignments submissions, access to learning resources, posts on forums, performance in assignments, performance in quizzes, survey data, task checklist activity.

Student-generated (Other tools) No other tools (e.g. social media, project management) were used to monitor students' participation in the course to reduce technical overload, thus no relevant data were generated.

Teacher-generated Final exam grades, forum posts.

System-generated Times spent on each unit of learning, navigation patterns, frequency of logins, activities accessed/used, number of participants per group, engagement levels.

ICT layer

The tools selected within the ICT layer regard PBL and LA tools that were subsequently launched within the three e-learning environments used to support the ten courses. These include:

PBL tools Forum, Wiki, Feedback, Quiz, Folder, Assignment, Social bookmarking, Student folder, Tasks checklist.

LA tools GISMO, Administration reports, Analytics graphs (Quiz submission, Wiki access, Folder access), Adaptive Quiz, Feedback responses analysis graph, Activity results(Groups with the highest average), Files uploaded, Forums graph, Anaconda tool for data analytics.

Data gathering tools Moodle, JIRA, yOUlearn (local university e-learning platform).

Relationships between layers

Once all framework layers were populated, the relationships between steps, activities, methods and tools were identified. For example, during the PBL *Problem formulation* step in the Danish University of Table 1, students perform activities such as literature search, literature storing, brainstorming, argumentation, and writing (Pedagogical Layer). Thus, the Analytics Layer was populated with relevant LA methods that are suitable for providing helpful insights to educators and learners. The discussions amongst experts revealed the most relevant LA methods for this PBL step, namely adaptive content, recommendations, warnings and mentoring.

As another example, discussions with educators concluded that the more relevant PBL-oriented activities for the Problem formulation step include brainstorming, literature searching, literature storing, argumentation, writing, and presenting. Another example includes the relationships of activities with ICT tools, e.g. Brainstorming can be carried out through a forum, a mind-map, Mindmeister etc.

This exercise culminated in 3862 relationships between "PBL step-activity-LA method

data-PBL tool-LA tool", making access to all this information challenging.

Consequently, a web-based application was developed that allowed educators to browse the framework contents in order to make informative decisions when designing the courses. For this purpose, the framework's content and relationships were recorded into a relational database. In addition, three main drop-down lists were created, which present PBL model'ssteps, activities and tools, as shown in Fig. 3.

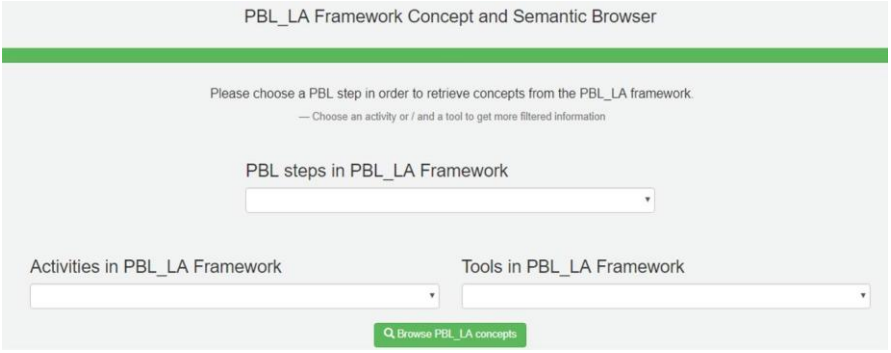

Fig. 3 PBL_LA browser interface (https://egov.dai.uom.gr:8080/RESTfulProject/Home.html)

Educators could choose a PBL step and view which other elements (i.e. activities, PBL tools to use, LA methods, data generated, LA technologies) they could consider when designing their courses for that PBL step. For more filtered results, users can specify an activity and/or a specific PBL tool they want to employ. This way, educators that wantto design, for instance, the *Problem Formulation* step are now aware of which types of activities can be employed (e.g. brainstorming, literature searching, literature storing etc.) and which ICT tool can support each PBL activity (e.g. forum, Google Docs, Mindmap etc.).

Activity	PBL tool
Writing	Google Docs
Writing	Wiki
Development	Google Docs
Development	Wiki
Production/testing	Google Docs
Production/testing	Wiki
Experimenting	Dabbleboard
Modeling	Dabbleboard
Reflecting	Quiz

Data generated	PBL tool
Frequency of log ins	Google Docs
Number of participants per group	Google Docs
Performance in quizzes	Google Docs
Frequency of interactions / engagement	Google Docs
Participation in group work	Google Docs
Survey data	Google Docs

LA Tool	LA method
CourseVis	Scaffolding
Google Analytics	Data visualization
Google Analytics	Data analysis
Google Analytics	Reflection
Google Analytics	Guidance
Google Analytics	Engagement analysis
Google Analytics	Feedback
Google Analytics	Scaffolding
LeMo	Association rule mining

Fig. 4 PBL_LA browser information for PBL step Design

We should note here that all educators were experienced in the use of technologies and were eager to adopt novel pedagogical and technological solutions. Therefore, this study's educators cannot be considered as representative sample. In real-life settings, educators would probably need training on the use of the PBL_LA framework and the relevant activities, methods and ICT tools. Investigating this aspect is however outside

This is a pre-print version of the following article:
Zotou, M., Tambouris, E. & Tarabanis, K. Data-driven problem based learning: enhancing problem based learning with learning analytics. *Education Tech Research Dev* **68**, 3393–3424 (2020).
https://doi.org/10.1007/s11423-020-09828-8

the scope of this study.

