Cross-Platform Analytics: A step towards Personalization and Adaptation in Education

Katerina Mangaroska
Norwegian University of Science and Technology
Trondheim, Norway
mangaroska@ntnu.no

Boban Vesin
University of South-Eastern Norway
Vestfold, Norway
boban.vesin@usn.no

Michail Giannakos
Norwegian University of Science and Technology
Trondheim, Norway
michailg@ntnu.no

ABSTRACT

Learning analytics are used to track learners' progress and empower educators and learners to make well-informed data-driven decisions. However, due to the distributed nature of the learning process, analytics need to be combined to offer broader insights into learner's behavior and experiences. Consequently, this paper presents an architecture of a learning ecosystem, that integrates and utilizes cross-platform analytics. The proposed cross-platform architecture has been put into practice via a Java programming course. After a series of studies, a proof of concept was derived that shows how cross-platform analytics amplify the relevant analytics for the learning process. Such analytics could improve educators' and learners' understanding of their own actions and the environments in which learning occurs.

CCS CONCEPTS

• **Applied computing → Interactive learning environments**;

KEYWORDS

learning analytics, multimodal systems, architecture

ACM Reference Format:
, Katerina Mangaroska, Boban Vesin, and Michail Giannakos. 2019. Cross-Platform Analytics: A step towards Personalization and Adaptation in Education. In *The 9th International Learning Analytics & Knowledge Conference (LAK19), March 4–8, 2019, Tempe, AZ, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3303772.3303825

LAK19, March 4–8, 2019, Tempe, AZ, USA

ACM ISBN 978-1-4503-6256-6/19/03…$15.00
https://doi.org/10.1145/3303772.3303825

1 INTRODUCTION

Pervasive technologies are used to allow learners and educators to take advantage of a learning ecosystem that includes various learning spaces, both physical and digital. On one hand, the potential of improving learning and teaching lies in the capacity to capture multifaceted authentic learner-generated data coming from diverse learning spaces (e.g., clickstream data, assessment data, grades, gaze data, physiological data). On another hand, researchers and educators need tools that could collect, harmonize, and integrate that data to harness the dormant potential.

Currently, learning analytics (LA) tools are mostly used to increase awareness and support different sorts of assessment employing single virtual learning environment (VLE) or utilizing one learning management system (LMS) [16]. However, distributed learning environments impose a need for combining data through seamless integration of cross-platform analytics (CPA) capabilities. Hence, the focus of the study is to enhance the analytics capacities of an existing system with an architecture that integrates three separate online learning systems and utilizes the CPA approach. In fact, the aim is to capture rich and authentic data across interconnected learning spaces, harmonize it, and visualize it as a support tool to learners and educators. Consequently, the study addresses the following research question (**RQ**): How cross-platform analytics can be combined to more accurately portrait learning, and support personalized and adaptive learning opportunities?

2 BACKGROUND AND RELATED WORK

Current research often relies on metrics such as: time spent in learning environments, clickstream data, self-reported data, or test performances, failing to include context-aware data collected across-spaces [17] or heterogeneous data coming from more than one platform [6, 16]. Contemporary learning ecosystems are consisted of several learning spaces, with those spaces to collect diverse learning analytics that are often analyzed and utilized in silos from other spaces of the ecosystem. Data from learning spaces is becoming easier to collect, difficult to interpret, and complex for teachers to understand. In addition, the more complex the data is, the harder it becomes to synchronize, analyze and utilize it. In fact, there are frameworks that describe how to capture data from various sources [10], but there is lack of tools to easily establish cross-platform and sometimes multi-modal fully inter-operable systems [6].

Nevertheless, one recent example is GLUEPS-MAASS, a conceptual model that attempts to collect and integrate data from multiple sources, and set up a multi-modal system [18]. Considering this example, the authors propose a cross-platform architecture that integrates analytics coming three digital learning environments (i.e., ProTuS, MasteryGrid, and Eclipse IDE) with the aim to enrich learning experiences and towards personalized feedback. On one hand, personalization could deliver more engaging and relevant learning activities, but it presents difficulties when it needs to be implemented within the learning process [11, 14]. On another hand, personalization holds the potential to spark the shift from a teacher-centered perspective to a learner-centered, competency-oriented, and from the concept of a "learning activity" to a "learning experience" Past research demonstrates development of tools [1, 5]; models [2, 22]; and adaptive learning systems [21] for monitoring, assessing, and predicting learner's behavior and performance. The results demonstrate that data integration from multiple sources is a successful way to improve prediction accuracy and design data-driven improvements accordingly.