Step II: Course design, delivery and assessment

Course design

In this step, we report the results of the design, delivery and assessment of a specific course, namely C1 from Table 1. We commenced by visiting the web application and received advice on which PBL activity, LA method and ICT tool could be used in each PBL step. For example, the information provided for the PBL step *Design* are illustrated inFig. 4.

Figure 5 illustrates how the course was deployed in Moodle based on the framework's suggestions.

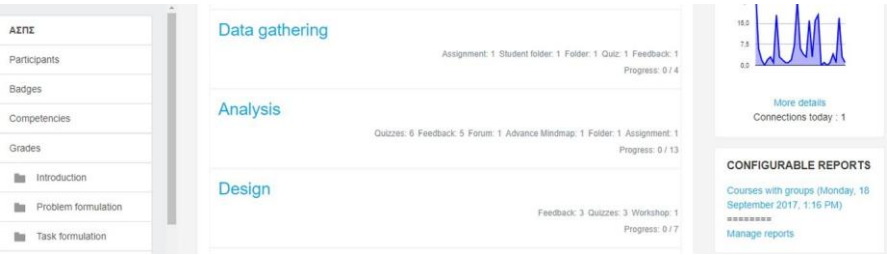

Fig. 5 Course delivery in Moodle based on the PBL_LA framework

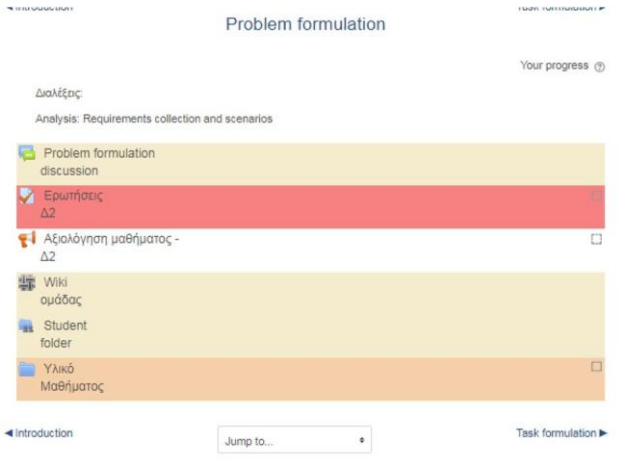

Fig. 6 Problem formulation step for C1 course

More specifically, each step includes a set of tools that allow the execution of the suggested PBL activities. That way, students were provided with the necessary means to learn by doing.

This is a pre-print version of the following article:
Zotou, M., Tambouris, E. & Tarabanis, K. Data-driven problem based learning: enhancing problem based learning with learning analytics. *Education Tech Research Dev* **68**, 3393–3424 (2020).
https://doi.org/10.1007/s11423-020-09828-8

Course delivery

Once the e-learning environment was set-up based on the design decisions, we delivered the course to the students. Figure 6 shows all different ICT tools that students were able to use within the course for the *Problem formulation* PBL step.

This figure shows that students were provided with a forum where they could discuss about the formulation of their problem and a wiki where they could document their work for review and monitoring. A folder included all learning material that was necessary for the execution of this PBL step, such as files on the analysis, requirements collection and scenarios design. Additionally, students could upload files relevant to their work in a "Student folder" tool and could reflect on the weekly course through a feedback tool. Finally, a quiz was provided to monitor their comprehension levels of thetaught materials.