Several conceptual frameworks and software architectures [7, 20, 23] are designed to effectively store and retrieve large amount of learner-generated data, while other to analyze that data [9]. Moreover, a lot of research in the field of technology-enhanced learning has been concentrated on enhancing e-learning interoperability [8]. Thus, several industrial solutions, such as The Learning Tools Interoperability [24] and the Experience API [13], are widely used to enhance systems and tools interoperability [9]. However, none at present fully supports a standardized approach of collecting data [4]. Consequently, the authors propose a innovative architecture that integrates multifaceted data and leverages CPA to support accurate predictions of behavior and performance, as well as to establish personalized and adaptive assessment of knowledge and skills.

3 CONCEPTUAL MODEL OF CROSS-PLATFORM ARCHITECTURE

This section briefly presents the technical and the design decisions for building a robust infrastructure that integrates data across various distributed platforms. The aim is to develop an integrated interface (i.e., VENT), avoiding the need to manually log in, gather, and synchronize data from different systems [25]. Thus, the integration encompasses two functional layers (see Figure 1):

- The dashboard layer (i.e., VENT) - aggregates, integrates, and generates visualized data.
- The data source layer - provides the content data. This layer consists of standalone applications.

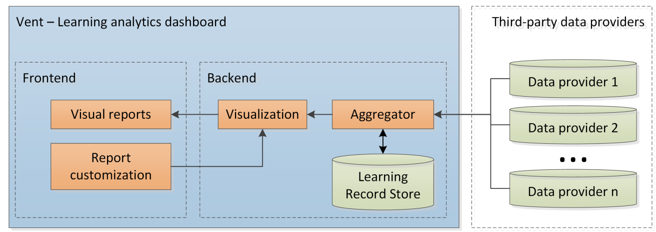

Figure 1: The architecture of LA dashboard

Figure 1 shows *the back-end* that consists of *Aggregator* and *Visualization* component. The *Aggregator* component mirrors every specific data type available to the Learning Record Store - LRS. The LRS is responsible for data storing, and providing access to aggregated learning records, containing both static data (e.g., personal) and dynamic data (e.g., clickstream) about every active learner. The xAPI specification [3] has been used as a standard vocabulary for communication with distributed educational data as it is inherently extensible to accommodate unforeseen data collection needs. The *Visualization* component creates visual representations of the aggregated data.

The *Front-end* offers access to categorized visual reports, and defines the format and the scope of the reports within the *Report customization* component. Moreover, visualizations are created in the *back-end* of the application, based on user input within the *Report Customization* component. Educators can customize the visualized reports by selecting activities and students (individuals or groups) for display, thus forming the vectors. The system uses the selected vectors to generate one or more graphs based on predefined templates.

Architectural layers

To support the integration of components from different systems (i.e., *ProTuS*, *MasteryGrids*, and *Eclipse IDE*), the authors implemented a layered architecture. The layered architecture aligns the business flow within the personalized learning environments, making it a natural choice for this implementation. The proposed architecture contains the following layers (Figure 2):

- **Presentation layer:** manages the interfaces and the browser communication logic. Users, such as learners and educators, submit their requests over this layer.
- **Application layer:** includes the *Data processing* engine, responsible for heterogeneous data collection and its redirection to the *Business* layer for data processing; and the *Visualization* engine that generates reports and visualize recommendations.

- **Business layer:** responsible for automation of the assessment process and generating recommendations.
- **Data layer:** stores learners' data and the content.