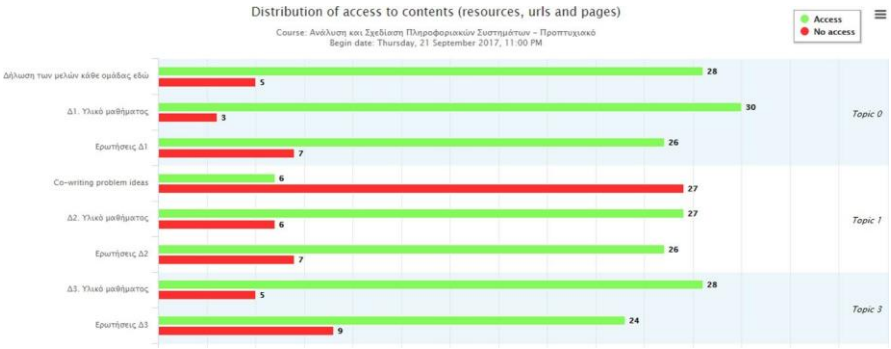

Fig. 7 Access of content frequency

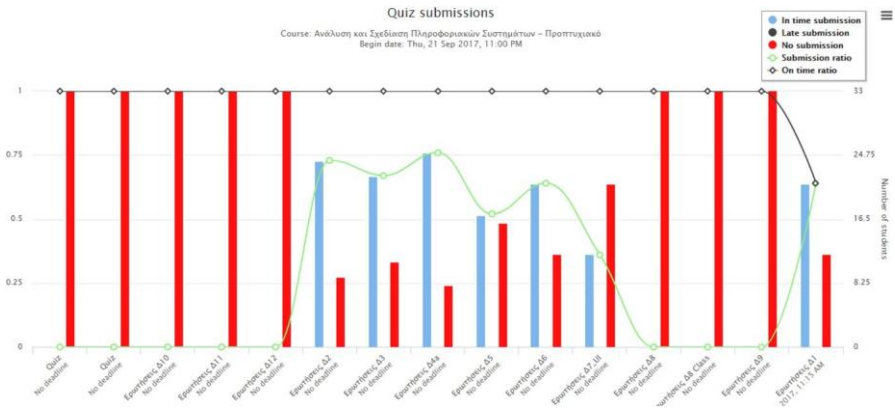

Fig. 8 Quizzes submission graph

Course assessment

Finally, we assessed the course for each PBL step, by consulting the LA plugins and relevant data visualizations based on the students' interactions with relevant activities.

As an example, Fig. 7 shows the total number of students that have accessed (top bar) and have not accessed (bottom bar) specific content or activities within the course. This urged us to contact students and encourage them to visit specific parts of the platform that include essential material to improve their performance.

Figure 8 shows how many and which students submitted each given quiz and whether the quizzes were submitted on time, late or not at all. This helped us to easily identify students that did not participate actively in each activity or had low performance and intervene accordingly, e.g. by adapting the quiz if it was too difficult, changing the material if the majority did not comprehend it etc.

Step III: PBL_LA framework application review

The review of all delivered courses is divided into three reflection steps, namely Remarks, Evaluation/Analysis and Conclusions/Action plans, representing the steps proposed by Vigentini et al. (2016). The aim of this review was to determine how educators used the PBL_LA framework and applied PBL and LA features in their classrooms, to report possible benefits and challenges, as well as to identify possible plans.

A summary of the reflection remarks by all educators is presented in Table 5.

Table 5 Summary of evaluation remarks on PBL_LA

Remarks	**Evaluation / Analysis**	**Conclusions / Action plans**
- There is an increase in student engagement close to the deadlines of assignments and on training days.	*- Educators could easily access previously uploaded material and communicate with students during each PBL step.*	*- A weekly or monthly report that informs groups of their status, their logs, and deliverables could add help students to have a better overview of their performance/tasks they need to fulfil.*
- The teacher could see how active the group is in discussing a particular item on each PBL step, however, there is no indication of total group activity.	*- Pre-defined assignments partly helped the group realizing their milestones and having better management of the given timeframe.*	*- A visualization of the group's progress in the form of a timeline in the header of the webpage as a reminder of the current phase of the project the group is currently working upon, as well as the forthcoming tasks and deliverables is needed.*

- The progress monitor does not help to follow different groups as they progress through different stages of the collaborative (PBL) task.	*- The ability to identify students that are being less engaged with the project at specific steps of PBL and potentially the ones that could be dropping out is very useful.*	*- Limited student participation mostly derives from their limited experience in working and collaborating in groups.*
- PBL tools such as wiki and assignment submission were mostly used by one representative of each group, making it difficult to monitor individual participation.	*- Extracting forum discussions for analysis purposes from Moodle was a cumbersome procedure. All forum discussions had to be extracted manually.*	*- Added guidelines and teacher guidance should be more increased so that students become more comfortable in participating and engaging in every PBL step.*
- Little opportunities for monitoring in platform such as yOUlearn.	*- Inconsistency between Moodle versions: LA tools that would be very useful to use (e.g. Engagement Analytics, Engagement and Recommendations) were not supported by the latest versions of the Moodle platform.*	*- A retrospective analysis for identifying the problematic PBL steps of the project would help in redesigning and improving these aspects, e.g. identified by the most commonly topics in the forum, while submission delay issues could be identified by analysing assignment delivery dates.*
- Participants preferred their own applications for making notes and did not use the wiki plugin.		
- The views of the provided resources decline with the proceeding of the PBL approach steps.		

Based on the reflection notes and discussions with educators, the benefits and challenges of combining PBL with LA can be identified. Students could absorb the weekly knowledge by having access to different modes of learning materials (slides, videos, e-lessons) and by reflecting on the weekly taught concepts through answering quizzes. The quizzes encouraged students to comprehend each lecture in depth and allowed educators to detect any specific concepts that were not understood by the

majority of the class. During each lecture, educators consulted the LA visualizations and discussed with students occurring issues, e.g. if the majority of students had low scores, if a question was answered incorrectly by a large percentage of students etc.