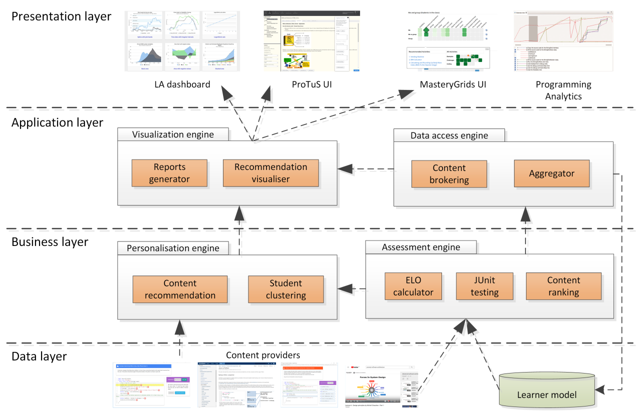

Figure 2: The overall architecture of the LA environment

4 METHODOLOGY

The approach adopted in the study is based on System Development Research [19] and the best practices in User-Centred Design enriched with the state-of-the-art approaches in Design Based Research (DBR) [26]. In fact, the overall study consists of three DBR cycles, each leading towards *development of cross-platform architecture among interconnected digital learning spaces*. This study focuses on the last DBR cycle, which is actually a low-scale but longitudinal evaluation of the proposed architecture.

Data collection

The study was carried out from mid January till March 2018 in an introductory object oriented programming course, part of the Computer Science Degree at Norwegian University of Science and Technology - NTNU. The study involved an experienced instructor of Java programming and 20 freshmen CS-major (12 males, 8 females), who had already taken an intro to programming in Python in their first semester. Before the start of the study, the students were introduced with NTNU's policy for ethical and data privacy issues.

The data was collected from examples, challenges, and coding exercises. Students could either select assignments by their own choice or follow recommendations generated by the system. Each example starts with a worked-out explanation how to write a code for a particular problem. Explanations and hints are available for almost all lines in the example, and the system records logs of all user's actions. Challenges, on the other hand, show a problem with blank lines that needs to be fill in by dragging and dropping the pieces of code to the blank fields. Immediate feedback and hints are also available for each challenge. At last, coding exercises require students to complete a given code to solve a problem. After students submit the code, it is being tested by a set of unit tests and the user receives a useful feedback.

Measures

The study looks into student's performance (i.e., dependent variable) measured from students' scores. Besides performance, the other traced measures are categorized under the system from which were generated:

- **Student's performance**: student's performance from coding exercises based on Elo-rating algorithm (performance.ProTuS);
- ***ProTuS* measures**: No. of visited content resources (visited.content), actions (actions.ProTuS), sessions (sessions.ProTuS), and solved coding exercises of average difficulty (coding.level3) in ProTuS;
- ***MasteryGrids* measures**: No. of sessions (sessions.-MasteryGrids), topics covered (topics.covered), sets completed (sets.completed), challenges solved in the 1st attempt (challenges.1st), time spend in statistics (MasteryGrids.stat) and total duration (total.duration);
- ***Programming analytics* measures**: No. of super easy assignments (SE.assignments); No. of easy assignments (E.assignments);

The selection of the measures does not attempt to offer an exhaustive set of data that can be harmonized, or combined to portrait learners' experiences. On the contrary, the selection is twofold: 1) providing a proof of concept of the proposed architecture, and 2) demonstrating that cross-platform analytics can indeed support researchers and educators to extract more valuable insights from learning experiences.

Data analysis

To get an initial understanding of the measures, descriptive statistics were calculated and the Shapiro-Wilk test was used to check for data normality. Next, a Spearman's correlation coefficient was calculated to investigate the relationship between the various measures. Finally, to identify the potential of combining CPA, the authors conducted a series of linear regressions; first, using analytics only from *ProTuS* and then adding analytics from the other platforms. The authors performed a stepwise regression, that is usually used for exploration, when researchers do not know which independent variables will create the best prediction equation [15].