Challenges identified regard the usage of specific PBL and LA tools within the e-learning environments, which in some cases were not very informative or were not working properly. This limits the choices educators can make regarding LA tools they can use within each PBL step. Another challenge regards the reluctance of students to interact with e-learning platforms. This seems to mostly derive from their limited knowledge on the PBL method. Thus, educators suggest an introduction to PBL and its benefits before the course starts would be beneficial for more efficient course execution. Additionally, the selection of a small only number of PBL tools and activities from the framework could help reducing students' workload and increasing their motivation for participation.

Conclusions

Current conditions compel rapid generation of new information, technologies, and professional domains. This requires that existing and future workforce is equipped with competences that allow them to be competitive and able to transfer across professional domains. Such settings could be employed with the PBL strategy, which supports learning by doing. This, sequentially, generates large amounts of educational data that, if recorded and analysed, e.g. with LA techniques, could provide insights on the learning progress and improve the quality of the courses. However, educators need guidance during this shift in their classroom dynamic as they usually feel overwhelmed by all decisions that need to be made and are not aware of how to apply changes in their courses. Existing research shows that limited studies have combined PBL with LA to explore their potential in offering such data-driven, student-centred courses.

In this paper, a new framework is proposed that combines PBL with LA to assist educators in designing and delivering more adaptable, data-driven and student-centred courses. The framework aims to bridge the gap between promising pedagogical and technological solutions and to empower educators to reap the benefits of employing LA within PBL. The construction of the framework followed a multi-phase methodology, as shown in Table 6.

Table 6 PBL_LA framework construction: Methodology, theory an implementation

Methodology phase	**Methodology step & theoretical background**	**Implementation results**
Literature review	PBL literature review (PBL models steps	PBL and LA research pros and cons
	LA literature review (LA methods, tools, data)	
	Combining PBL with LA	

	literature review	
PBL_LA framework construction (discussions with experts)	Choose topic (PBL, LA)	PBL framework layers (pedagogical, data, analytics, ICT)
	Choose contents from literature	PBL_LA framework contents (PBL activities, tools, LA methods, data etc.)
PBL_LA framework empirical evaluation	Needs-based PBL_LA framework customization	Customized PBL_LA framework
	Course design, delivery, assessment	Web application Reflection remarks
	Framework application review	

Detailed work carried out in each phase and step of the methodology was presented. More specifically, existing research on the PBL's combination with LA provides interesting and promising results that we capitulated on as a starting point for our work. The main concepts covered in the relevant literature were documented and consulted as a guide for the basic structure of our framework, aiming to be in conformity with existing relevant research. Concepts such as PBL model, Data, tools, LA tools, and LA methods were covered in the studies examined; however, each study focuses on a sub-set of these concepts, provides limited information on the pedagogical models employed, and does not provide guidelines on how other educators can design courses that combine PBL with LA. In summary, existing literature on combining PBL with LA constitutes ofempirical studies therefore missing a conceptual framework.

The framework also aims to address important challenges faced in PBL application, especially for novice educators in using PBL. Educators are usually not aware what changes they need to make in order to transform their course to PBL and how they will become facilitators instead of deliverers of knowledge. Educators are also concerned about what tools to provide their students to support their increased participation and how to successfully monitor this participation during the entire learning process.

The proposed framework consists of four layers, namely Pedagogical, Analytics, Data and ICT. Each layer is populated with contents to guide future endeavours of applying PBL and employing LA that will exploit the generated data and provide meaningful insights. These contents include specific PBL-oriented activities, LA methods, LA tools, PBL tools and educational data that were derived from relevant literature and discussions with PBL and LA experts.

The framework can be applied for any PBL model, provides a wide variety of contents that can help especially novice educators to find this otherwise overwhelming information together and combined in one place and can be applied in any educational

level, discipline and sector. These features incorporated in the framework aim to fulfil the proposed PBL_LA contributions stated in Table 4, compared to existing relevant research. On the other hand, educators that are not well experienced in PBL and/or LA would probably need training on the use of the framework and the relevant activities, methods and ICT tools.

An empirical evaluation of the PBL_LA framework's usage in real world settings was presented. In total, the framework was applied in ten courses of various disciplines. The customization of the framework from all courses based on the respective needs, as well as the design, delivery and assessment of one of the courses were presented in detail. In all ten courses, the Aalborg PBL model was used. Each educator proceeded to select a sub-set of items from the contents of each layer depending on technical background of educator and students, subject matter, and availability of technologies.

The framework's application was tested aiming to demonstrate how PBL was combined with LA in different settings and by different individuals. The testing was carried out in a variety of domains (computer programming, information systems design, modelling tools etc.), different sectors (academia, business training), different countries and cultures (Greece, The Netherlands, Austria, Spain, Denmark) and in different e-learning environments (two LMSs, one project tracking software). In all courses, students interacted with a variety of PBL-oriented tools during each step of the PBL model, e.g. forum, quiz, mind map, wiki. Educators could monitor students' actions and consult various LA visualizations that analyzed students' engagement within the e-learning platform. A more detailed description of the framework's application was provided for one of the ten courses. The course was carried out in Greece, it featured 32 postgraduate students and lasted for 13 weeks (one academic semester). All students were divided into groups, accessed the e-learning platform and PBL tools and followed the PBL steps towards the completion of their projects. Educators of the course accessed all available LA plugins during each PBL step, monitored each student's progress and adapted the course accordingly (e.g. provided additional content to passive students, configured the quizzes for questions with low gradesetc.).