5 RESULTS

First, the data was checked for normality. A Shapiro-Wilk test was performed due to the small sample size (n=20). The results showed that the data does not have normal distribution (p values were significant). Next, to investigate the relationship among the variables, the authors performed a pair-wise correlation analysis between the 12 extracted measures and the performance. A non-parametric Spearman's rank correlation was used to compute the correlations, due to the non-normal distribution and the highly skewed nature of the data. The results showed that student's performance has a positive relation with all four measures from *ProTuS*, with four measures from *MasteryGrids* (i.e., sessions.MateryGrids, topics.covered, MateryGrids.stat, and total.duration) and with none of the measures from *Eclipse IDE*.

Since linear regression analysis does not assume normality for either the predictor or the outcome variable, the non-normal distribution of the collected data was not an obstacle to perform regression analysis (as stated by Gauss–Markov theorem) [12]. Consequently, a stepwise linear regression for exploratory model building was performed to check which predictors entered into the model based on a purely mathematical criterion, best explain the variations in the dependent variable. Hence, as a first step in the regression analysis, the authors tested whether the four measures extracted from *ProTuS* as independent variables can predict student's performance. The results are presented in Model 1, as shown in Figure 3. As it can be observed, the model is significant (F(1,18)=23.97, p<0.000, R2=.57) and contains only one noteworthy predictor (i.e., coding.level3), while the other predictor variables have been excluded (Figure 4). Thus, the number of solved average difficulty exercises can explain 57% of the variation in student's performance.

Model Summary

Regression models	R^2	ΔR^2	ΔF	B	SE	β	p	ϱ
Model 1								
Constant				1307,53	5,18			
coding.level3	0,57	0,57	23,97	31,48	6,43	0,76	0,000	0,55
Model 2								
Constant				1302,75	5,03			
coding.level3				28,09	5,9	0,67	0,000	0,55
MasteryGrids.stat				1,99	0,83	0,34	0,028	0,54
coding.level3; MasteryGrids.stat	0,68	0,11	5,74	-	-	-	0,028	-
Model 3								
Constant				1294,37	5,91			
coding.level3				25,69	5,43	4,73	0,000	0,55
MasteryGrids.stat				2,12	0,75	2,81	0,013	0,54
sessions.Protus				1,73	0,79	2,21	0,042	0,60
coding.level3; MasteryGrids.stat; sessions.ProTuS	0,75	0,08	4,88	-	-	-	0,042	-

Figure 3: Models summary for student performance

Next, a hierarchical regression was performed accounting for the other eight variables, while controlling the extracted four measures from *ProTuS*. The results are presented in Model 2 (Figure 3) which contains two noteworthy predictors, while the the other ten variables were excluded (Figure 4). Model 2 is significant (F(1,17)=5.73, p<0.028, R2=.68) and accounts for 68% of the variation in student's performance. In other words, adding the total time spend in *MasteryGrids* statistics to the number of solved average difficulty exercises, accounts for additional 11% explanation in the variation in the dependent variable.

Finally, another regression was run including only three predictor variables (coding.level3, MateryGrids.stat and sessions.Protus) to estimate if all three combined can account for

Excluded variables	t-test	p	ϱ
Model 1			
visited.content	0,81	0,43	0,46
actions.ProTuS	1,93	0,07	0,68
sessions.ProTuS	1,69	0,11	0,60
Model 2			
visited.content	0,65	0,53	0,46
actions.ProTuS	1,40	0,18	0,68
sessions.ProTuS	2,21	**0,04**	0,60
sessions.MasteryGrids	1,28	0,22	0,54
topics.covered	0,53	0,60	0,50
sets.completed	-0,41	0,69	0,35
challenges.1st	-0,38	0,71	0,27
total.duration	-0,11	0,91	0,58
SE.assignments	1,05	0,31	0,17
E.assignments	0,42	0,68	0,03

Figure 4: Excluded variables

a greater proportion of the variance in student's performance. This regression was run due to the excluded but noteworthy predictor variable (i.e., sessions.Protus) from Model 2, (Figure 4). The results showed that adding a third predictor variable, (number of sessions) can explain 75% of the variation in student's performance. These results represent Model 3, which is significant (F(1,16)=4.88, p<0.042, R2=.75). Hence, combining analytics from different platforms (i.e., cross-platform analytics) has the capacity to significantly improve the prediction of student's performance.