Evaluation results showed positive feedback on all different testing settings, exhibiting reliability of the framework and potential across countries, disciplines and sectors.

More specifically, the framework's application review results show educators were more aware of how to design their courses and make the right decisions in regard to the purpose of each pedagogical PBL step as well as the activities and tools to use. Educators could also consult the LA results and become more aware of students' ongoing performances and intervene when necessary. Other remarks regard educators' ability to become aware of how a course can be structured in a PBL format by building it around a problem and allotting the learning process into the different PBL steps.

In this research, the PBL_LA framework was evaluated using the Aalborg PBL model. Future work includes the augmentation of the framework's contents and relationships as well as the utilization of the framework and browser by stakeholders in other sectors

to ensure ongoing configuration and enhancement. At the same time, empirical studies could be designed and executed using quantitative analysis methods.

The structure of the framework is general thus can be used with any other PBL model as well. Future work includes the configuration of the PBL_LA framework and usage with other PBL models to assess its effectiveness in facilitating educators and combing PBL with LA. In addition, it would be interesting to investigate the suitability of the framework to accommodate other student-centred pedagogical models in the future, such as flipped classrooms. Finally, future work will also investigate ways educators can be trained inboth the PBL method and LA methods and tools to more successfully follow the proposed approach.

Acknowledgements The work presented in this paper was carried out in the context of the PBL3.0 project[1] which is funded by the European Commission within the Erasmus+ Programme under grand agreement No. 562236. The authors would like to acknowledge the assistance of PBL and LA experts as well as additional educators for their valuable feedback throughout this research. The authors would also like to thank the editor and anonymous reviewers for their comments which allowed us to considerably enhance the quality of this paper.

Annex 1: Values of PBL_LA framework

See Tables 7, 8, 9 and 10.

Annex 2

See Table 11.

Table 7 PBL-oriented activities for PBL_LA framework

Brainstorming	Data collection
Group creation	Data analysis
Literature searching	Development
Literature storing	Production/testing
Argumentation	Experimenting
Writing	Modelling
Presenting	Application of solution
Scheduling	Perform evaluation techniques
Diagramming	Record evaluation results
Resource allocation	Report writing
Roles allocation	Report submitting
Tasks allocation	Publishing
Conducting surveys	Voting
Filtering and analysis of data	Getting feedback from surveys

[1] http://pbl3-project.eu/

This is a pre-print version of the following article:
Zotou, M., Tambouris, E. & Tarabanis, K. Data-driven problem based learning: enhancing problem based learning with learning analytics. *Education Tech Research Dev* **68**, 3393–3424 (2020).
https://doi.org/10.1007/s11423-020-09828-8

Design strategy	Documenting results
Performing user tests	Drawing conclusions
Getting feedback from interviews	Proposing future work
Reflecting	

Table 8 LA methods groups and LA methods

LA method groups	**LA methods**
Modelling	Learner modelling
	Behaviour modelling
	User profile design
	Natural language processing
	Assessment by matching learner's knowledge with knowledge domain
Relationship mining	Sentiment analysis
	Discourse analysis
	Association rule mining
Personalization & adaptation	Adaptive content to learners
	Recommendations on content, activities and interactions
Predictions	Changes in learner behaviour, identification of errors
	Early risk identification
	Interventions
	Classification
Structure discovery and analysis	Social network analysis
	Information flow analysis
	Semantic analysis
	Clustering
Traditional evaluation and monitoring	Feedback
	Reflection
	Guiding
	Scaffolding

Table 9 Data for PBL_LA framework

Assignments submitted	Relations of connected concepts in mind map
Courses enrolled	Version of problem formulated
Times spent on each unit of learning	Number of problem versions
Access to learning resources	Differences between problem versions
Navigation patterns	Structure of problem presentation
Frequency of log ins	Content of problem presentation
Activities accessed / used	Navigation patterns
Posts on forums	Engagement levels
Number of participants per group	Number of different roles allocated
Clusters of students who made specific mistakes	Allocation of roles across learners

Contributions to shared documents	Number of tasks
Facebook posts and interactions (replies, shares, tags)	Types of tasks
Twitter posts and interactions (replies, retweets)	Allocation of tasks across learners
Blog posts and interactions (comments, ratings)	Tasks level of completion
LinkedIn posts and interactions (networks, shares)	Tasks workload progress
Time spent in different webpages	Timesheets
Performance in assignments	Number / duration of meetings held
Performance in activities	Participants in each meeting
Performance in quizzes	Meeting minutes
Final exam grades	Survey data
Frequency of interactions / engagement	Interview data
Geo locations	Brainstorming / dialogues
Sensor data (e.g. movements, gestures, activities, physical state etc.)	Refined problem statement
Participation in group work	Number of design requirements
Students having difficulties with a concept	Content of design requirements
Learner profile	Design strategy
Types of resource sources	Testable prototype
Content of resources	Description of solution's implementation
Number of resources	Refined evaluation strategy
Content of discussions / posts	Evaluation results
Content of ideas	Report
Number of ideas formed by learner	Future work
Votes	Presentation
Mind maps	Brainstorming / dialogue posts
Number of concepts per mind map	Time spent on group creation