6 DISCUSSION AND CONCLUSION

To demonstrate and validate real-life examples of how and when learning is taking place, educators and researchers need to embrace the complexity of the learning process and its distributed nature across various spaces and contexts. Consequently, this study takes a humble approach to analysis, comparing LA across three online platforms utilizing correlation and regression analysis. The authors report, with considerable caution, a positive findings as a proof of concept for the feasibility and the potential of combining LA across platforms.

The reported findings present two models; model 1 that explains most of the variation in student's performance (i.e., 57%) only with one measure, and model 2 that shows an improvement by explaining 75% variation in the dependent variable. Thus, the results suggest that cross-platform analytics does account for a significant additional increase in the explanation of the variation in the outcome (i.e., 18%) or in other words, with an overall effect of 30%. This is a significant step towards building learner models that explain higher portions of variation in the outcome (e.g. student's performance) combining LA across platforms.

On the practical side, the authors managed to propose and implement in practice, architecture that integrates and interconnects CPA capabilities. This approach provides a unique opportunity to enrich the contemporary learner models, leaving it behind the exclusive focus on single source click stream data, while offering optimal learning designs for different user groups, needs, and circumstances.

Limitations

The study has several limitations that imply the future directions. First, the sample used in the third DBR cycle is relatively small (n=20); however, capturing and analyzing experiences of 20 students who had intensely use the learning ecosystem for three months provided a clear data-set, and supports a proof of concept for the proposed architecture. Second, the authors did not apply any rigorous technique for measures selection, but mainly selected 12 measures that most of the LMSs capture. Hence, selecting and crossing different analytics might enrich the benefits of cross-platform analytics. In addition, the selection of these three systems potentially limits the generalization of our findings.

ACKNOWLEDGMENT

This work was supported by the Research Council of Norway under the project FUTURE LEARNING (255129/H20).

REFERENCES

[1] Kimberly E Arnold and Matthew D Pistilli. 2012. Course signals at Purdue: Using learning analytics to increase student success. In *Proceedings of the 2nd international conference on learning analytics and knowledge*. ACM, 267–270.

[2] RS Baker, Jody Clarke-Midura, and Jaclyn Ocumpaugh. 2016. Towards general models of effective science inquiry in virtual performance assessments. *J. of Computer Assisted Learning* 32, 3 (2016), 267–280.

[3] Aneesha Bakharia, Kirsty Kitto, Abelardo Pardo, Dragan Gašević, and Shane Dawson. 2016. Recipe for success: lessons learnt from using xAPI within the connected learning analytics toolkit. In *Proceedings of the sixth international conference on LA & knowledge*. ACM, 378–382.

[4] Alan Berg, Maren Scheffel, Hendrik Drachsler, Stefaan Ternier, and Marcus Specht. 2016. The dutch xAPI experience. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*. ACM, 544–545.

[5] Matthew Berland, Don Davis, and Carmen Petrick Smith. 2015. AMOEBA: Designing for collaboration in computer science classrooms through live learning analytics. *International Journal of Computer-Supported Collaborative Learning* 10, 4 (2015), 425–447.

[6] Paulo Blikstein and Marcelo Worsley. 2016. Multimodal Learning Analytics and Education Data Mining: using computational technologies to measure complex learning tasks. *Journal of Learning Analytics* 3, 2 (2016), 220–238.

[7] Peter Brusilovsky. 2004. KnowledgeTree: A distributed architecture for adaptive e-learning. In *Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters*. ACM, 104–113.

[8] Luis De-La-Fuente-Valentín, Mar Pérez-Sanagustín, Davinia Hernández-Leo, Abelardo Pardo, Josep Blat, and Carlos Delgado Kloos. 2014. Technological support for the enactment of collaborative scripted learning activities across multiple spatial locations. *Future Generation Computer Systems* 31 (2014), 223–237.

[9] Juan Manuel Dodero, Enrique Juan González-Conejero, Guillermo Gutiérrez-Herrera, Sonia Peinado, José Tomás Tocino, and Iván Ruiz-Rube. 2017. Trade-off between interoperability and data collection performance when designing an architecture for learning analytics. *Future Generation Computer Systems* 68 (2017), 31–37.

[10] Federico Domínguez and Katherine Chiluiza. 2016. Towards a distributed framework to analyze multimodal data. In *Proc. of Workshop Cross-LAK–held at LAK '16*. 52–57.

[11] Elizabeth FitzGerald, Natalia Kucirkova, Ann Jones, Simon Cross, Rebecca Ferguson, Christothea Herodotou, Garron Hillaire, and Eileen Scanlon. 2018. Dimensions of personalisation in technology-enhanced learning: A framework and implications for design. *British Journal of Educational Technology* 49, 1 (2018), 165–181.

[12] Andrew Gelman and Jennifer Hill. 2006. *Data analysis using regression and multilevel hierarchical models*. Cambridge university press.

[13] Jonathan M Kevan and Paul R Ryan. 2016. Experience API: Flexible, decentralized and activity-centric data collection. *Technology, knowledge and learning* 21, 1 (2016), 143–149.

[14] Aleksandra Klašnja-Milićević, Boban Vesin, Mirjana Ivanović, Zoran Budimac, and Lakhmi C Jain. 2017. Personalization and Adaptation in E-Learning Systems. In *E-Learning Systems*. Springer, 21–25.

[15] Sébastien Lallé, Cristina Conati, and Giuseppe Carenini. 2016. Prediction of individual learning curves across information visualizations. *User Modeling and User-Adapted Interaction* 26, 4 (2016), 307–345.

[16] Katerina Mangaroska and Michail N Giannakos. 2018. Learning analytics for learning design: A systematic literature review of analytics-driven design to enhance learning. *IEEE Transactions on Learning Technologies* (2018).

[17] Roberto Martinez-Maldonado. 2016. Seeing learning analytics tools as orchestration technologies: Towards supporting learning activities across physical and digital spaces. In *CEUR Workshop Proceedings*, Vol. 1601. 70–73.

[18] Juan Alberto Muñoz-Cristóbal, María Jesús Rodríguez-Triana, Vanesa Gallego-Lema, Higinio F Arribas-Cubero, Juan I Asensio-Pérez, and Alejandra Martínez-Monés. 2016. Toward the Integration of Monitoring in the Orchestration of Across-spaces Learning Situations.. In *CrossLAK*. 15–21.

[19] Jay F Nunamaker Jr, Minder Chen, and Titus DM Purdin. 1990. Systems development in information systems research. *Journal of management information systems* 7, 3 (1990), 89–106.

[20] Thomas Rabelo, Manuel Lama, Ricardo R Amorim, and Juan C Vidal. 2015. SmartLAK: A big data architecture for supporting learning analytics services. In *Frontiers in Education Conference*. IEEE, 1–5.

[21] Leena M Razzaq and Neil T Heffernan. 2009. To Tutor or Not to Tutor: That is the Question.. In *AIED*. 457–464.

[22] María Jesús Rodríguez-Triana, Alejandra Martínez-Monés, Juan I Asensio-Pérez, and Yannis Dimitriadis. 2015. Scripting and monitoring meet each other: Aligning learning analytics and learning design to support teachers in orchestrating CSCL situations. *British Journal of Educational Technology* 46, 2 (2015), 330–343.

[23] Niall Sclater, Alan Berg, and Michael Webb. 2015. Developing an open architecture for learning analytics. *EUNIS Journal of Higher Education* (2015).

[24] Charles Severance, Ted Hanss, and Josepth Hardin. 2010. Ims learning tools interoperability: Enabling a mash-up approach to teaching and learning tools. *Technology, Instruction, Cognition and Learning* 7, 3-4 (2010), 245–262.

[25] Boban Vesin, Katerina Mangaroska, and Michail Giannakos. 2018. Learning in smart environments: user-centered design and analytics of an adaptive learning system. *Smart Learning Environments* 5, 1 (2018), 24.

[26] Feng Wang and Michael J Hannafin. 2005. Design-based research and technology-enhanced learning environments. *Educational technology research and development* 53, 4 (2005), 5–23.