Table 10 ICT tools for PBL_LA framework

LA tools	**PBL tools**
GLASS	LMS (Moodle)
CourseVis	Twitter
Google Analytics	Forum
SNAPP	Email
LeMo	Mindmap
StepUp!	Google Docs
LOCO-Analyst	Mindmeister
Netlytic	Digital library
eLAT toolkit	Google scholar
Gismo	Google / Bing
MOCLog	Dropbox
Learning Analytics Enhanced Rubric	Blogger
SmartKlass	Social bookmarking (Diigo)
Analytics and Recommendations	Digg
Adaptive quiz	Limesurvey
Orange Textable	Excel

Analytics Graphs	Linkedin
Configurable reports	Skype
Engagement Analytics	Facebook
Connect for Success (C4S)	Prezi
Automated Wellness Engine (AWE)	Google calendar
Personalized Adaptive Study Success (PASS)	Doodle
Anaconda analytics	Gliffy
NetlyticeLAT	Dabbleboard
	Task checklist
	Software for diagramming
	Task management
	Slideshare
	Quiz
	Wiki

Table 11 Aalborg PBL Steps

Step	**Description**
Group forming	Learners create groups.
Problem formulation	Each group member is asked to identify issues relevant to the domain that could be considered as problems that need solving and that could be solved within the designated timeframe.
Task formulation.	The groups divide the problem into specific tasks that need to be carried out towards solving the problem. This step also includes the allocation of tasks to members of the group.
Data gathering	Each group collects data that is relevant and essential for implementing the tasks they have formulated.
Analysis	Each group analyses the gathered information and reach specific decisions regarding the design of the solution.
Design	Groups proceed to design the strategy for the solution to their formulated problem.
Implementation	Groups proceed to implement the solution they designed. Implementations can vary depending on the solution designed and should be suitable for applying in test conditions for evaluation and validation.
Evaluation	The implemented solution is applied in various settings and evaluated using different evaluation methods, e.g. interviews, questionnaires, technical tests, simulations etc.
Reporting	The project work is reported. Here, each group posts their conclusions and proposes future work for the problem they solved.

References

Ali, L., Hatala, M., Gašević, D., & Jovanović, J. (2012). A qualitative evaluation of evolution of a learning analytics tool. *Computers & Education, 58*(1), 470–489.

Allen, D. E., Duch, B., Groh, S., Watson, G. B., & White, H. B. (2003). Professional development of University Professors: Case study from the University of Delaware. In International conference Docencia Universitaria en Tiempos de Cambio [University Teaching in Times of Change] at Pontificia Universidad Catσlica del Perï, Lima.

Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In J. Larusson & B. White (Eds.), Learning analytics (pp. 61–75). New York: Springer.

Buckingham, S. S. (2011). Learning analytics: Ascilite 2011 Keynote. https://simon.buckinghamshum. net/2011/12/learning-analytics-ascilite2011-keynote/.

Chalhoub-Deville, M., & Deville, C. (2008). Utilizing psychometric methods in assessment. Encyclopedia of Language and Education, 7, 211–224.

Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2012). A reference model for learning analytics.

International Journal of Technology Enhanced Learning, 4(5–6), 318–331.

Chen, J., Kolmos, A., & Du, X. (2020). Forms of implementation and challenges of PBL in engineering education: A review of literature. European Journal of Engineering Education. https://doi. org/10.1080/03043797.2020.1718615.

Chen, Y., Hogaboam, P., Hmelo-Silver, C. E., Lajoie, S. P., Wiseman, J., Bodnar, S., et al. (2016). Instructional dashboards to support deep learning in an online problem-based learning environment. Learning Environments for Deep Learning in Inquiry and Problem-Solving Contexts, 20, 6.

Davis, M. H., & Harden, R. M. (1999). AMEE Medical Education Guide No. 15: Problem-based learning: A practical guide. Medical Teacher, 21(2), 130–140.

Dimitracopoulou, A. (2015). Learning analytics: Concepts, methods, tools, achievements, research directions and perspectives [Synthesis and critical analysis of a new interdisciplinary field], Invited talk by the Master Program in “New Technologies for Communication and Learning”, Technological University of Cyprus, Limassol. Accessed on 20 January 2019 at https://ltee.aegean.gr/en/learning-analytics/.

Dimopoulos, I., Petropoulou, O., & Retalis, S. (2013). Assessing students’ performance using the learning analytics enriched rubrics. In Proceedings of the Third International Conference on Learning Analytics and Knowledge (pp. 195–199).

Dyckhoff, A.L., Lukarov, V., Muslim, A., Chatti, M.A, & Schroeder, U. (2013). Supporting action research with learning analytics. In Proceedings of the International Conference on Learning Analytics and Knowledge (pp. 220–229). New York: ACM Press.

Dyckhoff, A. L., Zielke, D., Bültmann, M., Chatti, M. A., & Schroeder, U. (2012). Design and implementation of a learning analytics toolkit for teachers. Journal of Educational Technology & Society, 15, 58–76.

El Alfy, S., Gómez, J. M., & Dani, A. (2019). Exploring the benefits and challenges of learning analytics in higher education institutions: A systematic literature review. Information Discovery and Delivery, 47(1), 25–34.

Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. International Journal of Technology Enhanced Learning, 4(5–6), 304–317.

Fortenbacher, A., Beuster, L., Elkina, M., Kappe, L., Merceron, A., Pursian, A., et al. (2013). LeMo: A learning analytics application focussing on user path analysis and interactive visualization. Proceedings of the IEEE 7th International Conference on Intelligent Data Acquisition and Advanced Computing Systems, 2, 748–753. https://doi.org/10.1109/IDAACS.2013.6663025.

Foster, E., & Siddle, R. (2019). The effectiveness of learning analytics for identifying at-risk students in higher education. Assessment & Evaluation in Higher Education, 45(6), 842–854.

Gong, L., & Liu, Y. (2019). Design and application of intervention model based on learning analytics under blended learning environment. In Proceedings of the 2019 7th International Conference on Information and Education Technology (pp. 225–229).

Göhnert, T., Ziebarth, S., Malzahn, N., & Hoppe, H. U. (2014). Enriching (learning) community platforms with learning analytics components. In N. Baloian, F. Burstein, H. Ogata, F. Santoro, & Z. Zurita (Eds.), Collaboration and technology (pp. 177–184). Springer. https://doi.org/10.1007/978-3-319- 10166-8_16.

Hershkovitz, A., de Baker, R. S. J., Gobert, J., Wixon, M., & Pedro, M. S. (2013). Discovery with models: A case study on carelessness in computer-based science inquiry. American Behavioral Scientist, 57(10), 1480–1499.

Kolmos, A., Fink, F. K., & Krogh, L. (Eds.). (2004). The Aalborg PBL model-progress, diversity and challenges. Aalborg: Aalborg University Press.

Koschmann, T. D., Myers, A. C., Feltovich, P. J., & Barrows, H. S. (1994). Using technology to assist in realizing effective learning and instruction: A principled approach to the use of computers in collaborative learning. The Journal of the Learning Sciences, 3(3), 227–264.

Kotsiantis, S., Tselios, N., Filippidi, A., & Komis, V. (2013). Using learning analytics to identify successful learners in a blended learning course. International Journal of Technology Enhanced Learning, 5(2), 133–150.

Leony, D., Pardo, A., de la Fuente Valentín, L., Sánchez de Castro, D., Delgado Kloos, C. (2012). GLASS: A learning analytics visualization tool. In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (pp. 162–163). ACM, New York.

Leung, C. (2012). Qualitative research in language assessment. The Encyclopedia of Applied Linguistics. https://doi.org/10.1002/9781405198431.wbeal0979.

Lias, T. E., & Elias, T. (2011). Learning analytics: The definitions, the processes, and the potential. Long, P. D., & Siemens, G. (2011). Penetrating the fog: analytics in learning and education. TD Tecnolo-

gie Didattiche, 22, 132–137.

MacNeill, S., Campbell, L. M., & Hawksey, M. (2014). Analytics for education. Journal of Interactive Media in Education. https://doi.org/10.5334/2014-07.

Mauffette, Y., & Poliquin, L. (2002). PBL in science education: A curriculum reform in biology at University of Quebec in Montreal. PBL Insight, 4(1), 1–5.

Mazza, R., Bettoni, M., Faré, M., & Mazzola, L. (2012). MOCLog—Monitoring Online Courses with log data. In Proceedings of the 1st Moodle Research Conference (pp. 132–139), Heraklion, Greece. Mazza, R., & Dimitrova, V. (2007). CourseVis: A graphical student monitoring tool for supporting instructors in web-based distance courses. International Journal of HumanComputer Studies, 65,

125–139.

McGaghie, W. C., Bordage, G., & Shea, J. A. (2001). Problem statement, conceptual framework, and research question. Academic Medicine, 76(9), 923–924.

Neame, R. L. (1989). Problem-based medical education: The Newcastle approach. In H.G. Schmidt, M.

J. Lipkin, M. W. Vries, & J. M. de Greep (Eds.) New directions for medical education (pp. 112– 146). New York: Springer.

Neville, A. J. (2009). Problem-based learning and medical education forty years on. Medical Principles and Practice, 18(1), 1–9.

Otte, E., & Rousseau, R. (2002). Social network analysis: A powerful strategy, also for the information sciences. Journal of Information Science, 28, 441–453.

Ørngreen, R., Knudsen, S. P., Kolbæk, D., & Jensen, R. H. S. (2019). Investigating the use of Moodle at a PBL University: Design factors and experiences. In 18th European Conference on e-Learning (p. 444). Academic Conferences and publishing limited.

Papamitsiou, Z., & Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. Journal of Educational Technology & Society, 17, 49–64.

Pardo, A. (2014). Designing learning analytics experiences. In J. A. Larusson & B. White (Eds.), Learning analytics: From research to practice (pp. 15–38). New York: Springer.

Picciano, A. G. (2012). The evolution of big data and learning analytics in American higher education.

Journal of Asynchronous Learning Networks, 16(3), 9–20.

Pistilli, M. D., Willis, J. E., & Campbell, J. P. (2014). Analytics through an institutional lens: Definition, theory, design, and impact. In J. Larusson & B. White (Eds.), Learning analytics (pp. 79–102). New York: Springer.

Prinsloo, P., Slade, S., & Galpin, F. (2012). Learning analytics: Challenges, paradoxes and opportunities for mega open distance learning institutions. In 2nd International Conference on Learning Analytics and Knowledge (pp. 130–133). Vancouver.

Rojas, I. G., & Garcia, R. M. C. (2012). Towards efficient provision of feedback supported by learning analytics. In 2012 IEEE 12th International Conference on Advanced Learning Technologies (ICALT) (pp. 599–603). IEEE.

Romero, C., & Ventura, S. (2013). Data mining in education. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 3, 12–27.

Saarinen-Rahiika, H., & Binkley, J. M. (1998). Problem-based learning in physical therapy: A review of the literature and overview of the McMaster University experience. Physical Therapy, 78(2), 195–207.

Santos, J. L., Govaerts, S., Verbert, K., & Duval, E. (2012). Goal-oriented visualizations of activity tracking: A case study with engineering students. In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (pp. 143–152).

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

Schultz, N., & Christensen, H. P. (2004). Seven-step Problem-Based Learning in an interaction design course. European Journal of Engineering Education, 29(4), 533–541.

Siemens, G. (2013). Learning analytics: The emergence of a discipline. American

Behavioral Scientist, 57(10), 1380–1400.

Sohmen, V. S. (2020). Project-based learning (PBL) in a higher education project: Introduction of an accelerated PBL (A-PBL) model. In M. C. P. O. Okojie & T. C. Boulder (Eds.), Handbook of research on adult learning in higher education (pp. 118–150). IGI Global.

Spikol, D., Ruffaldi, E., Dabisias, G., & Cukurova, M. (2018). Supervised machine learning in multimodal learning analytics for estimating success in project-based learning. Journal of Computer Assisted Learning, 34(4), 366–377.

Tambouris, E., Panopoulou, E., Tarabanis, K. A., Ryberg, T., Buus, L., Peristeras, V., et al. (2012). Enabling problem based learning through Web 2.0 technologies: PBL 2.0. Journal of Educational Technology & Society, 15, 238–251.

Tempelaar, D. T., Rienties, B., & Giesbers, B. (2015). In search for the most informative data for feedback generation: Learning analytics in a data-rich context. Computers in Human Behavior, 47, 157–167.

Ünal, E. (2019). Web 2.0 technologies supporting problem based learning: A systematic literature review.

Journal of Problem Based Learning in Higher Education, 7(1), 25–50.

Van Harmelen, M., & Workman, D. (2012). Analytics for learning and teaching. CETIS Analytics Series, 1(3), 1–40.

Verbert, K., Duval, E., Klerkx, J., Govaerts, S., & Santos, J. L. (2013). Learning analytics dashboard applications. American Behavioral Scientist, 57(10), 1500–1509.

Vigentini, L., Mirriahi, N., & Kligyte, G. (2016). From reflective practitioner to active researcher: Towards a role for learning analytics in higher education scholarship. Learning, Design, and Technology: An International Compendium of Theory, Research, Practice, and Policy, 1–29.

West, D. M. (2012). Big data for education: Data mining, data analytics, and Web dashboards. Brookings Institution. Governance Studies. https://www.brookings.edu/~/media/research/files/paper s/2012/9/04%2520e ducat ion%2520t echno logy%2520w est/04%2520e ducat ion%2520t echno logy%2520west.pdf.

Wong, B. T. M. (2019). The benefits of learning analytics in open and distance education: A review of the evidence. In M. S. Khine (Eds.), Emerging trends in learning analytics (pp. 65–81). Brill Sense.

Wijnia, L., Loyens, S. M., & Rikers, R. M. (2019). The problem-based learning process: An overview of different models. In M. Moallem, W. Hung, & N. Dabbagh (Eds.), The Wiley handbook of problem-based learning (pp. 273–295).

Zhou, C., & Zhu, Z. (2019). Fostering problem-based learning (PBL) in Chinese universities for a creative society. In Z. Zhu & C. Zhu (Eds.), Global perspectives on fostering problem-based learning in Chinese Universities (pp. 1–31). IGI global.

Zotou, M. (2015). Enhancing students' skills and capabilities to exploit Open Government Data. Innovation and the Public Sector, 24, 327.

Zotou, M., & Tambouris, E. (2014). Data-driven blended problem based learning towards enhancing transversal skills. In 2014 IEEE 14th International Conference on Advanced Learning Technologies (ICALT) (pp. 762–764). IEEE.